



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Název:	Analýza domény Data Management Planning a souvisejících nástrojů
Student:	Tereza Machačová
Vedoucí:	Ing. Marek Suchánek
Studijní program:	Informatika
Studijní obor:	Informační systémy a management
Katedra:	Katedra softwarového inženýrství
Platnost zadání:	Do konce letního semestru 2020/21

Pokyny pro vypracování

Plánování správy dat je důležitou součástí různých vědeckých i komerčních projektů. Cílem práce je vysvětlit základní pojmy a souvislosti v této doméně a analyzovat nástroje pro tvorbu plánů. Hlavním výstupem práce bude detailní analýza nástrojů včetně jejich zhodnocení a porovnání.

- Seznamte se s doménou plánování správy dat (Data Management Planning), analyzujte klíčové pojmy a procesy s využitím metod konceptuálního modelování.
- Proveďte stručnou rešerši existujících nástrojů pro tvorbu plánů správy dat.
- Detailně analyzujte vybrané nástroje z pohledu různých typů uživatelů (persony), zaměřte se na funkce i na UI/UX. Na základě funkcí porovnejte nástroje a popište, jak funkce podporují činnosti v doméně.
- Popište technologie použité v nástrojích a možnosti použití nástrojů v rámci organizace i jednotlivci.
- Analyzujte a popište business modely nástrojů a vyhodnoťte jejich přínosy pro uživatele a organizace z ekonomicko-manažerského hlediska.

Seznam odborné literatury

Dodá vedoucí práce.

Ing. Michal Valenta, Ph.D.
vedoucí katedry

doc. RNDr. Ing. Marcel Jiřina, Ph.D.
děkan

V Praze dne 11. prosince 2019



**FAKULTA
INFORMAČNÍCH
TECHNOLGIÍ
ČVUT V PRAZE**

Bakalářská práce

Analýza domény Data Management Planning a souvisejících nástrojů

Tereza Machačová

Katedra softwarového inženýrství
Vedoucí práce: Ing. Marek Suchánek

1. června 2020

Poděkování

Ráda bych poděkovala vedoucímu své práce, Ing. Marku Suchánkovi, za vedení práce, konzultace, poskytnuté materiály a i všechnu další pomoc. Zejména za ochotný a přátelský přístup a velmi rychlé zodpovídání dotazů.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principu při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisu, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 1. června 2020

.....

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2020 Tereza Macháčová. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení na předchozí straně, je nezbytný souhlas autora.

Odkaz na tuto práci

Macháčová, Tereza. *Analýza domény Data Management Planning a souvisejících nástrojů*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2020.

Abstrakt

Plánování správy dat je důležitou součástí různých vědeckých i komerčních projektů. Je potřeba rozhodnout, jaká data budou vytvořena a jak s nimi bude nakládáno. Jak a jestli vůbec budou sdílena, kdo k nim bude mít přístup a jak všechny tyto požadavky zajistit. Práce se zabývá vysvětlením základních pojmů a souvislostí v této doméně, a především detailní analýzou nástrojů. Jedná se o online nástroje sloužící k tvorbě plánů správy dat. Nástroje jsou v práci popsány a zhodnoceny.

Klíčová slova DMP, plánování správy dat, plán správy dat, DMPTool, DM-POne, Data Stewardship Wizard, DS Wizard, planning tool, FAIR data, data steward, machine-actionable DMP

Abstract

Data management planning is an important part of scientific and commercial projects. It is needed to decide what type of data will be produced and how the data will be treated. If they will be shared and in which way, who will have an access to the data and how to ensure that all these requirements are fulfilled. This bachelor thesis deals with explanation of basic terms and context in this domain and mostly detail analysis of tools. These mentioned tools are online tools used to create data management plans and are described and evaluated in the thesis.

Keywords DMP, data management planning, data management plan, DMP-Tool, DMPOnline, Data Stewardship Wizard, DS Wizard, planning tool, FAIR data, data steward, machine-actionable DMP

Obsah

Úvod	1
1 Cíl práce	3
2 Analýza domény plánování správy dat	5
2.1 Úvod do plánování správy dat	6
2.2 Požadavky investorů na plány správy dat	8
2.2.1 Šablony	9
2.3 Machine-actionable DMP	10
2.4 Aktivní plány správy dat	16
2.5 Úvod do dat	16
2.6 FAIR data	17
2.6.1 Nalezitelná	17
2.6.2 Přístupná	18
2.6.3 Interoperabilní	19
2.6.4 Znovupoužitelná	20
2.6.5 FAIRifikační proces	22
2.6.6 The FAIR Funder	23
2.7 Data stewardship	27
2.7.1 FAIR data stewardship pro life science	28
2.8 DMP Common Standards Working Group	29
3 Analýza nástrojů	33
3.1 DMPOnline	36
3.1.1 Vytváření plánu	36
3.1.2 Reference	39
3.1.3 Technická stránka a business model	40
3.2 DMPTool	41
3.2.1 Vytváření plánu	41

3.2.2	Technická stránka a business model	42
3.3	Data Stewardship Wizard	43
3.3.1	Vytváření plánu	43
3.3.2	Editor znalostního modelu	46
3.3.3	Vyhodnocení ceny úložiště	48
3.3.4	Technická stránka a business model	49
3.4	DataWiz	50
3.4.1	Vytváření plánu	50
3.4.2	Studie	52
3.4.3	Materiály pro projekt	53
3.4.4	Technická stránka a business model	53
3.5	easyDMP	54
3.5.1	Vytváření plánu	54
3.5.2	Technická stránka a business model	57
3.6	ezDMP	58
3.6.1	Vytváření plánu	58
3.6.2	Technická stránka a business model	59
3.7	QUT DMP	60
3.7.1	Vytváření plánu	60
3.7.2	Technická stránka a business model	60
3.8	DMP OPIDoR	62
3.8.1	Vytváření plánu	62
3.8.2	Technická stránka a business model	63
3.9	CEDAR	64
3.9.1	Technická stránka a business model	65
4	Zhodnocení nástrojů	67
4.1	Přízpůsobení	68
4.1.1	Výběr šablony	69
4.1.2	Úprava šablony nebo vlastní šablona	70
4.1.3	Jazyk nástroje	72
4.1.4	Možnosti exportu	72
4.1.5	Správa jednotlivcem či organizací	74
4.1.6	Přidání dalších informací	75
4.1.7	Registrace a přihlašování	77
4.2	Machine-actionability	79
4.3	Zhodnocení UI/UX dle desatera Jakoba Nielsena	82
4.3.1	Viditelnost stavu systému	83
4.3.2	Shoda mezi systémem a realitou	84
4.3.3	Uživatelská svoboda	85
4.3.4	Konzistence a standardy	87
4.3.5	Prevence chyb	88
4.3.6	Rozpoznání před vzpomnutím	89
4.3.7	Flexibilita a efektivita použití	91

4.3.8	Estetický a minimalistický design	92
4.3.9	Pomoc s pochopením chyb a jejich vyřešením	93
4.3.10	Pomoc a dokumentace	95
4.4	Závěrečné hodnocení	97
	Závěr	101
	Literatura	103
	A Obsah přiloženého média	107

Seznam obrázků

2.1	Diagram popisující fáze projektu The FAIR Funder	24
2.2	Model vytvořený DMP Common Standards Working Group	30
3.1	Úvodní otázky nástroje DMPOnline	36
3.2	Nástroj DMPOnline.	38
3.3	Přehled šablon investorů v DMPOnline	40
3.4	Nástroj DMPTool	42
3.5	Nástroj Data Stewardship Wizard	44
3.6	Příklad grafu metrik v Data Stewardship Wizard	45
3.7	Vyhodnocení ceny úložiště v Data Stewardship Wizard	48
3.8	Nástroj Datawiz	52
3.9	Studie v Datawiz	53
3.10	Nástroj easyDMP	55
3.11	Nástroj ezDMP	58
3.12	Nástroj QUT DMP	61
3.13	Nástroj DMP OPIDoR	62
3.14	Nástroj CEDAR	65

Seznam tabulek

4.1	Shrnutí možností přizpůsobení	78
-----	---	----

Úvod

Plánování správy dat je důležitou součástí různých vědeckých i komerčních projektů. Je potřeba rozhodnout, jaká data budou vytvořena a jak s nimi bude nakládáno. Jak a jestli vůbec budou sdílena, kdo k nim bude mít přístup a jak všechny tyto požadavky zajistit. [1]

Plány správy dat jsou s postupem času stále více vyžadovány a jsou na ně kladeny vyšší nároky. Statické dokumenty, ke kterým se již lidé nevraceli, se mění na aktivní dokumenty spolupracující mezi několika systémy. Do práce na těchto dokumentech se zapojuje více lidí, mezi kterými jsou také experti z různých oblastí. Práce vysvětluje, proč je tento přechod žádaný, jaké přinese výhody, a jak lze tuto změnu realizovat. [2]

Plány lze vytvářet a spravovat v online nástrojích. Takovýchto nástrojů existuje mnoho a stále se objevují nové. Nástroje mohou být přizpůsobené potřebám konkrétních investorů a obsahovat tak šablony těchto investorů, které zjednodušují a zpřehledňují tvorbu plánu. [3]

V práci je analyzováno a hodnoceno celkem osm nástrojů. Jsou hodnoceny nejen funkcionality nástrojů, ale i uživatelská přívětivost či podporování investorů. Práce tak prezentuje nástroje a jejich silné a slabé stránky, dle kterých je jednodušší si zvolit vhodný nástroj pro konkrétní potřeby.

Termín „účastník projektu“, používaný zejména v první části práce, je do češtiny přeložený anglický výraz „stakeholder“. Stakeholdery jsou přímí účastníci projektu, jako například výzkumníci, ale také investoři, osoby spravující repozitáře nebo zabývající se licencemi pro projekt. Účastníky projektu jsou tak v práci myšleny všechny osoby, které splňují označení stakeholder.

Z důvodu delšího názvu pojmu „plán správy dat“ a častého opakování je používáno i kratší označení „plán“. Pokud není řečeno jinak, plánem je myšlen plán správy dat. Výjimečně je použit i anglický ekvivalent se zkratkou DMP, a to převážně v názvech různých nástrojů či projektů a souvislosti s nimi.

Cíl práce

1. Seznámit se s doménou plánování správy dat (Data Management Planning), analyzovat klíčové pojmy a procesy s využitím metod konceptuálního modelování.
2. Provést stručnou rešerši existujících nástrojů pro tvorbu plánů správy dat.
3. Detailně analyzovat vybrané nástroje z pohledu různých typů uživatelů (persony), zaměřit se na funkce i na UI/UX. Na základě funkcí porovnat nástroje a popsat, jak funkce podporují činnosti v doméně.
4. Popsat technologie použité v nástrojích a možnosti použití nástrojů v rámci organizace i jednotlivce.
5. Analyzovat a popsat business modely nástrojů a vyhodnotit jejich přínosy pro uživatele a organizace z ekonomicko-manažerského hlediska.

Analýza domény plánování správy dat

V teoretické části bakalářské práce bude popsána doména plánování správy dat. Bude vysvětleno k čemu slouží, jaké jsou její výhody a jaké jsou naopak její slabiny. Následující sekce budou více přibližovat konkrétní aspekty plánů správy dat, se zaměřením na nedostatky a jejich možným řešením. Různé projekty, iniciativy a pracovní skupiny dnes globálně pracují na odstranění těchto slabin s cílem změnit současné nepříliš efektivní plány. Výsledkem by měly být efektivní plány propojené s různými systémy a nástroji, umožňující pokročilou automatizaci, která ulehčí práci všem účastníkům projektu a zároveň pomůže i osobám mimo projekt. Ať už jiným výzkumníkům, kteří mohou využít použitou šablonu plánu správy dat nebo projektem získaná data, tak i široké veřejnosti, která se buď sama a nebo přes publikaci čerpající z dat projektu, dozví o dosažených výsledcích.

2.1 Úvod do plánování správy dat

Plánování správy dat se zabývá plány správy dat, což jsou dokumenty obsahující informace o datech projektu.

Plán typicky obsahuje a zodpovídá tato témata:

- účel získání dat,
- typ dat,
- způsob získání dat,
- nakládání s daty během výzkumu a po jeho skončení,
- místo uložení dat,
- doba uložení dat,
- použité licence,
- sdílení dat, případně open access data,
- osoby mající přístup k datům,
- platnost přístupu,
- použité standardy,
- případná pravidla a požadavky od tvůrců dat či investorů.

Plány pro správu dat plní mnoho funkcí, bez kterých by se mnoho projektů v dnešní době nejspíše ani neuskutečnilo. Popisují, jaká data budou získána, jakým způsobem, a jak s nimi bude nakládáno. Jak během projektu, tak i po jeho skončení, řeší tedy i uchování dat. Zohledňují také formu dat a z toho vyplývající nejvhodnější přístup. Zajišťují celkový přehled a zaručují kvalitu a efektivitu práce z důvodu předcházení chybám. Definují možnosti sdílení a přístupu pro využití dat a poznatků i po skončení projektu, které mohou být využity pro budoucí projekty, jež tak mohou být rozsáhlejší. Obsahují informace o tom, kde budou data archivovaná a pod jakými licencemi, za jakým účelem a kým byla získána. Zároveň je možné díky plánům pro správu dat ověřit, zda je projekt v souladu s legálními a etickými normami. Plán také zjednodušuje proces obměny personálu, jelikož nový personál má k datům přístup a rychleji se tak zapojí do fungování projektu. Dobře zpracovávaná a ukládaná data také vylučují možnost, že by se stejná práce dělala víckrát než je nutné a vznikaly by duplicity. [1]

Důležitým faktorem je i informovanost a přehled investorů, aby věděli, na co budou využity jejich finanční prostředky, mohou průběh a výsledky projektu s plánem porovnávat. Mnoho investorů vyžaduje plán správy dat nejen

v průběhu financování projektu, ale už ve fázi domluvy grantu. Poskytovatelem financí neboli investorem jsou typicky nějaké regionální, státní nebo mezinárodní instituce, např. ministerstvo nebo Evropská unie. Případně také organizace vytvořené za účelem podpory vědy, které získávají finance pro přerozdělování. [1]

Plánování správy dat je, jak vyplývá z výše uvedených poznatků, důležitou součástí projektu. Plán bohužel často není považován za nedílnou součást výzkumného procesu, ale spíše za administrativní nutnost, která doplňuje žádost o grant. Plán tak bývá vytvářen jako statický dokument ještě před začátkem projektu a nepodporuje k žádným aktivitám. Některé otázky mohou zůstat nezodpovězené, některé zůstávají příliš nejasné a obecně zodpovězené kvůli možné otevřené formě odpovědi. [2]

Plán pro správu dat by ale měl být nedílnou součástí projektu, a to ve všech jeho fázích. Již jen z důvodu, že většina projektů ve všech odvětvích využívá nějakým způsobem digitální úložiště či digitální nástroje. Plán by s nimi mohlo být možné propojit a tím usnadnit práci. V plánu by měly být popsány metody, poznatky, pravidla a další informace, a dle toho, jak by se v průběhu výzkumu měnily, by se měl měnit i plán, aby stále zaznamenával aktuální informace. Neměl by být statickým dokumentem. Stále je potřeba, aby plány byly pohodlně čitelné a srozumitelné lidem, ale zároveň se stále častěji a intenzivněji vyskytuje požadavek na *machine-actionable plány správy dat*, které by strojově dokázaly podstatnou část práce zautomatizovat. Zavedení machine-actionable plánů by tak mělo přínos pro všechny účastníky projektu, a i pro další osoby. Jelikož by všechna získaná data byla dohledatelná a přístupná veřejnosti (alespoň když je projekt hrazen z veřejných prostředků je toto zvykem), o výsledcích projektu se dozví více lidí a projekt tak může mít větší dopad a efekt. V praxi tomu tak bohužel zatím nebývá. S těmito změnami souvisejí komplikace z důvodu potřebného zajištění možností přístupu, sdílení a licencí. Nejde jen o uložení dat, aby byla k dispozici, ale také o zajištění veškeré údržby, a to i po skončení projektu. [1, 2]

Otázkami týkajícími se praktické realizace a rozšíření machine-actionable plánů se zabývají různé pracovní skupiny, projekty, či pilotní program, které budou popsány v následujících sekcích.

2.2 Požadavky investorů na plány správy dat

V této sekci jsou informace o požadavcích investorů, ať už jde o formu či obsah plánu. Je zde popsáno, jak těmto požadavkům vyhovět, co přitom lze použít a na které skutečnosti by si měli žadatelé o grant dát pozor nebo naopak které části zdůraznit, aby měli větší šanci získat financování pro svůj projekt.

Některé evropské grantové agentury již požadují plán správy dat a někteří investoři požadují plán správy dat už před samotným financováním projektu, jako součást žádosti o grant. Je nežádoucí, aby se pracně získaná data financovaná z veřejných zdrojů ztratila kvůli nedostatečné správě. Plán správy dat je tak zároveň „pojištěním“, že data nezůstanou uložena pouze lokálně v počítači jednoho z výzkumníků a poté zapomenuta či smazána. To totiž není vůbec výjimečným případem a těmto situacím lze dodržováním plánu správy dat zabránit. Nestačí v plánu pouze uvést informace o zacházení s daty, je nutné jejich dodržování vynucovat a kontrolovat. [4]

Investoři většinou nespecifikují požadovaný formát či standardy, pracovník projektu vytvářející plán správy dat tak zvolí nejvhodnější způsob a popíše, proč se rozhodl zrovna takto, proč je zvolený způsob pro danou situaci vhodný. Podobný případ nastává, pokud by data neměla být přístupná všem, ale jen nějaké skupině osob. Je to možné, ale je nutné v plánu vysvětlit, za jakým účelem tomu tak je. Je vhodné zdůraznit, jaké kroky se podniknou pro ochranu osobních údajů účastníků průzkumu, jak dosáhnout anonymizace, a přitom mít detailní data přístupná co nejvíce lidem. Tímto lze investorům dát najevo snahu o ochranu osobních údajů, zároveň snahu poskytnout výsledky projektu co nejvíce lidem a že jsou tyto dva důležité body v rovnováze, což je ceněno. [1]

Z plánu by mělo být jasné patrné, kdy budou data zveřejněna, kde, a jakým způsobem. Pokud by neměla být zveřejněna všem nebo by měla být zveřejněna se zpožděním, je nutné vysvětlit, z jakého důvodu, a na jak dlouho by tento stav případně byl. [5]

Jak již bylo zmíněno, v současné době probíhají projekty a programy snažící se změnit volnou statickou formu plánu na specifikovanější, spolupracující se stroji. Součástí jsou i projekty, zaměřující se přímo na investory, kteří nemusí mít potřebné znalosti na vyžadování takového plánu a jeho kontrolu. Například náplní programu The FAIR Funder je dokázat, že i bez větších znalostí těchto plánů, jejich fungování a plánování správy dat samotného je možné, aby investoři tyto plány správy dat reálně vyžadovali a dokázali i kontrolovat, zda jsou plněny. Více o tomto projektu je v bakalářské práci v sekci 2.6.6 The FAIR Funder.

2.2.1 Šablony

V plánech je zapotřebí zodpovědět všechny otázky, které se mohou v závislosti na konkrétním investorovi či projektu lišit, ale zároveň je obvykle vyžadováno, aby plány byly stručné a přehledné. Investoři chtějí spíše shrnutí, příliš dlouhé opakující se texty jsou pro ně nežádoucí. S tvorbou plánů správy dat tak pomáhají nástroje, které už mají i předpřipravené různé šablony požadavků a investor tak může vyžadovat plán dle určité šablony, které mohou mít různé požadavky či omezení.

Mezi šablony patří pro představu BBSRC, kde je vyžadován plán na sdílení dat, který může obsahovat informace a detaily typů dat a jejich formátů, standardů a metod sdílení dat. Dále například NERC, který má povinnou jedinou stránku a pro každé NERC je vyžadován Outline Data Management Plan (ODMP) sloužící k archivaci a znovupoužití dat po ukončení spolupráce či skončení grantu. Šablona British Heart Foundation má zase podmínkou grantu, že získaná data musí být předána organizaci British Heart Foundation přes jejich systém Researchfish. [1]

2.3 Machine-actionable DMP

Tato sekce se věnuje popisu machine-actionable plánů správy dat dle [2]. Zároveň vysvětluje obvyklý pohled na klasické plány a proč je současný způsob jejich využití neefektivní. Hlavním cílem těchto plánů je vylepšit tradiční plány správy dat, a kvalitu dat a *metadat* (strukturovaná data popisující další data a zdroje) díky automatizaci, čímž se navíc sníží administrativní náročnost. Jsou popsány i další vyplývající výhody a kroky, jak této změny dosáhnout. Pojem machine-actionable lze vysvětlit jako „informace strukturovaná konzistentním způsobem tak, že stroje a počítače mohou být naprogramované pro snadné zpracování a vyhodnocování této struktury a jejího obsahu“.

V souvislosti se zmíněným pohledem na plány správy dat pouze jako na administrativní nutnost se rozšiřuje názor, že místo klasických plánů by se měly využívat spíše machine-actionable plány, což zvýší jejich hodnotu pro všechny zúčastněné strany – výzkumníky, investory či různé managery a administrátory. Výzkumná komunita se snaží dosáhnout cíle, kdy plány budou machine-actionable a budou umožňovat výměnu informací napříč nástroji a systémy. Také se plány správy dat stanou součástí již existujících pracovních postupů a určité části plánů tak budou automaticky generovány a sdíleny, čímž se zvýší kvalita informací a sníží náročnost a množství administrativní práce. Součástí projektu a plánu je i případný souhlas k použití dat, který se uvádí v začátcích projektu, ještě před samotným sběrem dat. S využitím machine-actionable plánů mohou ostatní uživatelé tyto informace jednoduše dohledat a citovat, jelikož součástí dat je i informace o jejich tvůrcích.

Aby bylo možné machine-actionable plány správy dat zrealizovat, je potřeba do plánů začlenit další klíčové informace o projektu a jeho výstupech. Je možné začít na lokální úrovni, od konkrétní výzkumné instituce, což se i předpokládá, jelikož okamžité využití machine-actionable plánů globálně je téměř nereálné. Postupná realizace by započala využitím datových modelů pro výměnu informací a sdíleným ekosystémem služeb s možností zaslání notifikací. Bude nutné zavést machine-actionable persistentní identifikátory. Realizací výše zmíněných lze dosáhnout snížení množství práce většiny účastníků a zároveň zlepšit kvalitu zaznamenávaných informací. Například, po zadání dostatku informací bude možné využít služby, které navrhnou licenci a odhadovanou cenu úložiště, což opět usnadní práci. Bude možné využít stávající služby či se využijí služby nové, které v souvislosti s novými plány teprve vzniknou.

Výhod pro každého z účastníků je opravdu mnoho. Přínosem pro investora jsou aktuální strukturované informace o tom, kdo bude data vytvářet, kde budou uložena a pod jakou licencí. machine-actionable plány správy dat dokáží navíc mnoho z těchto informací monitorovat a díky tomu se zjednodušuje proces kontroly dodržování podmínek. Výzkumník se přes projekt může připojit k expertům a zeptat se na radu. Zautomatizované procesy mohou velmi usnadnit tvorbu plánu správy dat a umožní ostatním ho aktualizovat

a měnit. Plán bude zároveň důležitým zdrojem informací k designu a implementaci. Plán také umožní automatickou generaci přehledu o dostupnosti dat, kde vydavatel nalezne informace o tom, za jakých podmínek může data použít a kde je získat, což je užitečné pro uživatele, kteří mohou chtít získaná data využít. Přehled totiž obsahuje odkazy na *datasety* (datové sady) a umožňuje tak citaci článků a datasetů. A díky již zmíněnému navrhování odhadovaných cen úložiště, licence a množství dat se usnadní práce i správci úložiště, protože plán s dostatečnými informacemi umožňuje kapacitní plánování a usnadňuje uchování dat.

Realizaci machine-actionable plánů správy dat popisuje deset zásad. Ty také popisují, jaké akce jednotliví účastníci již podstupují, nebo by podstupovat měli, aby bylo možné společně pracovat napříč výzkumnými komunitami za dosažením vyšších cílů. Všechny zásady jsou stejně důležité a nezáleží na jejich pořadí. Zásady jsou nezávislé na nástrojích a použitých technologiích, taktéž nesouvisí s žádnou konkrétní šablonou plánu správy dat nebo konkrétní organizací. Nevyžaduje se uskutečnění všech zásad najednou globálně, uskutečnění je možné postupně, od jednotlivých organizací a projektů až po vytvoření rozsáhlé sítě služeb.

1. Začlenit plány správy dat s pracovními procesy všech účastníků do ekosystému dat.

Vytvářet a spravovat plán správy dat by nemělo být záležitostí jediné osoby, měla by to být skupinová činnost, kde jednotliví účastníci sdílejí znalosti ze svého oboru a zároveň kontrolují informace poskytnuté ostatními, že jsou navzájem v souladu. Tak se zajistí, že jsou přítomny všechny informace a že jsou správné. Obvykle jsou plány vytvořeny před začátkem projektu a poté už se nemění ani nepoužívají. Zde se ale k plánům budou účastníci vracet, aktualizovat je a nadále využívat informace v nich obsažené.

2. Povolit automatizovaným systémům jednat za účastníky.

Některé z informací zachycených v plánech správy dat jsou již dostupné elektronicky, takže místo opětovného zadávání by bylo výhodné použít informace z vhodných zdrojů, tedy tyto informace dohledat a porovnat ještě s dalšími zdroji pro zaručení kvality.

S potřebnými informacemi by systém měl být schopen předvyplnit plány správy dat, odhadnout náklady na úložiště a jeho správu a toto úložiště případně i zarezervovat. Také by měl dokázat navrhnout vhodnou licenci a průběžně ověřovat, zda jsou data dostupná tak jak je uvedeno v plánu a pod správnou licenci. Automatizace je tedy v různých částech projektu možná pro většinu účastníků, což ušetří čas i náklady a poskytne přesnější informace.

2. ANALÝZA DOMÉNY PLÁNOVÁNÍ SPRÁVY DAT

Pokud je vyžadován zásah člověka, systém by měl zaslat notifikaci či vytvořit ticket se zadaným úkolem (v systémech pro hlášení chyb a správu úkolů, ticket obsahuje informace o konkrétním úkolu či problému) a přiřadit ho určité osobě, která buď bude umět daný problém vyřešit nebo takovou osobu kontaktuje.

3. Mít podmínky i pro stroje.

Některá pravidla a podmínky mohou být dána nejasně, obvykle příliš obecně, a je tak těžké ověřit jejich plnění i pro lidi, natož pro stroje, které potřebují specifičtější informace. Pokud je podmínka zadána jako „data budou po celou dobu projektu spravována pod nejvyššími standardy“, jak poznat, zda jsou data v souladu s touto podmínkou? Je proto vhodné pravidla nejen formulovat přesněji, ale formulovat je tak, aby se jimi mohly řídit i stroje. Ty se jimi budou při správné formulaci řídit za každých podmínek.

4. Popsat komponenty ekosystému správy dat.

Zvolit vhodný *repozitář* (úložiště) pro uložení dat je obvyklým problémem. Na výběr je řada repozitářů které se liší náklady, licencemi, typem akceptovaných dat a množstvím dat které pojmu. Tyto informace o repozitáři ale nejsou na jednotném místě, mohou být obsažené v podmínkách užití, často pokládaných dotazech (FAQ), či nejsou vůbec viditelně specifikovány a je možné se je dozvědět až po vyžádání.

Pokud ale budou systému popsány specifické komponenty těchto ekosystémů, bude dle druhé zásady možné vytvořit službu, která s využitím informací obsažených v DMP navrhne vhodný repozitář. Nejednalo by se o vytváření celého výše popsaného úplně od začátku bez jakýchkoliv základů. Existují registry, re3data¹ a OpenDOAR² které obsahují rozhodující informace o tisících datových repozitářích, jako jsou typy obsahu a podmínky. Každý registr je spravován manuálně a každý repozitář je před přidáním do seznamu řádně zkontrolován. Re3data poskytuje informace otevřeně přístupné machine-actionable formou a v současnosti pracuje ve spolupráci na doporučovací službě, která přispěje k umožnění projektů s daty dle *iniciativy FAIR*, která říká, že data mají být naleziitelná, přístupná, interoperabilní a znovupoužitelná. Iniciativa FAIR je popsána v sekci 2.6.

5. Používat PID a kontrolované slovníky.

Aby plány správy dat byly srozumitelné a jasné pro všechny účastníky, je nutné, aby se používaly správně a přesně definované termíny a precizní identifikace zdrojů. Zde ale mohou nastat dvě různé komplikace, protože

¹<http://re3data.org/>

²<http://v2.sherpa.ac.uk/opensoar/>

do plánů se informace většinou zapisují jako otevřený text. Může tak nastat situace, kdy jsou informace příliš obecné, nebo příliš specifické. Pokud jsou informace příliš obecné, bývá nejasné, jak se tedy s daty má zacházet a je složité posoudit splnění kritérií. Pokud jsou ale informace příliš specifické, je často vyžadována dobrá znalost zmiňovaných termínů a souvislostí, kterou mohou někteří účastníci, obvykle ti s odlišným zaměřením, postrádat. Informace tak mohou být srozumitelné pouze pro úzkou skupinu osob.

Z tohoto důvodu je potřeba používat *kontrolované slovníky* a *persistentní identifikátory*. Kontrolované slovníky obsahují seznam dobře definovaných pojmů, které lze použít k popisu dat a omezují tak výběr slov na popsání dat, jelikož účastníci vybírají z tohoto limitovaného seznamu. Persistentní identifikátory (se zkratkou PID) identifikují a lokalizují zdroje. Mohou se vztahovat k osobám, publikacím, datasetům, repozitářům, organizacím či dalším elementům z datového ekosystému. Výzkumníci mohou být popsáni pomocí ORCID, což je identifikátor vědeckých a dalších akademických autorů, a existují i další specializované identifikátory. V machine-actionable plánech správy dat by namísto odpovědí formou otevřeného textu, jak je tomu v tradičních plánech, mělo být zadávání informací umožněné vybráním ze seznamu možností. Předejde se tak obecnému a často používanému spojení, jako je například „data budou po celou dobu projektu spravována nejosvědčenějšími metodami a nejvyššími standardy“.

6. Využívat společný datový model.

Datový model je médiem pro výměnu informací mezi účastníky. Informace jsou poskytovány machine-actionable formou, což umožňuje vzájemnou spolupráci různých systémů a dovoluje službám jednat za účastníky. Datový model není určen žádnou šablonou nebo dotazníkem, vzhledem k širokému rozsahu témat by měl být modulární. Jádro modelu však bude pro všechny plány stejné a model by měl využívat existující standardy a kontrolované slovníky. Datový model je využíván nástroji pro zapisování a čtení informací do a z plánu a schopnost jednat na základě informací v něm uvedených.

Vytvořením společného datového modelu se zabývá pracovní skupina *DMP Common Standards Working Group*, která je více popisována přímo v sekci 2.8. Skupina se s pomocí společného datového modelu snaží dosáhnout zavedení společných standardů. Výsledkem bude lepší schopnost systémů vzájemně spolupracovat a možnost kontrolovat a sdílet informace obsažené v plánech. Skupina se skrze společné informační modely snaží zrealizovat vizi, kdy plány správy dat budou plně začleněny do systémů i pracovních postupů a s tímto vědomím budou vytvářeny a udržovány. Modulární návrh umožní přizpůsobení a rozšíření

dle současných standardů a slovníků, aby co nejlépe vyhovoval potřebám různých komunit. Nástroje a systémy budou schopné z plánu číst a zapisovat do něj. [6]

7. Umožnit použití plánu správy dat pro lidi i stroje.

Plány správy dat v současné době nejenže nebývají využívány v průběhu celého projektu ale jen na začátku, ale někdy nebývají využívány vůbec. Ne všichni investoři je považují za nutnou součást žádosti o grant a když jsou plány přítomny, jsou přečteny několika málo účastníky a poté odloženy a dále nevyužity.

Převedením plánu na takzvaný *živý dokument* (z anglického „living document“), tedy dokument, který se mění a je aktualizovaný, se zvýší pravděpodobnost využití plánu a jeho diskutování i v průběhu projektu. A to zejména pokud nebude dostupná pouze nejnovější verze plánu, ale budou dostupné i předchozí verze a budou moci být posouzeny rozdíly mezi nimi, a to jak lidmi, tak stroji. Dalším přínosem by byly nastavitelné notifikace změn v DMP, které by bylo možné povolit různě pro jednotlivé účastníky, například jen pro některé části plánu. To opět vede k využití co nejstrukturovanějších informací a vyhnutí se otevřenému textu. Některý obsah pravděpodobně bude muset zůstat formou otevřeného textu, napsán člověkem pro člověka, a nikoliv pro stroj, tím ale bývá obvykle obsah, který stroj nedokáže zpracovat, nebo není zpracování strojem doporučováno, jelikož je nedokáže dostatečně dobře posoudit, typicky etické záležitosti.

8. Podporovat vyhodnocování a monitorování správy dat.

Investoři a tvůrci pravidel řídí poptávku plánů správy dat a z těchto důvodů musí plány obsahovat i nějaké monitorování dodržování podmínek. Plány musí být vyhodnocovány jak na začátku projektu při žádosti o grant, tak i v jeho průběhu. Mnoho účastníků stále potřebuje informace srozumitelnou formou psanou otevřenou formou člověkem, ale pokud budou alespoň pravidla a podmínky machine-actionable, přispějí k automatizaci monitorování. Nástroje budou kontrolovat, zda je vybraná licence pro sdílení dat v souladu s podmínkami investora, zda jsou data ukládána do uvedeného repozitáře, zda jsou opravdu přístupná. Plány správy dat by vždy měly odrážet skutečný stav, i když by byl odlišný od doporučených osvědčených metod na správu dat.

9. Změnit plány správy dat na aktualizovatelné verzované dokumenty.

Plány správy dat by neměly být jen statickým plánem ale aktualizovatelným dokumentem mapujícím současný stav správy dat. Aktualizování plánu nemusí vždy vyžadovat zásah člověka, některé změny mohou být zautomatizované. Součástí by měly být automatické notifikace na změny, zmiňované v předchozích bodech.

10. Zpřístupnit plány správy dat veřejnosti.

Z předchozích bodů je již známo, jaké přínosy bude mít sdílení plánu pro účastníky projektu. Další přínosů lze dosáhnout zpřístupněním plánu i za hranice projektu. Umožní se porovnávání plánů mezi sebou, posouzení podle konkrétních parametrů, vyvození závěrů, jejich využití, informování ostatních o těchto závěrech. Informace mohou být obecnější a přispěje se tak ke standardizaci plánů a obeznámení s nejlepšími osvědčenými praktikami pro správu dat, nebo specifitější, určené pro konkrétní publikum. Sdílení plánu je také vhodné pro lepší orientaci v datech. Výsledky projektu budou díky zveřejnění veřejnosti zveřejněny i osobám, které se v získaných datech nemusí zorientovat. Nemusí mít potřebné znalosti a nemusí se jim podařit všechna potřebná data dohledat. Data mohou být různé velikosti, typů, mohou se lišit zásadami sdílení, a je tak možné, že nebudou uložena všechny zřetelně pohromadě, ale budou nějak rozdělena. Je složité zjišťovat, jaká data byla projektem nasdílena a kde je najít. Pomocí machine-actionable plánu správy dat by ale mělo být snadné dohledat ostatní části, protože pokud uživatel najde nějaký výstup projektu, součástí něj bude i plán, kde bude vše popsáno a bude obsahovat odkazy na jednotlivé části.

Machine-actionable plány by tedy měly oproti klasickým plánům tyto výhody:

- ucelenější správnější informace,
- informace jsou ve vhodném rozsahu a popsány srozumitelnými definovanými pojmy,
- snížení nutné (nejen) administrativní práce,
- jednodušší dohledání dat,
- jednodušší dohledání informací o projektu,
- vždy aktuální informace,
- předvyplnění plánu,
- monitorování dodržování plánu,
- služby pomáhající s výběrem úložiště a licence,
- lepší kapacitní a finanční plánování,
- možnost sledovat změny díky verzování plánu,
- využití notifikací,
- standardizace plánů.

2.4 Aktivní plány správy dat

Aktivní plány správy dat jsou další reakcí na klasické neefektivní plány správy dat. Data by dle skupiny ADMP IG (Active Data Management Plans Interest Group) měla být zachycena v aktivním DMP, který je možno rozšiřovat a doplňovat do něj algoritmy, původ dat a další informace. Zpětná vazba od uživatelů prokázala, že většina současných plánů je pro jejich potřeby nedostačující a nepodporuje opětovné použití dat v delším časovém horizontu. Předběžné nápady o konceptu aktivních plánů správy dat zahrnují zhodnocení existujících osvědčených způsobů a identifikaci požadavků na aktivní DMP z kompletního spektra účastníků – požadavky investorů, výzkumníků, ale i repozitářů a dalších. Zahrnují také specifikaci nástrojů a služeb na podporu tvorby a monitorování plánů a rozhraní, a výměnné formáty umožňující vzájemnou spolupráci těchto nástrojů a systémů. ADMP IG bude zároveň spolupracovat s dalšími skupinami se stejným zájmem. [7]

2.5 Úvod do dat

Dat, která jsou potřeba zpracovávat a uchovávat neustále přibývá. Nejen díky rozsáhlejšímu projektům a lepším možnostem výzkumu oproti dřívějším desetiletím, ale také díky digitalizaci. Výpočetní technika se rozšířila do všech oblastí lidského života. Všudypřítomné chytré telefony a Internet of Things také hodně přispívají k tomu, že je mnohem více informací než dříve a jsou i lépe a snadněji dostupné. [4]

Dá se také čerpat z dřívějších projektů a jejich poznatků, data potřebná pro daný průzkum mohou být sdílena jiným projektem, a lze tak využít tyto již dokázané závěry, než aby týmy ztrácely čas tím, že by zkoumaly stejný ověřený fakt znovu nezávisle na sobě. Obzvláště pokud nejsou ona potřebná data stěžejní, ale pouze vedlejší či podpůrnou záležitostí projektu. [1]

Větší množství dat a požadavky na sdílení činí plány pro správu dat stále důležitějšími, složitějšími a náročnějšími. Je důležité ukládat vše, není moudré některá data „zahodit“. Tato data se mohou ukázat jako podstatná až později a mohou vést k vyřešení závažných problémů, například k léčbě dosud nevyléčitelných nemocí. Některá data mají navíc takovou povahu, že je nelze získat znovu. [4]

Řešením těchto problémů se zabývá iniciativa FAIR a doména správy dat, kterou je *data stewardship*. FAIR popsaná v následující sekci specifikuje, jaké požadavky by data měla splňovat, aby byla co nejlépe využitelná. Data stewardship je proces zodpovědného nakládání s daty nejen během projektu, ale i po jeho skončení. Je mu věnována sekce 2.7.

2.6 FAIR data

Následující sekce vysvětluje principy iniciativy FAIR týkající se dat, jejich uložení a zacházení s nimi. Popisuje i z toho vzešlou *iniciativu GO FAIR* a proces, kterým se data upraví, aby splňovala principy FAIR. Druhá část sekce je zaměřená na pilotní program The FAIR Funder, jehož cílem je rozšířit užití FAIR dat. Podkladem první části je [8].

Data dle FAIR mají být:

- nalezitelná (findable)
- přístupná (accessible)
- interoperabilní (interoperable)
- znovupoužitelná (reusable)

Pojem byl použit poprvé v roce 2014 v Holandsku a principy vzešly z pracovní skupiny FORCE11. Tyto principy byly v roce 2016 publikovány v online časopise Scientific Data. Principy by měly být splnitelné počítačovými systémy s žádným či co nejmenším zásahem člověka, jelikož lidé stále více spoléhají na pomoc strojů se zpracováváním rostoucího množství dat. [4]

V následující části jsou vysvětleny principy FAIR. Popis každého principu se skládá ze seznamu bodů a detailnějšího vysvětlení vyplývajícího z těchto bodů. Dále jsou ke každému z principů připsány některé typové otázky dle [5], které by plán správy dat měl zodpovídat. FAIR nepokrývá aspekty týkající se kvality dat a etiky.

2.6.1 Nalezitelná

1. (Meta)data mají přiřazeny globálně unikátní persistentní identifikátory.

Data i metadata mají přiřazeny globálně unikátní identifikátory. Přiřazení těchto identifikátorů je zároveň nejdůležitějším bodem FAIR, jelikož ostatní body principů vychází z existence těchto identifikátorů. Je nutné, aby identifikátory byly persistentní a byly tak platné co nejdéle. Udržet webové odkazy aktivní je náročné, jelikož to stojí peníze a může se tak stát, že odkazy již nadále nebudou funkční. FAIR se snaží zachovat odkazy udržitelné, aspoň do určité míry.

2. Data jsou bohatě popsána metadaty.

Metadata musí obsahovat identifikátory dat jež popisují, a zároveň by měla popisovat kontext, kvalitu, stav a charakteristiku dat. Dostatek detailních metadat umožní počítači převzít zdoluhavou a náročnou práci třídění a správy kterou jinak musejí zastávat lidé. I bez identifikátoru dat by mělo být možné data dohledat dle informací získaných z metadat.

3. Metadata obsahují identifikátory popisovaných dat.

Metadata a jimi popisované datasety jsou obvykle různé oddělené soubory, a proto je potřebné, aby metadata obsahovala i identifikátory popisovaných dat.

4. (Meta)data jsou v přístupných zdrojích.

Identifikátory a bohatá metadata samy o sobě nezaručí, že data budou nalezitelná, pokud nikdo nebude vědět o jejich existenci. Je několik způsobů, jak zaručit, aby data byla objevitelná, jedním z nich je indexace, kterou používá například Google. Pro odborná data je ale zapotřebí pokročilejší, specifitější indexace.

Otázky, které by měl plán zodpovídat:

- Jaké jmenné konvence budou použity?
- Jaká metadata budou vytvořena?

2.6.2 Přístupná

1. (Meta)data lze získat skrze standardizovaný komunikační protokol.

Uživatelé většinou přistupují k datům kliknutím na odkaz, což je vlastně pokročilé rozhraní pro základní protokol TCP. K přístupu k FAIR datům by neměl být potřeba žádný specializovaný nástroj či komunikační metody, většina dat bude přístupná právě pomocí TCP.

- a) Protokol je přístupný, zdarma.

Protokol by měl být volně přístupný (open-access) a zdarma, aby bylo možné využít dat co nejvyšší. Každý by tak měl být schopen přistupovat alespoň k metadatům.

- b) Protokol umožňuje autorizační a ověřovací postupy tam, kde je to potřeba.

Možnost autorizačních a ověřovacích postupů je klíčový, avšak bohužel často nepochopený, chybně vyložený bod. Přístupná data nemusí nutně znamenat přístupnost pro každého, spíše popisuje, komu jsou data přístupná, a za jakých podmínek. Případně, jak lze tento přístup získat. Data mohou být přístupná jen vyhrazené skupině osob, v závislosti na povaze projektu či jeho fázi, a i přesto být FAIR. V ideálním případě by přístupnost měla být popsána tak, aby stroj pochopil požadavky a dokázal na jejich základě reagovat. Může tak například vyzvat uživatele k registraci či přihlášení.

2. Metadata jsou přístupná, i když již data nejsou dostupná.

Z důvodu finanční náročnosti na uchování a správu dat online se může stát, že data po nějakém časovém období nebudou dostupná. Stále by

ale měla být dostupná alespoň související metadata, která navíc bývají jednodušší a levnější na údržbu. Dle metadat lze dohledat informace o zdrojích dat, tedy instituce, publikace či konkrétní osoby, což může být velice užitečné.

Otázky, které by měl plán zodpovídat:

- Jaké metody nebo softwarové nástroje jsou potřeba k přístupu k datům?
- Je možné dodat příslušný software?
- Kde budou data, metadata, dokumentace a kód uloženy?
- Pokud bude užití dat nějak omezené, jak budou data přístupná?

2.6.3 Interoperabilní

1. (Meta)data pro reprezentaci znalostí používají formální přístupný sdílený jazyk s širokým použitím.

Interoperabilita je schopnost systémů vzájemně spolupracovat. Pro reprezentaci dat a metadat by měl být použit formální široce rozšířený jazyk, tak, aby různí lidé byli schopni si vyměňovat data a interpretovat je (je tedy nevhodné používat mrtvé jazyky). To stejné platí i pro počítače, i ty by měly být schopné data číst bez nějakých specializovaných algoritmů či překladačů. Každý systém by měl mít dle interoperability alespoň znalost formátů výměny dat druhého systému.

2. (Meta)data používají slovníky v souladu s FAIR principy.

Používané kontrolované slovníky popisující datasety by měly být zdokumentovány, tato dokumentace by měla být snadno nalezitelná a přístupná všem, kteří používají dataset. Měly by mít také přiřazeny identifikátory, stejně jako data a metadata.

3. (Meta)data obsahují kvalifikované reference na další (meta)data.

Kvalifikované reference jsou křížové reference (nebo také křížové odkazy), které vysvětlují svůj záměr. Reference tedy nejsou obecné, jako například vztah „X souvisí s Y“. Popisuje se, že daný dataset vychází z určitého jiného datasetu, nebo ve kterém datasetu lze dohledat doplňkové informace. Cílem je vytvořit mezi zdroji co nejvíce výstižných referencí.

Otázky, které by měl plán zodpovídat:

- Jaké standardy budou použity?
- Jaké slovníky pro data a metadata budou použity? Bude se jednat o standardní slovníky?

2.6.4 Znovupoužitelná

1. (Meta)data jsou bohatě popsána s množstvím přesných relevantních atributů.

Hlavním cílem FAIR je maximálně využít znovupoužitelnost dat. Data i metadata by tak měla být dobře popsána, aby mohla být replikována i za odlišných podmínek. Dobře značená data jsou lépe nalezitelná a znovupoužitelná. Metadata by neměla umožňovat jen data dohledat, ale měla by i obsahovat informace o podmínkách, za jakých byla data získána. A to, za jakým účelem byla shromážděna, kdy se tak stalo, nastavení parametrů, název a verze použitého softwaru. Dále je dobré zmínit, zda se jedná o surová, tzv. raw data, nebo byla zpracována, a ujistit se, že všechny názvy jsou dobře vysvětlené. Je vhodné upozornit na různé zvláštnosti či omezení dat, o kterých by uživatel, jenž bude data chtít použít, měl být informován.

- a) (Meta)data jsou vydána s jasnou licencí pro jejich využití.

Nejasnosti v právech pro použití dat by mohla silně a negativně ovlivnit práci s daty. Proto by veškeré licence i práva měly být jasně a zřetelně popsány, a to jak pro lidi, tak pro stroje.

- b) (Meta)data obsahují i původ dat.

Pro možné využití dat dalšími lidmi jsou důležité i informace o původu dat. Možní uživatelé by měli vědět, odkud data pochází, jak citovat zdroj. Součástí by tak měl být popis pracovního procesu. Kým byla data vygenerována či shromážděna? Byla již dříve publikována? Jsou zde obsažena i data z jiných zdrojů, která mohla být případně změněna či doplněna? Tento proces by měl být ideálně popsán machine-readable formou.

- c) (Meta)data splňují komunitní standardy příslušné domény.

Datasety které jsou si podobné, mající například stejný typ, způsob organizace a používající stejný společný slovník, je mnohem jednodušší využít. Pokud existují nějaké komunitní standardy či osvědčené postupy, měly by být dodrženy.

Otázky, které by měl plán zodpovídat:

- Kdy budou data uvolněna pro znovupoužití?
- Pokud data nebudou přístupná hned jak to bude možné, proč? Za jakým účelem? Na jak dlouho bude tento stav?
- Jak dlouho je zamýšleno, že data zůstanou znovupoužitelná?

FAIR není nový revoluční koncept, některé organizace už tyto principy používaly, například evropská organizace pro jaderný výzkum CERN. FAIR

tyto principy spíše přiblížil veřejnosti, obeznámil instituce se svými principy a přesvědčil o výhodách z nich vyplývajících. Iniciativa FAIR si tak rychle začala získávat pozornost a vznikla z jejích základů nová iniciativa GO FAIR, která se tyto principy snaží uvést do praxe v různých disciplínách. Součástí GO FAIR je i většina původních designerů FAIR, iniciativa GO FAIR tedy není od původního návrhu nijak vzdálena. Aby data splňovala principy FAIR, je zapotřebí práce *data stewardů* (správců dat, více v následující sekci 2.7 Data Stewardship). Těch bude v následujících letech zapotřebí značné množství, dle odhadů až 500 000 data stewardů světově. GO FAIR proto vytváří výukové materiály a workshopy zaměřené právě na data stewardship. [4]

Často se vyskytuje otázka, jaký je rozdíl mezi FAIR daty a open data (otevřená data). Tyto dva pojmy bývají často chápány jako synonyma, ale není tomu tak. Zatímco open data by měla být přístupná všem a každý je může používat a sdílet, druhý bod FAIR, kterým jsou přístupná data, znamená něco zcela jiného. Data mohou být přístupná jen vyhrazené skupině lidí, v závislosti na povaze projektu, jeho fázi, či účelu dat. Přístup se také může průběžně měnit, často právě v závislosti na fázi projektu, kdy je nejdříve přístupný jen projektovým pracovníkům, poté investorům a až nakonec široké veřejnosti. FAIR tak není synonymem pro open data.

FAIR principy souvisí s daty, metadaty a infrastrukturou (webové prohlížeče). Většina požadavků pro naležitelnost a přístupnost spadá na úroveň metadat. Interoperabilita a znovupoužitelnost je spíše záležitostí dat. Skupina DMP Common Standards WG (více v 2.8) vytvořeným modelem a společnými standardy přispěla k realizaci FAIR plánů správy dat. Jejich model již je implementován v mnoha nástrojích, což umožní výměnu informací mezi systémy a nástroji, a v dalších nástrojích bude teprve implementován. Tím bude možné pokrýt principy „I“ a „R“ – interoperabilitu a znovupoužitelnost. Principy naležitelnosti a přístupnosti by měly být dosaženy ukládáním plánů správy dat do vhodných repozitářů, což se zatím příliš nepraktikuje. [3]

2.6.5 FAIRifikační proces

FAIRifikační proces definovaný GO FAIR popisuje přeměnu dat na data dle iniciativy FAIR. FAIRifikační proces se skládá z následujících sedmi kroků:

1. Získání dat – získání přístupu k datům, která se budou FAIRifikovat.
2. Analýza získaných dat – analyzování struktury dat a relací mezi nimi.
3. Definice sémantického modelu – definuje se sémantický model datasetu, který jednoznačně popisuje entity, relace a jejich význam. Je možné využít již existující modely, které mohou obsahovat užitečné slovníky.
4. Vytvoření dat s identifikátory – aplikováním sémantického modelu z předchozího kroku budou data „linkable“ – každý objekt bude mít svůj vlastní identifikátor a je možné se takto na jednotlivé objekty odkazovat. Tento krok umožní interoperabilitu a znovupoužití dat.
5. Přiřazení licence – i když je informace o licenci součástí metadat, specifikuje se přiřazení licence jako samostatný krok pro zdůraznění důležitosti. Absence licence může znemožnit uživatelům data použít, i když byla data zamýšlena jako přístupná všem, a proto by se zvolení správné licence nemělo opomíjet.
6. Definování metadat pro dataset – pro splnění většiny FAIR bodů je potřeba, aby metadata byla důkladně popsána.
7. Nasazení zdroje dat – zpřístupnění či publikování FAIRifikovaných dat spolu s metadaty a licencí. GO FAIR popisuje, jak lze data FAIRifikovat a jak postupovat, ale aktivně se nepodílí na implementaci komponent potřebných ve FAIRifikačním procesu.

2.6.6 The FAIR Funder

The FAIR Funder je pilotní program, jehož cílem je rozšířit užití FAIR dat. Investorům usnadní proces vyžadování a kontroly FAIR dat a výzkumníkům usnadní jejich produkci. Zároveň je vysvětleno sedm fází pracovního procesu, popisujících přípravu šablon a metrik, sběr a uchování dat a kontrolu splnění požadavků. Text vychází z [9].

S poznáním, že vysoce kvalitní metadata jsou nezbytná k FAIRifikaci dat, založili členové iniciativy GO FAIR a *RDA* (nezisková organizace Research Data Alliance³) sérii *workshopů M4M*. Workshopy M4M, neboli Metadata for machines, podporovaly vytváření metadat pro stroje, tzv. machine-actionable metadata. Výsledkem M4M se stal pilotní program The FAIR Funder. The FAIR Funder představuje pracovní postup o sedmi fázích pro správu dat.

Po M4M počet mezinárodních účastníků, mezi nimi i dvě investorské agentury, spustil tento pilotní program. Program má ukázat, jak rychle lze uskutečnit reálné vyžadování FAIR výstupů investory. Program se zaměřuje na posouzení kvality data stewardship plánu a posouzení stupně FAIRifikace výstupů projektu. Na obě tato posouzení je nutná vysoká technická znalost a profesní zkušenosti, kterých je v současné době málo a obvykle se nenachází v investorských organizacích. Pilotní program umožní svým členům studovat mechanismy, přičemž agentury budou moci definovat očekávané parametry. Investoři mohou díky tomuto projektu reálně běžně vyžadovat machine-actionable metadata, zatímco pracovníci projektů mohou reálně a běžně požadavkům investorů vyhovět, a to bez nějakého specializovaného zaškolení či vyšší znalosti metadat.

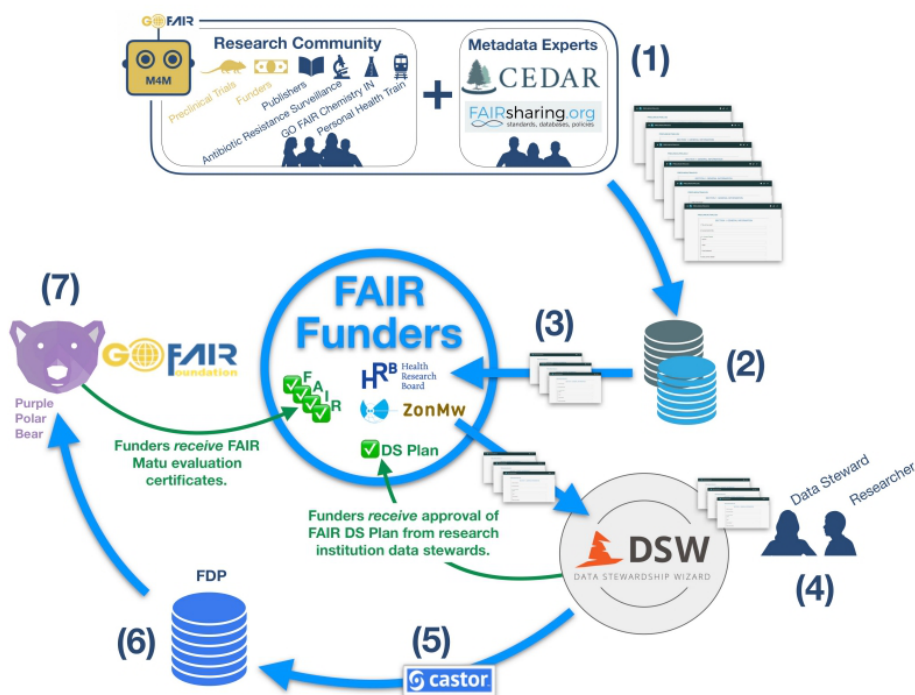
Pilotní program popisuje proces skládající se ze sedmi fází a jejich možné poskytovatele a řešení, čímž ukazuje proveditelnost svých cílů. Obecný kontext FAIR data stewardship obsahuje a používá machine-actionable metadata. Pilotní program pro tyto potřeby využívá vysoce konfigurovatelný nástroj pro data stewardship plánování, kterým je *Data Stewardship Wizard* (nebo také DS Wizard). Obecně se slovem *wizard* nazývá funkce softwaru nebo webového rozhraní, která funguje jako průvodce a dokáže vysvětlit konkrétní problematiku.

Data Stewardship Wizard byl zahájen společně infrastrukturami ELIXIR NL a ELIXIR CZ, jehož součástí je i ČVUT v Praze. Jedná se o chytrý nástroj na plánování správy dat pro FAIR Open Science. Open Science je hnutí, snažící se zpřístupnit vědecký výzkum všem vrstvám společnosti. DS Wizard je založen na hierarchickém machine-actionable znalostním modelu zohledňujícím osvědčené postupy, převážně formou otázek a odpovědí. Znalostní model DS Wizarda je volně přístupný a vysoce konfigurovatelný. Tento nástroj učí své uživatele standardy použití a znovupoužití machine-actionable metadat. Proces, kterým vznikají šablony pro machine-actionable metadata, může být

³<https://www.rd-alliance.org/>

2. ANALÝZA DOMÉNY PLÁNOVÁNÍ SPRÁVY DAT

jednoduše replikován pro možné budoucí potřeby v jakékoliv výzkumné oblasti.



Obrázek 2.1: Diagram popisující fáze projektu The FAIR Funder: K tvorbě a sdílení FAIR výstupů slouží pracovní proces skládající se ze sedmi fází. Ve středu diagramu jsou investoři (modrý kruh). Zelené šipky značí řešení současných hlavních problémů investorů s FAIR daty. Těmi jsou posouzení celkové kvality plánu a posouzení stupně FAIR výstupů projektu. Navrhovaný proces odebírá investorovi zodpovědnost za tato technická posouzení a předává ji vhodným účastníkům výzkumné komunity, kteří mají potřebné znalosti. Externí služby pro posouzení a vyhodnocování vyžadují kombinaci zaškolených expertů (data stewardů), obsahu (zaručeným FAIRsharingem), technologie (CEDAR, Castor EDC a DS Wizard), a nakonec důvěryhodné organizace třetích stran, mezi které patří GO FAIR a Purple Polar Bear.

1. Výzkumné komunity provozují M4M workshopy.

M4M workshopy jsou populární události, díky kterým mohou experti spolupracovat a společně tvořit. Pro potřeby pilotního programu tak mohou definovat metadatová schémata splňující FAIR, z nichž budou poté vytvořeny machine-actionable šablony, které bude moci využít širší komunita lidí.

2. Definované šablony pro metadata a FAIR metriky jsou uloženy v otevřených repozitářích.

Technické výstupy M4M workshopů jsou dohledatelné a sdílené, což přispívá k rychlému rozšíření FAIR metadat a metrik. Informace o standardech metadat budou vypsány na stránkách FAIRsharing⁴. FAIRsharing zároveň pracuje na vytvoření metadat standardů čitelných stroji (machine-readable), čímž bude možné zajistit, že je dataset popsán dle definovaných požadavků.

3. Investoři využívají machine-actionable šablony pro metadata a FAIR metriky v CEDAR repozitáři.

CEDAR (Center for Expanded Data Annotation and Retrieval) byl založen v roce 2014 za účelem zřízení ekosystému na vývoj, vyhodnocování a využití biomedicínských dat. Soustředí se na metadatové šablony a FAIR data. Informace v repozitáři CEDAR využijí investoři k vytvoření nových požadavků a začlení je do znalostního modelu Data Stewardship Wizard. Pro výzkumníky vypadají CEDAR formuláře podobně jako jim známé webové formuláře při klasických žádostech. Na rozdíl od nich ale poskytují CEDAR formuláře s machine-actionable metadaty jisté výhody, například mohou díky kontrolovaným slovníkům pomáhat s vyplňováním formulářů nabízením souvisejících termínů a dokončováním rozepsaných slov. FAIRsharing poskytuje definované standardy metadat a splnění FAIR principů je hlášeno na konci projektu automatizovanými službami.

4. Výzkumníci a data stewardi vytváří plány pro investory obsahující požadovaná zadaná metadata.

Plány budou vytvářeny pomocí CEDAR formulářů obsažených přes odkazy v Data Stewardship Wizard a FAIRsharing. I když již existuje mnoho nástrojů na tvorbu plánů správy dat, součástí tohoto pilotního programu je právě i nasazení dle potřeb vyvíjeného nástroje Data Stewardship Wizard. Investoři nebudou plány aktivně hodnotit, to bude úkolem data stewardů místních výzkumných organizací. Ti budou mít za úkol ověřit, že plán je vyhovující. Výstupem DS Wizard bude machine-actionable, ale zároveň lidem srozumitelný, FAIR plán správy dat.

5. Výzkumníci spustí projekty a začnou se sběrem dat.

V průběhu projektů bude používán CEDAR a Castor EDC, což je elektronický systém na tzv. case-report form (zpráva o výzkumu nebo také případová studie, týkající se dat). CEDAR, Castor EDC a DS Wizard budou usilovat o interoperabilitu napříč platformami.

⁴<https://fairsharing.org>

6. Machine-actionable data a metadata budou uchována v repozitářích, které automaticky kontrolují splnění FAIR metrik.

Na závěr projektu budou data a metadata uložena do jednoho či více repozitářů podporujících FAIR data. Součástí pilotního programu bude rozšíření repozitářů o funkce, které budou automaticky spouštět služby třetích stran kontrolující splnění FAIR metrik.

7. Služby důvěryhodných třetích stran vyhodnocující FAIR metriky.

Díky službám třetích stran monitorujících splnění FAIR metrik není potřeba, aby investoři měli specializaci a znalosti potřebné ke kontrole. Mohou tak FAIR data vyžadovat a mít jistotu, že jsou v souladu s jejich požadavky. Služba zároveň poskytne přehled obsahující informace o splnění jednotlivých principů FAIR.

Téměř připravené k užití jsou již technické komponenty pilotního programu, různé nástroje a služby, FAIR repozitáře a FAIR metriky. Náročnější jsou sociální komponenty, jako změny v organizacích a motivace těchto organizací ke změně. K tomu ale pozitivně přispívají M4M workshopy a iniciativy pro vývoj FAIR certifikací.

2.7 Data stewardship

Tato sekce pojednává o procesu data stewardship. Je zde vysvětleno, k čemu slouží, proč jeho důležitost stále roste a jaké může mít přínosy pro organizaci, jež tento postoj zastává a má svého data stewarda, tedy správce dat. Součástí je i sekce popisující projekt, který se snaží vyřešit problém s nedostatkem data stewardů v určité oblasti výzkumu.

V [4] je definován pojem data stewardship jako:

„Proces a postoj, který vede k zodpovědnému nakládání s vlastními i cizími daty během, po, i před vědeckou prací a obecně cyklem práce s daty.“

Prof. Barend Mons, specialista na FAIR data

Jedním z důvodů, proč jsou data stewardi oproti dřívějšku v současné době žádanější, je nárůst množství i typů dat. Pokud v organizaci nejsou přesně stanoveny podmínky zacházení s daty, lehce se stane, že většina dat bude nepoužitelná. Budou na různých platformách, nebude sjednocené názvosloví, identifikátory nebudou jednotné nebo nebudou vůbec použité, data mohou být v různých databázích... V datech tak není přehled, není možné komunikovat napříč platformami, nelze data dobře analyzovat a vyvozovat z nich závěry. Každé oddělení tak s největší pravděpodobností použije k potřebné analýze jen svá vlastní data, která mohou být neúplná a nepřesná. Prací data stewardů je právě starat se o to, aby k těmto situacím nedošlo. Data by měla být detailně a sjednoceně popsána, propojena mezi sebou a s odpovídajícími metadaty. Díky tomu bude možné s nimi dobře pracovat, data by proto měla být stále aktuální a dostupná. S dobře spravovanými daty se zvyšuje efektivita práce jednotlivých zaměstnanců, kteří je začnou více a více využívat, a tak se zvyšuje i efektivita celé společnosti. S dobře spravovanými daty lze také předejít některým chybám a dosáhnout lepšího stanovení priorit. [10]

Data stewardi dohlížejí i na to, aby data neporušovala žádná práva a normy. Tento aspekt získal zvýšenou pozornost zejména se zavedením GDPR. V souvislosti s ním se otázky související s daty zbytečně neodkládají tak jako tomu bylo dříve, řeší se časné a s dostatečným finančním zajištěním, v některých případech již před spuštěním projektu. Společnosti si dávají pozor, aby veškerá data byla v pořádku a nemusely se tak řešit soudní spory či pokuty. Součástí práce data stewardů je tedy zároveň zajistit, aby data byla v souladu s regionálními a mezinárodními právy a zákony a také s podmínkami dané organizace. [10, 11]

V jiných evropských zemích je pozice data stewarda na plný úvazek již běžnou záležitostí. V České republice zatím není známo, že by nějaká organizace měla data stewarda na plný úvazek, ale je to jen otázka času (informace datována k 4. 6. 2019). [4]

2.7.1 FAIR data stewardship pro life science

Tento text vychází z [12] a pojednává o projektu, jehož cílem je řešení nedostatku data stewardů v oblasti *life science*, což je odvětví vědy zabývající se studii organismů – mikroorganismů, rostlin, zvířat i člověka.

Pro správu FAIR dat jsou zkušení data stewardi nezbytní. Těch ale může být na trhu nedostatek nejen obecně, ale i lokálně, nebo nemají zkušenosti v dané oblasti. Jeden z takovýchto problémů začalo řešit úložiště Zenodo. Název projektu je „Towards a community-endorsed data steward profession description for life science research“, což se dá volně přeložit jako „Vstřícní komunitou podporovanému popisu profese data stewarda pro výzkum life science“. Cílem projektu je zvýšit profesionalitu data stewardů v oblasti life science s důrazem na implementaci FAIR principů. Ze získaných znalostí a dovedností bude vyvíjen školicí materiál, doplněn o návrh eLearningového modulu.

Některá centra již začala sdílet své osvědčené postupy pro data stewardship, ale například v oblasti life science takovéto informace nejsou dostupné, nejsou popsány vyžadované znalosti a dovednosti. Vhodný přístup s potřebným zaměřením a školení v tomto oboru mají nedostatky na národní úrovni. Tím se limituje možnost reálného uplatnění data stewardů v life science výzkumech, ale zároveň se tím velmi komplikuje nábor data stewardů a jejich zaškolení pro uplatnění v institucích a projektech. Projekt cílí na zvýšení profesionality data stewardů v oblasti life science v Nizozemsku, a to jak zvýšení kvality, tak kvantity. Kvalita se zajistí dalším školením již existujících data stewardů a kvantita zaškolením nových.

Součástí projektu bude popis funkce data stewarda, obzvláště v oblasti life science. Z toho budou vyvozeny související povinnosti, znalostní báze a dovednosti. Ke školení bude sloužit návod s doporučenou vzdělávací osnovou odkazující na existující kurzy, popisující potřebné vzdělání a znalosti data stewarda. Zároveň budou vyvíjeny zmíněné modulární výukové materiály, s jejichž pomocí se poté uskuteční kurzy pro zaškolení data stewardů. Tyto kurzy budou zároveň i otestováním materiálů a zaškolení samotného v praxi. Na závěr bude navrhnout eLearningový modul připraven na plnou implementaci, která by proběhla v další fázi po získání potřebného financování.

Funkční popis a definice pro life science data stewardy společně s dokumentací vzdělávací osnovy a školicích materiálů budou veřejně přístupné. Přístupné budou na několika internetových stránkách, například na výukovém portálu ELIXIR TeSS a na komunitě Zenodo.

2.8 DMP Common Standards Working Group

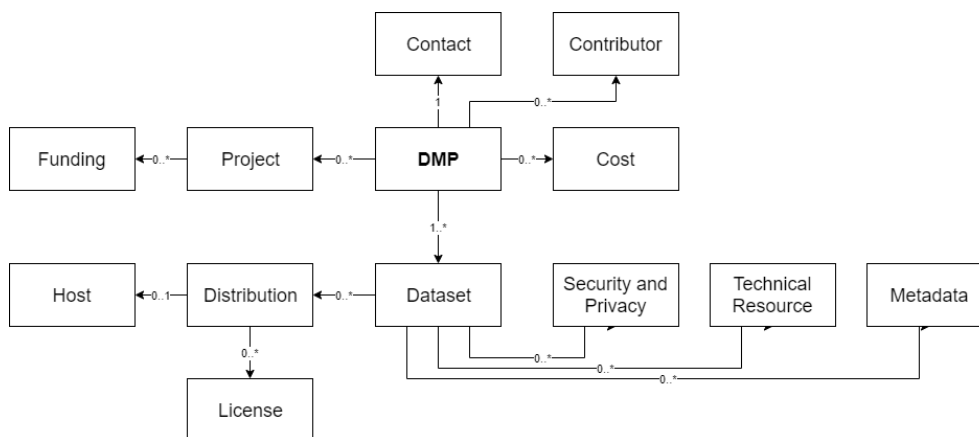
Následující sekce dle [6] se zabývá pracovní skupinou řešící problematiku související s absencí zavedených společných standardů pro plány správy dat. Jsou popsány cíle, kterých chce skupina dosáhnout a i způsoby, jakými tak chce učinit. Taktéž je vysvětleno, proč je absence těchto standardů překážkou pro další rozvoj plánů správy dat.

Společnost již došla k závěru, že machine-actionable plány jsou potřebné pro další rozvoj plánů a na toto téma se pořádají četné workshopy a diskuze. Je žádané, aby nástroje a systémy byly navzájem propojené a spolupracovaly. Vzhledem k vysokému stále se zvyšujícímu počtu existujících nástrojů pro tvorbu plánů správy dat se tento požadavek ale stává problematičtější. Pro jeho splnění je zapotřebí větší koordinace a stanovení standardů. A právě workshop pro machine-actionable plány správy dat organizovaný Digital Curation Center (DCC) a University of California Curation Center odhalil, že v současném stavu jsou pro další postup nezbytné společné standardy pro machine-actionable plány. Výstupem diskuzí a workshopů je vytváření specifických pracovních skupin zabývajících se identifikovanými problémy. Jednou z těchto skupin je i zde popisovaná DMP Common Standards WG, která řeší zmíněný problém s vysokou prioritou, a to společné standardy.

Skupina bude otevřená všem, bude spolupracovat s dalšími skupinami zabývajících se podobnými tématy, a jejími členy budou všichni účastníci z celého spektra správy dat. Vezmou se tak v úvahu požadavky každé skupiny účastníků. Spolupráce bude probíhat převážně online, ale skupina bude využívat událostí, jako jsou konference či workshopy, které členové navštěvují, k osobním setkáním. Učiněná rozhodnutí skupiny budou dostupná na internetové stránce RDA, stejně jako důležitá zjištění. Činnosti mohou být publikovány na online fórech pro podporu zapojení širšího okruhu lidí, pokud to povaha činností dovolí. Někteří z předních reprezentantů různých skupin z oblasti plánování správy dat jsou již členy této skupiny – Data Stewardship Wizard, DMPonline a DMPTool. Všichni reprezentanti odsouhlasili zvažování implementace standardů této pracovní skupiny do svých nástrojů, někteří se již aktivně účastní. Skupina se však i nadále bude snažit vyhledávat další reprezentanty z řad výzkumných komunit, aby byly její výsledky široce přijaty.

Hlavní zaměření skupiny je vývoj společného informačního modelu a specifikování mechanismů které činí plány machine-actionable. Výsledkem bude lepší schopnost systémů vzájemně spolupracovat a možnost kontrolovat a sdílet informace obsažené v plánech. Bude možné ověřit, že PID identifikátory skutečně odkazují na existující dataset a že byla specifikována licence. Společné informační modely nejsou zamýšlené jako šablony ani dotazníky, pouze by měly poskytovat opětovně použitelné způsoby, jak reprezentovat machine-actionable informace. Skupina se skrze společné informační modely snaží zrealizovat vizi, kdy plány správy dat budou plně začleněny do systémů i pracovních postupů a s tímto vědomím budou vytvářeny a udržovány. Modulární

návrh umožní přizpůsobení a rozšíření dle současných standardů a slovníků, aby co nejlépe vyhovoval potřebám různých komunit. Nástroje a systémy budou schopné z plánu číst a zapisovat do něj.



Obrázek 2.2: Model vytvořený DMP Common Standards Working Group: Aplikační profil pro metadata, poskytující základní interoperabilitu systémům pracujících s machine-actionable plány správy dat.

Profil pokrývá širokou škálu případů užití, neobsahuje specifické požadavky. Informace popisující tento profil vychází z [13].

Popis entit:

- DMP – poskytuje informace o DMP, jedná se o nejdůležitější část diagramu a většina informací v tomto poli je k zadání povinná.
- Project – popisuje projekt, který může být mnoha typů, kromě financovaných projektů i interní projekt nebo disertační práce.
- Funding – popisuje detaily financování projektu.
- Contact – specifikuje skupinu která poskytuje informace pro plán, může se jednat o jednotlivce i organizaci. Nemusí to být tvůrce plánu.
- Contributor – obsahuje všechny skupiny podílející se na vytvoření a správě plánu.
- Cost – obsahuje seznam nákladů v souvislosti se správou dat.
- Dataset – dataset lze chápat jako logickou entitu popisující data a informace o nich.
- Distribution – popisuje určitou instanci datasetu a jeho vlastnosti, jako je formát a velikost.
- License – uvádí licenci, pod kterou budou data dostupná.

- Host – popisuje informace o systému, kde jsou data uložena. Může popisovat repozitář, ale také počítače, které jsou používány v průběhu projektu pro zpracovávání a dočasné uložení dat.
- Security and Privacy – popisuje specifické požadavky týkající se zabezpečení a ochrany údajů specifického datasetu. Například, že data nejsou anonymizována.
- Technical Resource – specifikuje potřebné/použité vybavení a nástroje nutné k vytvoření či zpracování dat.
- Metadata – poskytuje ukazatel na standard metadat použitý pro popsání dat. Neobsahuje metadata datasetů.

Již předběžné návrhy se budou kvůli získávání zpětné vazby vyhodnocovat s komunitou. Existující plány se vyjádří pomocí vyvíjeného modelu a budou se zkoumat možnosti podpory úpravy machine-actionable plánů různými nástroji zahrnutými v procesu správy dat. Součástí je i kontrola správně zaznamenaných informací o původu dat a verzování. Na závěr se vytvoří prototypy, na kterých se detailně prozkoumají možnosti integrace, automatizované validace, a spouštění činností a upozornění napříč systémy.

Během práce se budou mezi výzkumnými komunitami hledat kandidáti pro pilotní studie a předání získaných znalostí. Ke konci práce skupiny budou spuštěny pilotní projekty s modely, které budou přizpůsobeny potřebám a požadavkům jednotlivých komunit. V pilotních studiích tak bude díky modelu pro integraci systémů demonstrováno, jak machine-actionable plány správy dat fungují. Výstupy této pracovní skupiny by měly napomoci ke zvýšení kvality dat a jejich reprodukovatelnosti a zároveň ke snížení náročnosti administrativní práce. Výsledky práce, obsahující pokyny k použití modelu a dokumentaci, budou prezentovány na konferencích a publikovány na sociálních médiích.

Výstupem práce této skupiny je také další krok k tomu, aby plány správy dat splňovaly iniciativu FAIR, konkrétně principy interoperability a znovupoužitelnosti. [3]

Analýza nástrojů

Tato kapitola se věnuje analýze nástrojů, převážně jejich průchody při vytváření plánu. Jsou popsány funkce nástrojů, jejich rozdělení, a na úvod informace jako jsou okolnosti jejich vzniku nebo zaměření. Z této kapitoly vychází následující kapitola 4, která se věnuje hodnocení nástrojů dle zadaných kritérií.

Jak bylo zmíněno, k tvorbě plánu správy dat lze využít online nástroje. Existuje mnoho celosvětově dostupných nástrojů pro správu dat, ale prvními nástroji byly DMPOnline a DMPTool, které byly spuštěny ve Velké Británii a Spojených státech amerických již v roce 2010 a 2011. V posledních letech bylo vydáno mnoho dalších nástrojů. Většina poskytuje možnosti pro vytvoření a sdílení plánu, ale objevují se i odlišné preference a jiné pohledy na činnost nástrojů. Některé se zaměřují na omezení otevřeného textu formou uzavřených otázek, což umožní implementovat služby, které na základě slovníků a znalostníchází budou moci uživatelům radit a pomáhat. Jiné nástroje se více zaměřují na to, aby uživatele naučily plán a nástroj používat a obeznámily ho s vyplývajícími výhodami. Některé nástroje se zaměřily na konkrétní problém a zabývají se tak požadavky konkrétního investora. [3]

V souvislosti s rozsáhlejšími projekty probíhajícími na více institucích či projekty financovanými více investory se objevuje nutnost sjednocení plánů. Každý může mít trochu odlišné požadavky, a je nepraktické a obtížné vytvářet plán správy dat zvlášť pro každou instituci a investora. A ještě obtížnější je všechny tyto plány spravovat. Na řešení otázky sjednocení plánů se zaměřil projekt DMPRoadmap. Ten vytvořil nástroj, který umožňuje vytvářet a spravovat různé verze plánů a do již hotových šablon umožňuje přidat další požadavky. Není tak nutné spravovat pro jeden projekt několik plánů. DMPRoadmap je open-source nástroj, který je již využíván v DMPOnline, DMPTool a dalších. Sjednocení datových politik a standardů zaručí interoperabilitu nástrojů, nikoliv však dosažení používání jediného upřednostněného nástroje a šablony. Tento scénář je vysoce nepravděpodobný a zároveň nechtěný. Plán správy dat by měl být vytvářen a optimalizován pro kontext konkrétního výzkumu, jelikož každý uživatel či oblast výzkumu může mít od-

lišné požadavky a potřeby. [3]

Nástroje DMPOnline, DMPTool, QUT DMP a DMP OPIDoR (a další, kterým se ale tato práce nevěnuje) jsou takzvané „forky“ z DMPRoadmap. Jedná se o kopie repozitáře, které mohou být libovolně upravované. Jelikož všechny patří do stejné rodiny nástrojů založené na DMPRoadmap, jsou si velmi podobné a jejich popis bude zaměřen spíše na odlišnosti. Detailně popsán je první nástroj, DMPOnline (Velká Británie), s ním je porovnán DMPTool (Spojené státy americké) a na závěr krátké porovnání DMPOnline s australským QUT DMP a francouzským DMP OPIDoR. Zmiňovaná porovnání jsou pro představu o funkcích, odlišnostech a společných vlastnostech těchto příbuzných nástrojů. Hodnocení a porovnání dle stanovených kritérií jsou popsány v následující kapitole.

Pokud nástroj nenabízí více šablon a otázky jsou ve většině případů stejné, práce se jím zabývá více. Jsou vyjmenovány některé otázky, popsány formáty vstupu. Pokud ale nástroj nabízí větší přizpůsobitelnost, a množství a typ otázek v kategoriích by se lišil v závislosti na zvolené šabloně či modelu, tak se práce nezabývá konkrétními průchody dle šablon, kterých mohou být desítky. Průchod je popsán dle výchozí šablony s případnými komentáři, co může být odlišné, pokud by pro plán byla vybrána jiná šablona než výchozí.

Seznam analyzovaných nástrojů vychází z konzultací s vedoucím:

- DMPOnline
- DMPTool
- Data Stewardship Wizard
- DataWiz
- easyDMP
- ezDMP
- QUT DMP
- DMP OPIDoR
- CEDAR

Poslední z nástrojů, CEDAR, nebude v práci hodnocen a nebude tedy obsažen v následující kapitole Zhodnocení. CEDAR totiž není nástrojem na vytváření plánů správy dat jako ostatní nástroje, je určen primárně na metadata. Umožňuje vytvářet formuláře a má dobrou podporu machine-actionability. I z důvodu zmínění v předchozí kapitole bude popsán spolu s ostatními nástroji, ale kvůli své odlišnosti nebude hodnocen a srovnáván.

Všechny nástroje jsou k použití zdarma. Poznátky o nástrojích v této i následující kapitole vychází ze zdrojů uvedených v této kapitole u jednotlivých nástrojů. Analýza je doplněna o snímky z těchto nástrojů. Jelikož smyslem

snímků je více přiblížit nástroje, ukázat rozložení prvků a vizuálně doplnit psaný popis, nebyl kladen důraz na dobrou čitelnost všech textů a pro odpovědi byl použit vygenerovaný text. Snímky se tedy nezabývají zněním otázek a jejich vhodným zodpovězením, ale celkovou podobou nástroje a rozvržením prvků.

3.1 DMPOnline

Vydání: 2010

Provozovatel: Digital Curation Center (DCC)

Funkčnosti: Vytváření, editování, sdílení a publikování plánů správy dat. Vícejazyčný, uzpůsobený pro potřeby výzkumu, institucí a investorů.

Zdroj: Nástroj [14], GitHub [15], publikace [16]

DMPOnline obsahuje mnoho šablon plánů založené na konkrétních specifických požadavcích investorů a institucí. Použití DMPOnline je tak některými investory povinně vyžadováno, jiní to pouze vnímají jako užitečnou alternativu. Uživatelům jsou na začátku položeny tři otázky sloužící pro vybrání vhodné šablony. S interpretací a zodpovídáním otázek pomáhají sami investoři a univerzity. Šablony jsou spravovány DCC, je ale přesto doporučováno, aby výzkumníci vždy konzultovali požadavky přímo s investory.

3.1.1 Vytváření plánu

Pro vytvoření plánu je nejdříve zapotřebí vyplnit název projektu, hlavní výzkumnou organizace a hlavní investorskou organizace.

* What research project are you planning?

Zkušební projekt

* Select the primary research organisation

Swedish School of Sport and Health Sciences

* Select the primary funding organisation

ca|

Biotechnology and Biological Sciences Research Council (BBSRC)

Cancer Research UK (CRUK)

Engineering and Physical Sciences Research Council (EPSRC)

Medical Research Council (MRC)

Obrázek 3.1: Úvodní otázky nástroje DMPOnline

Ke každé z otázek je přidružen checkbox. Pokud plán slouží jako žádost o financování, je nutné uvést název projektu přesně. V případě, že se nejedná o reálný projekt, ale uživatel si chce jen něco vyzkoušet či vytvořit plán pro vzdělávací účely, označí checkbox pro tzv. „mock“ projekt. Výběr organizace v dalších dvou otázkách probíhá pomocí výběru ze seznamu. Pokud potřebná organizace není v seznamu nebo s plánem není spojena žádná organizace a otázka by tak měla zůstat nezodpovězená, lze místo odpovědi označit checkbox sloužící právě pro tyto dva případy. Dle zvoleného investora se může zobrazit ještě čtvrtá otázka nabízející šablony. Po zodpovězení těchto otázek je vytvořen plán. A to buď na základě výchozí šablony, nebo na základě šablony zvolené dle investora.

Dále je popsán průchod plánem dle výchozí šablony, se zmíněním odlišností oproti investorským šablonám. Plán a jeho tvorba jsou rozděleny do následujících sekcí:

1. Detaily projektu

V první sekci je možné vybrat až šest organizací poskytujících instrukce a rady pro psaní plánu. V následující sekci, ve které se vyplňují otázky, budou zobrazované instrukce k jednotlivým otázkám od vybraných organizací. Standardně je zvolen provozovatel DMPOnline, Digital Curation Center. Jedinou povinnou otázkou je zde název projektu, který je již vyplněn z předchozí stránky. Dále lze vyplnit investora, číslo grantu, abstrakt projektu a kontaktní údaje hlavního výzkumníka. Všechny odpovědi jsou formou otevřeného textu a u některých otázek jsou stručné vysvětlující texty.

2. Přehled projektu

V přehledu projektu není vyžadována žádná akce uživatele, nejsou zde otázky. Jedná se pouze o přípravu k dalšímu kroku, kterým je napsání plánu. Je zde popsáno, jaká šablona je vybrána, jaké specifikace by měl plán dle této šablony splňovat a jaké otázky by měl zodpovídat. Tyto otázky jsou podrobně rozepsány v další sekci, ve které se na ně bude odpovídat.

3. Napsání plánu

Tato sekce obsahuje otázky rozdělené do několika témat. Jednotlivé otázky i celá témata se liší v závislosti na vybrané šabloně, ale ve většině případů jsou podobné otázkám ve výchozí šabloně. Investorské šablony se stejnými či podobnými otázkami obvykle zabývají také, pouze s odlišnými prioritami a jejich formulace se liší. Odpovědi jsou psány formou otevřeného formátovaného textu s možností formátování – lze tak jako odpověď zadat tabulku či číslovaný seznam. Po zodpovězení otázky je u otázky uloženo, kdo ji kdy zodpověděl. Ke každé z otázek lze přidat i komentář pro spolupracovníky.

3. ANALÝZA NÁSTROJŮ

Pokud si uživatel zvolil organizaci DCC pro pomoc s vyplněním plánu, ke každé z otázek je poskytnut detailní návod s vysvětlením. Jaké otázky při psaní odpovědi zvážit, jak je vhodně a jasně zodpovědět a co znamenají zmíněné pojmy.

Například, první otázka ve výchozí šabloně zní "Jaká data budou shromážděna nebo vytvořena?".

DCC zde navrhuje popsat typ, formát a objem dat. K formátu i objemu dat jsou přítomny další nápovědy, obvykle formou delšího popisu. V nápovědě formátu dat je uvedena otázka, zda budou data ve formátu prostého textu, tedy txt, nebo jiné, a následují příklady dalších formátů jako je csv nebo tif. Dále návod obsahuje vysvětlení, proč je vhodné používat otevřené formáty, a že součástí odpovědi by mělo být i vysvětlení a obhájení vybraného formátu, z jakého důvodu byl zvolen.

Pokud se nejedná o výchozí šablonu, k otázkám jsou přidány ještě rady a návody související s konkrétní vybranou šablonou. Tyto návody nemusí být přítomné u každé otázky. Pokud by si uživatel nebyl u otázky jistý, může stále použít návod od DCC, který je u každé z otázek také připojen a lze mezi nápovědami libovolně přepínat v rámci otázky.

The screenshot displays the DMPOnline tool interface. At the top, there are two orange progress bars: "Data Collection (0 / 2)" and "Documentation and Metadata (1 / 1)". Below these is a question: "What documentation and metadata will accompany the data?". The question is contained within a rich text editor with a toolbar showing bold, italic, list, link, and table icons. The editor contains the text "Lorem ipsum dolor sit amet, consectetur adipiscing elit." followed by a numbered list: "1. Fusce aliquam vestibulum ipsum." and "2. Aenean id metus id velit ullamcorper pulvinar.". Below the editor is an orange "Save" button. At the bottom, a blue bar indicates the question was "Answered 23 hours ago by machacova.terka@gmail.com".

Obrázek 3.2: Nástroj DMPOnline.

4. Sdílení

Ve Sdílení lze spravovat spolupracovníky, přidat nové, a nastavit viditelnost plánu:

- Soukromý plán vidí jen tvůrce plánu, přidání spolupracovníci a administrátoři.
- Viditelný pro organizaci.
- Veřejný plán vidí kdokoliv.

Viditelnost pro veřejnost a organizaci je určena pro dokončené plány, a pro nastavení těchto možností je tak nutné, aby bylo alespoň 50 % otázek zodpovězeno.

Spolupracovníky lze přidat pozváním přes email. Pokud již nástroj DMPOnline používají, uvidí daný plán v menu mezi svými plány. Pokud jej nepoužívají, obdrží email s pozvánkou. Při přidání je povinné vybrat práva spolupracovníka vybráním jedné z následujících možností (každá možnost obsahuje i právo přístupu a čtení):

- Spoluvlastník, má admin práva, může tedy odpovídat na otázky, psát komentáře a přidávat spolupracovníky.
- Editor, může odpovídat na otázky a psát komentáře.
- Jen pro čtení, může psát komentáře.

5. Stažení

V možnostech stažení plánu si lze zvolit formát (pdf, docx, html...), a pokud byl zvolen formát pdf tak i možnosti formátování. Zde lze vybrat velikosti jednotlivých okrajů dokumentu a font a velikost písma. Zároveň se uživatel rozhodne, zda chce mít ve stahovaném souboru i texty otázek a nadpisy sekcí a zda tam mají být i nezodpovězené otázky. První stránku lze vyhradit shrnutí, které obsahuje jméno tvůrce a název šablony. V případě jiné než výchozí šablony si uživatel zvolí, zda chce mít v plánu obsaženy i doplňkové sekce, které investorem nejsou vyžadovány.

3.1.2 Reference

Kromě přehledu plánů a vytvoření nového plánu nástroj ještě nabízí záložku Reference, obsahující veřejné plány a požadavky investorů. Veřejné plány jsou vytvořené službou DMPOnline. Z přehledu lze vyčíst šablonu, organizaci a autora pro každý plán a tento plán lze stáhnout. Stránka s požadavky investorů vypadá velice podobně, pouze místo plánů popisuje jednotlivé šablony a umožňuje z této stránky plán dle dané šablony začít vytvářet. Poskytuje též hotové ukázkové plány dle těchto šablon.

3. ANALÝZA NÁSTROJŮ

Template Name	Download	Organisation Name	Last Updated	Funder Links	Create a new plan	Sample Plans (if available)
AHRC Data Management Plan	 	Arts and Humanities Research Council (AHRC)	14-08-2019	Data Management Plan guidance Data Management Points	 	Religious studies DMP from Bristol Language studies DMP from Glasgow UK and German International Criminal Co-operation example from Robert Gordon University
BBSRC Template	 	Biotechnology and Biological Sciences Research Council (BBSRC)	16-05-2019	BBSRC policy on DMPs	 	TRDF Grant DMP from Cambridge Drosophila Genetics DMP from Glasgow
CESSDA list of Data Management Questions	 	Data Archiving and Networked Services	11-03-2020		 	
Data Management Plan NWO (January 2020)	 	Netherlands Organisation for Scientific Research (NWO)	07-02-2020	NWO Data management protocol NWO	 	
Datamanagement ZonMw-template 2016-2018	 	ZonMw (Nederlands)	16-05-2019	ZonMw FAIR data management (2016-2018)	 	Ischemic heart disease example
Data management ZonMw-template 2019	 	ZonMw (Nederlands)	25-04-2019	ZonMw FAIR data management (2019)	 	
DCC Template	 	Digital Curation Centre	18-05-2018		 	
EPSRC Data Management Plan	 	Engineering and Physical Sciences Research Council (EPSRC)	16-05-2019	Policy framework on research data	 	Synthetic chemistry example from Glasgow

Obrázek 3.3: Přehled šablon investorů v DMPOnline

3.1.3 Technická stránka a business model

Jelikož jsou DMPOnline, DMPTool, QUT DMP a DMP OPIDoR forky DMPRoadmap, mají mnoho společného. Jsou open-source a GitHub každého z nich je přístupný z [15], což je seznam forků DMPRoadmap. DMPRoadmap je Ruby on Rails aplikace, což je framework pro vývoj webových aplikací napojených na databázi. Je vyžadováno MySQL nebo PostgreSQL.

Od roku 2018 je business modelem DMPOnline tzv. freemium. Všechny výše popsané funkce jsou přístupné zdarma. Pokročilejší funkce jsou dostupné až po zaplacení předplatného, na rok nebo na tři roky. To umožní přidávat nové šablony, upravovat stávající, přiřadit jim fáze pro vyplnění a přidávat návody. Je také možné přidat vlastní logo a barvy. Cena je vypočítávána dle velikosti instituce a intenzitě výzkumu.

3.2 DMPTool

Vydání: 2011

Provozovatel: California Digital Library (CDL)

Funkčnosti: Vytváření, editování, sdílení a publikování plánů správy dat. Vícejazyčný, uzpůsobený pro potřeby výzkumu, institucí a investorů.

Zdroj: Nástroj [17], GitHub [15]

DMPTool poskytuje proklikávací wizard pro vytváření plánů správy dat dle požadavků a potřeb investorů. Obsahuje přímé odkazy na stránky investorů a různé typy a nápovědy.

Původní nástroj DMPTool začal v roce 2011 s osmi partnerskými institucemi jako reakce na poptávané požadavky investorských agentur. Cílem bylo sjednotit odborné znalosti a snížit náklady potřebné na záležitosti správy dat. Mezi původních osm institucí patří i již zmiňovaný Digital Curation Center (DCC-UK). První verze nástroje sklídila velký úspěch a po dalším financování byla v roce 2014 vydána druhá verze.

3.2.1 Vytváření plánu

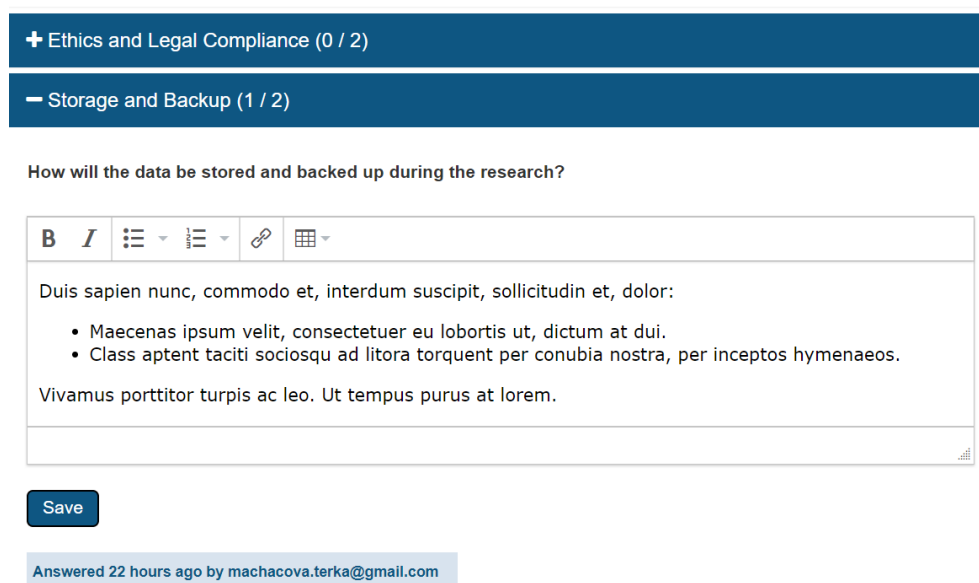
Nástroj je velice podobný nástroji DMPOnline. Počáteční stránka s otázkami je stejná, pouze jsou v odpovědích nabízené jiné organizace a celý nástroj je laděn do modré barvy místo oranžové. Tvorba plánu se skládá ze stejných sekcí – Detaily projektu, Přehled projektu, Napsání plánu, Sdílení, Stažení.

Taktéž stránka Detaily projektu je téměř totožná, pouze zde chybí otázka na vyplnění identifikátoru ORCID, a místo návodu od DCC jsou zde návody od DMPTool. Zatímco v DMPOnline je ale možné si zvolit návody od DMPTool, v tomto nástroji si není možné zvolit nápovědy od DCC.

Stránka Přehledu projektu je též stejná. Pokud nebyla vybrána žádná určitá šablona a plán byl vygenerován dle výchozí šablony, pak budou otázky v následující části nástroje stejné jako u výchozí šablony DMPOnline a i popisy pro nápovědy budou velmi podobné. Oba nástroje totiž využívají shodnou výchozí šablonu od DCC. V nástroji DMPTool tak budou nápovědy k otázkám od DMPTool (pokud uživatel výběr nezrušil či nevybral něco jiného) a od organizace dle investora. V případě výchozí šablony je návod poskytován právě organizací DCC. Rozdělení popisu otázek v nástroji je poněkud odlišné od DMPOnline a nejsou zde všechny texty, ale poskytované rady jsou ve většině případů shodné. Stejně jako u DMPOnline, i zde jsou informace o zodpovězení otázky, komentáře, a odpovědi formou otevřeného textu, který lze formátovat.

Stránky Sdílení a Stažení jsou obsahově také stejné jako u DMPOnline. Záložka Reference, obsahující veřejné plány a požadavky investorů která byla u DMPOnline, zde není, ale podzáložky jsou umístěny v pravém horním rohu pod záložkou „Learn“, spolu s dalšími informacemi.

3. ANALÝZA NÁSTROJŮ



The screenshot shows a web interface for DMPTool. At the top, there are two navigation tabs: '+ Ethics and Legal Compliance (0 / 2)' and '- Storage and Backup (1 / 2)'. Below the tabs, a question is displayed: 'How will the data be stored and backed up during the research?'. The question is followed by a rich text editor with a toolbar containing icons for bold (B), italic (I), bulleted list, numbered list, link, and table. The text in the editor reads: 'Duis sapien nunc, commodo et, interdum suscipit, sollicitudin et, dolor:' followed by a bulleted list with two items: 'Maecenas ipsum velit, consectetur eu lobortis ut, dictum at dui.' and 'Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos.' Below the list is the text 'Vivamus porttitor turpis ac leo. Ut tempus purus at lorem.' At the bottom of the editor is a 'Save' button. Below the editor, a light blue box contains the text 'Answered 22 hours ago by machacova.terka@gmail.com'.

Obrázek 3.4: Nástroj DMPTool

3.2.2 Technická stránka a business model

Stejně jako ostatní z této rodiny nástrojů je DMPTool open-source a GitHub je přístupný ze seznamu forků z [15]. DMPRoadmap je Ruby on Rails aplikace, což je framework pro vývoj webových aplikací napojených na databázi. Je vyžadováno MySQL nebo PostgreSQL.

DMPTool k použití se všemi funkcionalitami zcela zdarma, jedná se o komunitou podporovaný nástroj.

3.3 Data Stewardship Wizard

Vydání: 2015

Provozovatel: Dutch Techcentre for Life Sciences (ELIXIR NL) a ČVUT v Praze (ELIXIR CZ)

Funkčnosti: Vytváření, editování, sdílení a publikování dotazníků a znalostních modelů. Hierarchické dotazníky s uzavřenými otázkami, konfigurovatelnost znalostního modelu, vyhodnocení FAIR metrik, sestavení plánu správy dat.

Zdroj: Nástroj [18], GitHub [19]

Data Stewardship Wizard je nástroj vyvíjený jako součást pilotního programu The FAIR Funder. Byl zahájen společně infrastrukturami ELIXIR NL a ELIXIR CZ, jehož součástí je i ČVUT v Praze. Jedná se o chytrý nástroj na plánování správy dat pro FAIR Open Science. DS Wizard je založen na hierarchickém machine-actionable znalostním modelu zohledňujícím osvědčené postupy, převážně formou otázek a odpovědí. Znalostní model DS Wizarda je volně přístupný a vysoce konfigurovatelný. Tento nástroj učí své uživatele standardy použití a znovupoužití machine-actionable metadat. Proces, kterým vznikají šablony pro machine-actionable metadata, může být jednoduše replikován pro možné budoucí potřeby v jakékoliv výzkumné oblasti.

Jelikož je tento nástroj rozsáhlejší a má mnoho funkcí a možností, je popsán nejpodrobněji.

Nástroj nabízí několik uživatelských rolí, dle kterých nemusí být viditelné všechny funkcionality nástroje.

Demo s rolí data stewarda slouží pro vyzkoušení a poznání nástroje, není zamýšleno pro seriózní použití. Registrace i přihlášení jsou jednoduché a instance je sdílena s dalšími uživateli. Role výzkumníka má taktéž jednoduchou registraci a přihlášení. Není zde viditelný editor znalostního modelu určený pro data stewardy, k použití jsou připraveny hotové znalostní modely poskytnuté nástrojem DS Wizard. Pro organizace je možnost vlastní instance distribuované přes Docker, kde se sama organizace stará o zřízení a správu. Po domluvě je další možností využití cloudu, kde vlastní instanci organizace spravuje Data Stewardship Wizard.

3.3.1 Vytváření plánu

Většina pojmů a funkcionalit zmíněná v této podsekci bude vysvětlena znovu a více do detailu v následující sekci, která se věnuje editoru znalostního modelu.

Vytváření plánu probíhá přes vytvoření dotazníku. Jedná se o interaktivní formulář založený na znalostním modelu. Dotazník obsahuje informace o dodržení FAIR metrik a vyplněný formulář lze exportovat. Při vytváření se

3. ANALÝZA NÁSTROJŮ

zadává jméno dotazníku, verze znalostního modelu, volí se přístupnost a případně tagy. Dotazníky mohou mít soukromou přístupnost, veřejnou, a veřejnou pouze pro čtení.

Jelikož odpovědi na jednotlivé otázky nemusí být uživateli ihned známé, jsou otázky rozděleny do *fází*. Otázky jsou tak rozděleny do tří fází, dle toho, kdy jsou žádoucí:

- před odesláním žádosti,
- před odesláním plánu správy dat,
- před dokončením projektu.

Při vyplňování se vybere současná fáze dotazníku. Otázky, které by měly být v dané fázi vyplněny, budou označeny červenou barvou.

Obvykle je součástí otázky popis, vhodná fáze pro vyplnění, reference pro nápovědu, přiřazený expert, a odpověď. Po zvolení odpovědi se zobrazí případné rady a další otázky dle stromovité struktury modelu. Pro každou otázku lze odeslat zpětnou vazbu, že je například nejasná nebo je v ní chyba. Na základě *štítků* u odpovědí se počítají FAIR metriky. Ty poté značí, v jaké míře jsou jednotlivé body FAIR splněny. Tyto statistiky nejsou certifikované, mají spíše informativní charakter a indikují možné nedostatky v plánu, kde je potřeba zlepšení.

1.b.5 Did you choose the workflow engine you will be using? + !

Desirable: *Before Submitting the DMP*

Data Stewardship for Open Science: *ydj*

a. Workflows will be run from a computer command line
Good DMP Practice

b. Yes, we will be using what we always use
Good DMP Practice

c. Yes, we have selected one based on what features we need
Good DMP Practice

d. Explore ☰

Obrázek 3.5: Nástroj Data Stewardship Wizard

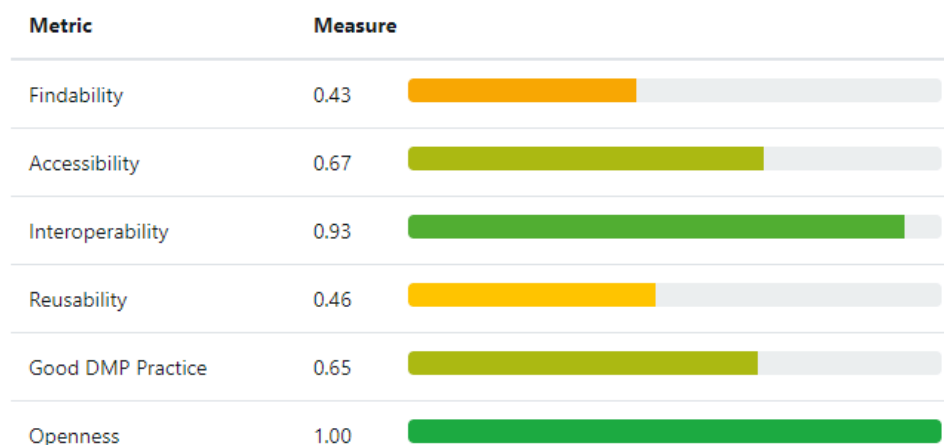
Odpovědi mohou být označeny štítky:

- Good DMP practice – osvědčené postupy
- Openness – otevřenost
- Findability – naležitelnost
- Accessibility – přístupnost
- Interoperability – interoperabilita
- Reusability – znovupoužitelnost

Kromě obecných osvědčených postupů a otevřenosti jsou zde tedy i všechny body FAIR. Pokud je štítek zelený, značí, že tato odpověď by měla být preferována – jedná se o vhodný postup a je moudré ho dodržovat. V případě červeného štítku se jedná o opačnou situaci – mělo by se zvážit, zda by tento postup měl být opravdu vybrán. Plán by ale měl vždy odrážet skutečný stav, a to i v případě, že by měl být v rozporu s preferovanými osvědčenými postupy.

Otázky mohou být označeny jako *TODO*, pokud uživatel neví, jak je zodpovědět nebo zatím nemá potřebné informace. Všechny takto označené otázky jsou přístupné z *TODO* seznamu, který se nachází v menu v levé části obrazovky. Seznam se zobrazí po označení první otázky.

V menu je kromě *TODO* seznamu, výběru fáze a rozcestníku na kapitoly otázek i celkový přehled. Po jeho otevření se uživatel dozví, kolik otázek je již zodpovězeno, a míru splnění metrik ze štítků. Ty jsou reprezentovány grafy. Z těchto informací lze přehledně zjistit, ve kterých metrikách plán vyniká či zaostává.



Obrázek 3.6: Příklad grafu metrik v Data Stewardship Wizard

Data Stewardship Wizard má pro lepší práci s metadaty integrován FAIR-sharing.org. Pomocí této služby se referenční datasety mohou přidat výběrem

z širokého seznamu. Automaticky k nim budou přiřazeny odkazy. Je tak jednoduché a pohodlné najít zdroj a informace o něm.

Po vytvoření dotazníku je možné ho exportovat dle různých šablon do zvoleného formátu. Dotazník lze vyplňovat opětovně a tento iterativní způsob je také zamýšlen. Uživatel obvykle nemá všechny požadované informace ihned a některé odpovědi se v průběhu času změny. Existující dotazník je také možné naklonovat, tedy nakopírovat. To je vhodné zejména v souvislosti s veřejnou přístupností pouze pro čtení. Může tak být vytvořen předvyplněný dotazník, který je poté sdílen a ostatní uživatelé si ho mohou naklonovat a vyplnit. Při vydání nové verze modelu je potřeba dotazníky migrovat – upgradovat na novou verzi modelu. Migrace není povinná, ale bez ní bude dotazník stále vázán na starší model a bude postrádat provedené změny.

3.3.2 Editor znalostního modelu

Znalostní model je šablonou pro dotazníky. Je to stromovitá struktura skládající se z kapitol, otázek a odpovědí. Otázek v modelu je mnoho, v řádu stovek, ale při odpovídání se prochází pouze určitá větev dle zvolených odpovědí, a uživatel tak přijde do kontaktu pouze se zlomkem z celkového počtu. Nástroj navíc nenutí k vyplnění všech otázek, čímž se může konečný počet zodpovídaných otázek ještě snížit. Model není pevně daný, v průběhu času pravděpodobně nastane situace, kdy bude potřeba něco změnit, například přidat otázku.

DS Wizard nabízí možnost vytvářet vlastní znalostní modely. Při vytváření nového znalostního modelu se zadává název, ID a případně model, na kterém je založen. Znalostní model může být vytvořen úplně od začátku nebo rozšířením či změnou stávajícího modelu.

Zodpovědnou osobou za vytváření a správu modelů je obvykle data steward, výzkumníci tuto možnost nepotřebují a nevyužijí, pracují s již hotovými modely.

Vytváření modelu sestává z tvorby otázek a odpovědí. Ty mají mnoho nastavitelných funkcí a vlastností, text lze formátovat pomocí značkovacího jazyka Markdown.

Otázky

Otázky jsou rozděleny do kapitol s možným popisem. Otázka má název, popis a eventuálně i přiřazené experty, které lze v případě nejasností kontaktovat. Také může obsahovat referenci s odkazem na webovou stránku s více informacemi, které jsou poskytovány knihou Data Stewardship For Open Science, kterou napsal prof. Barend Mons, nebo externími zdroji.

Pro přidání otázky se zadá typ, název (obvykle podobou otázky), popis, vhodná fáze a *tag*. Tagy určují, jaké otázky budou zobrazeny. Lze je využít například k tomu, aby se nemusely vytvářet dva znalostní modely, které by byly velmi podobné, pouze by jeden z nich byl detailnější a měl více otázek. Vytvoří

se jeden model a jednotlivé otázky se označí tagy. Při vytváření dotazníku se v nabídce vybere vyhovující tag a dotazník se bude skládat z otázek spadajících pod vybraný tag. Dále lze přidat větší počet odpovědí, referencí a expertů.

Odpovědi

Odpovědi mohou nabývat několika typů:

- *Možnosti* – uživatel vybírá jednu z možností jako odpověď. Tento typ se používá pro větvení, dle odpovědi se mohou zobrazit další související otázky a rady.
- *Seznam* – odpovědi může být několik položek, ale všechny jsou stejného typu. Pokud mají navazující otázky, jsou tyto otázky u každé z položek.
- *Hodnota* – zadávaná uživatelem, může to být číslo, datum, string (krátký text) nebo text. Těchto typů je v nástroji použitých málo, je zde snaha se tomuto typu vyvarovat, hlavně pokud by měl být odpovědí text. Tyto odpovědi mohou být nejasné, příliš stručné či zbytečně detailní a nelze je použít pro měření splnění metrik.
- *Integrace* – podobné typu Hodnota, ale při zadávání textu nabízí odpovědi, obvykle převzaté z nějakého externího zdroje jako je FAIRsharing. Součástí je i odkaz.

Díky tomu, že odpovědi typu Možnosti mohou obsahovat rady, další navazující otázky a štítek splnění metrik, jsou nejvíce používané. Rady i navazující otázky se zobrazí pouze při vybrání odpovědi, ke kterým jsou vázané. Rada je blok textu, který může popisovat, že je scénář s touto odpovědí nepravděpodobný a zda uživatel něco neopomněl. V případě odpovědi, která není příliš v souladu s FAIR, může více vysvětlit souvislosti, aby uživatel zvážil vhodnější postup.

Při přidání štítku se zvolí metrika, váha a míra. Metrikami jsou body FAIR, váha i míra jsou čísla na stupnici od nuly do jedné. Váha určuje důležitost odpovědi, od nedůležité po velmi důležitou. Míra určuje, jak bude výsledek touto odpovědí ovlivněn, zda k horšímu, nebo k lepšímu, a tedy jestli je v souladu s FAIR. Výsledek metrik se počítá jako vážený průměr ze všech zvolených odpovědí.

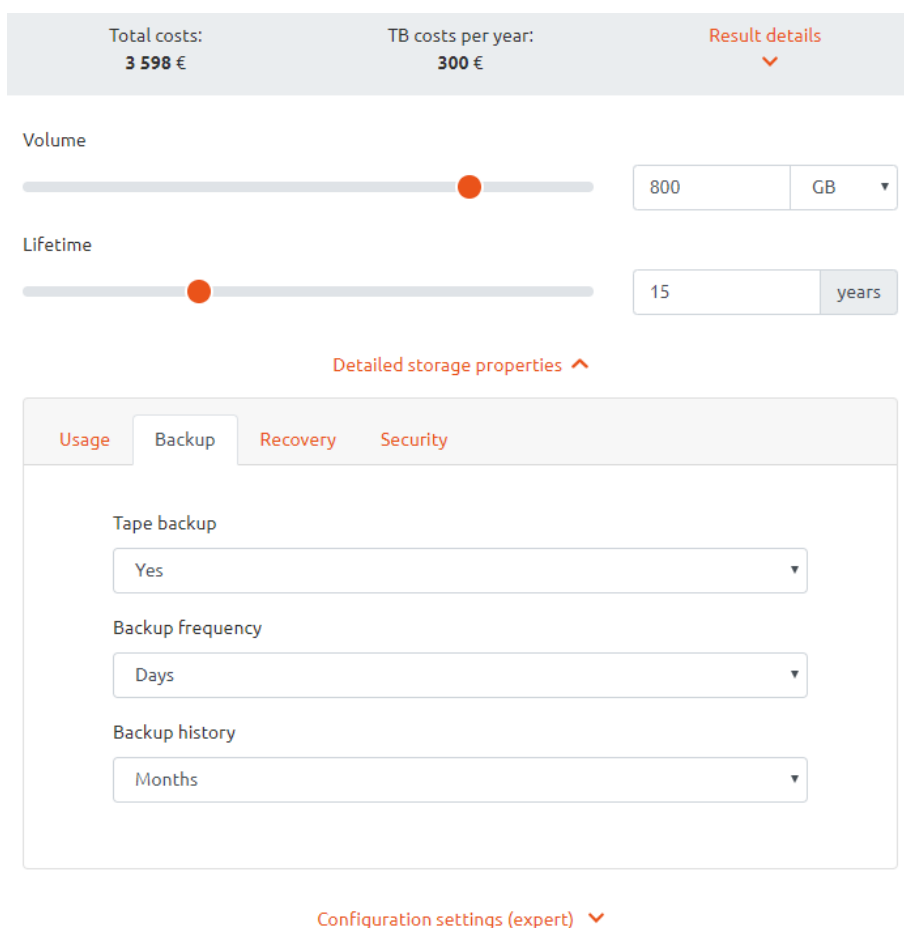
Kdykoliv během práce na modelu lze pomocí náhledu zjistit, jak bude dotazník vypadat. V náhledu fungují veškeré prvky jako je zobrazení nápovědy a navazujících otázek při vybrání příslušné odpovědi.

Než může být znalostní model použit, musí být nejdříve publikován. Při publikaci se zadává verze, popis, licence a readme. Po publikování již nemůže být model měněn. Pokud je potřeba provést nějaké změny, je potřeba model publikovat znovu pod novou verzí. Po publikaci je model zařazen do záložky Znalostní modely a je možné dle něj vytvářet dotazníky. Znalostní modely mohou být importovány a exportovány pomocí registru.

3.3.3 Vyhodnocení ceny úložiště

Data Stewardship Wizard má mimo jiné i funkci spočítání odhadované ceny úložiště. Výsledkem je cena za rok a celková cena, obě zobrazeny v eurech. Je možné otevřít podrobnější informace, jak nástroj toto číslo vypočítal. Součástí jsou jednotlivé položky jako disky, servery, nastavení, elektřina. Většina z těchto položek obsahuje další detaily s příslušnými náklady. Na stránce této funkcionality je uvedeno, že se jedná o odhad dle zadaných vlastností a aktuální konfigurace. Výpočet má tři úrovně nastavení dle množství poskytnutých detailů.

Základní úroveň nastavení se skládá ze dvou parametrů, objemu dat (zadaného v GB, TB nebo PB) a doby v letech. Hodnoty lze zadat pomocí posuvníků, nebo přímým zadáním číselné hodnoty.



Total costs: **3 598 €**

TB costs per year: **300 €**

Result details ▾

Volume

800 GB ▾

Lifetime

15 years

Detailed storage properties ▲

Usage Backup Recovery Security

Tape backup

Yes ▾

Backup frequency

Days ▾

Backup history

Months ▾

Configuration settings (expert) ▾

Obrázek 3.7: Vyhodnocení ceny úložiště v Data Stewardship Wizard

Detailnější úroveň nastavení má otázky převážně formou výběru jedné z několika možností a je rozdělena na čtyři sekce:

- Použití – nastavuje se počet změn denně a zda je obsahem více malých souborů, méně velkých, či databáze.
- Záloha – volí se, jak často se mají provádět zálohy a po jakou dobu mají být přístupné.
- Obnova – určuje, do jaké doby bude případný problém vyřešen a po jakou dobu může být systém nedostupný.
- Bezpečnost – nastavení přístupových práv.

Nejdetailnější úroveň nastavení se skládá pouze ze zadávání číselných hodnot a je rozdělena do pěti sekcí:

- Obecné – cena elektřiny, energetická účinnost.
- Pevné disky – cena, velikost, spotřeba a životnost disku.
- Servery – cena serveru, cena racku, spotřeba a životnost serveru.
- Zálohovací pásky – rychlost, kapacita, cena pásek a cena páskové jednotky.
- Vytváření sítí – cena sítě, firewallu a jeho údržby.

3.3.4 Technická stránka a business model

Data Stewardship Wizard má velmi flexibilní exporty. Existuje na to separátní worker v Pythonu, který používá Jinja2 šablony a flexibilně dle specifikace v JSONu poskytuje převody do jiných formátů díky Pandoc a wkhtmltopdf. Pro nástroj byl použit jazyk Elm. Nástroj je distribuován přes Docker pomocí GitHub Actions. Pro lepší práci s metadaty je integrován FAIRsharing.org.

Nástroj je k použití se všemi funkcionalitami zcela zdarma. Nabízí vlastní cloud, který je díky infrastruktuře ELIXIR zdarma. Vývoj se financuje z projektů.

3.4 DataWiz

Vydání: 2017

Provozovatel: Leibniz Institute for Psychology Information and Documentation (ZPID)

Funkčnosti: Vytváření, editování a sdílení plánů správy dat. Vícejazyčný, zaměřený na investory a jejich potřeby. Podpora zpracování a dokumentace psychologických dat dle standardů.

Zdroj: Nástroj [20], GitHub [21]

Webový asistent DataWiz pomáhá s přípravou a dokumentací výzkumných dat v průběhu celého projektu. Umožňuje vytvářet plány správy dat dle různých požadavků investorských agentur, a dle těchto požadavků je i exportovat. Je možné exportovat celý projekt či konkrétní studie, nebo je naopak nahrávat. Umožňuje vysoce kvalitní dokumentaci dle současných standardů a zajišťuje dlouhodobou znovupoužitelnost výzkumných dat.

3.4.1 Vytváření plánu

Tento nástroj neumožňuje výběr šablony dle investora ani přizpůsobení. Jedinou podobnou vlastností je značení u jednotlivých otázek, které upozorňují na povinné otázky některého ze tří investorů. U každé otázky je uvedeno alespoň stručné vysvětlení s případnými odkazy. Tyto odkazy neopouští DataWiz, přesměrovávají pouze na jiné části, kde jsou pojmy podrobně vysvětleny. U vysvětlivek jsou zároveň uvedeny i použité zdroje. Většina otázek je formou otevřeného textu.

Vytváření plánu je rozděleno celkem na šest sekcí. Čtyři z nich se zabývají klasickou tvorbou, zodpovídáním otázek, sdílením, exportem. Zbylé dvě se věnují vytváření studií a nahrávání materiálů. Mezi zmiňované čtyři sekce pro tvorbu patří:

1. Dokumentace projektu

První stránka obsahuje více otázek, než tomu je zvykem v jiných nástrojích. Otázky jsou téměř stejné jako u předchozích nástrojů, ale všechny jsou formou otevřeného textu.

2. Plánování správy dat

Sekce obsahuje nejprve vysvětlení tří symbolů. Tyto symboly mohou být uvedeny u jednotlivých otázek a značí, že některý z investorů tuto otázku vyžaduje. Investory reprezentovanými symboly jsou program EU Horizon 2020, BMBF (Spolkové ministerstvo pro vzdělání a výzkum v Německu) a DFG (Německá výzkumná společnost). Tato hlavička je viditelná ve všech osmi skupinách, do kterých jsou otázky rozděleny:

- a) Administrativní data
Opakuje se několik otázek zodpovězených v Dokumentaci projektu, společně se zadanými odpověďmi, které zde ale nejdou změnit. Pokud je chce uživatel změnit, musí se vrátit na předchozí stránku. Dále se zde nachází otázky o trvání projektu, jeho partnerech a cílové skupině, které je projekt určený.
- b) Výzkumná data
Tato část se zabývá otázkami mnoha typů. Jsou zde otázky na použití již existujících dat, způsob získání dat, výzkumné metody, zajištění kvality, důvody pro uchování dat a zároveň i otázky na úložiště. Většina otázek v této části má již odpovědi formou výběru.
- c) Dokumentace
Otázky v této části mají o něco rozsáhlejší popisy, než tomu bylo v předchozích skupinách. První otázkou je účel dokumentace formou výběru, ostatní otázky se týkají převážně standardizace a monitorování dokumentace. Dokumentace je důležitá převážně pro investora EU Horizon 2020, který má všechny tyto otázky označeny jako vyžadované.
- d) Sdílení
Sdílení dat obsahuje jen tři otázky. Zda je povinné data sdílet, zamýšlené a očekávané využití dat a přístup k datům pro třetí strany.
- e) Úložiště a infrastruktura
Tato skupina otázek je důležitá pro program EU Horizon 2020 který vyžaduje všechny z těchto otázek, a také BMBF, které vyžaduje až na jednu taktéž všechny. Otázky se netýkají jen samotného úložiště, ale i jmenných konvencí a speciálních technických požadavků, jakým může být specifický software.
- f) Organizace a opatření
Pro BMBF jsou otázky Organizace a opatření vyžadovány, ostatními investory příliš ne.
- g) Etické a legální aspekty
Většina otázek je vyžadována všemi investory a jsou formou výběru ano/ne.
- h) Náklady
Nejstručnější skupinou otázek jsou Náklady, které obsahují jedinou otázku ohledně nákladů na správu dat, s odpovědí formou výběru. K otázce se mohou vztahovat dvě další podotázky na posouzení odhadované ceny.

3. ANALÝZA NÁSTROJŮ

The screenshot shows the DataWiz interface with three main sections:

- Reuse of already existing data:** Includes a search bar with the text "Search for already existing data was not successful" and a dropdown arrow. Below it is an information icon and a paragraph of text explaining the importance of checking for existing data at the start of a project.
- Research Method:** Features four checkboxes: "survey data" (checked), "experimental data", "test data", and "other". Below the checkboxes is an information icon and a note: "Characterize the type of data by means of research method (multiple answers are possible)".
- Reproducibility of the data collection:** Contains a large empty text area for input. Below it is an information icon and a note: "To what extent is the data collection process reproducible? Data collections for secondary analyses (reuse of existing research data, meta analyses, bibliographical analyses) or simulation studies typically show a high degree of reproducibility."

Obrázek 3.8: Nástroj Datawiz

3. Spolupracovníci

V této sekci by měli být vyjmenováni všichni, kteří mají přístupová práva do projektu v DataWizu. Administrátor projektu může přes email pozvat nové spolupracovníky/příspěvatele, a určit jim role. Přirazuje se práva ke konkrétním studiím.

4. Export

Při exportování plánu si uživatel zvolí, zda chce s plánem exportovat i materiály. Tedy použité studie a datasety. Ty jdou exportovat buď kompletně, nebo jen určité části. Po navolení lze stáhnout exportovaný plán v zip formátu.

3.4.2 Studie

Tato sekce umožňuje vytvořit novou studii či upravit existující. Studie slouží k popsání postupu získání dat a správě získaných dat.

Nejdříve je potřeba studii vytvořit, aby bylo možné využít další možnosti. Studie je tvořena otázkami jako byla tvorba plánu. Místo symbolů investorů jsou zde ale symboly informující o fázi, kdy by otázka měla být vyplněna. Zda již před začátkem sběru dat, nebo až po jeho dokončení.

Stejně jako u základních otázek při vytváření plánu správy dat se i zde vyplňuje název, ID, autoři a popis. Dále je možné přidávat software a publikace, ale až po uložení studie. Otázkami, které se vyplní až po dokončení sběru dat, jsou typicky publikace a autoři. U většiny otázek jsou ale přítomny oba symboly.

Po tom, co je studie vytvořena, je možné ji uzamknout, aby nemohla být upravována. Lze ji opět odemknout. Zároveň je možné přidat dataset či nahrát

soubory. Přidání datasetu probíhá vyplněním názvu a popisu a následným nahráním souboru. Poté je možné spravovat číselník (codebook) datasetu.

Je také možné data studie exportovat. Navzdory tomu, že je možné nástroj přepnout do angličtiny, i se všemi otázkami a popisy, je podstatná část exportovaného dokumentu v němčině.

Research Design: ψ ✓

Cross-sectional study ▼

i State the time dimension of your data collection, whether it is a single or multiple measurement occasion. Please note that a cross-sectional data collection can also extend over a longer period of time. Please code longitudinal study only when there are two or more measurement occasions for each subject, and if findings regarding the time dimension are of interest (e.g. change in parameters between two measurement points). More information on [the essentials of transparent science](#).

Research Method: ψ ✓

Survey Test development Experiment

i Characterize the study type by means of research method (multiple answers are possible).
 Surveys collect information on human experience and behavior (no intervention).
 Test data aim at the development or standardization of psychological instruments or tests.
 (Quasi-)Experiments manipulate or control certain conditions in order to analyze (causal) effects.
 More information on [the essentials of transparent science](#).

Obrázek 3.9: Studie v Datawiz

3.4.3 Materiály pro projekt

Slouží k nahrávání souborů. DataWiz poskytuje oblasti pro nahrání materiálů na úrovni projektu a na úrovni studií odděleně. Maximální velikost nahrání je 1,5 GB. Soubory nahrané ke konkrétním studiím tedy odtud nebudou viditelné.

3.4.4 Technická stránka a business model

DataWiz je implementován v Javě (Enterprise Edition) společně s populárními technologiemi, jako je jQuery a jeho knihovny. Pro činnost DataWizu je zapotřebí databázového serveru (MySQL) a případně Apache serveru s PHP.

Nástroj je k použití se všemi funkcionalitami zcela zdarma.

3.5 easyDMP

Vydání: 2015

Provozovatel: EUDAT and UNINETT Sigma2

Funkčnosti: Vytváření plánů správy dat podle šablon, sdílení. Zaměření na snadnou ovladatelnost a umožnění vyplnění i s případnou minimální znalostí správy dat.

Zdroj: Nástroj [22], GitHub [23], dokumentace [24]

easyDMP má za cíl poskytnout nástroj umožňující jednoduchým způsobem vytvořit plán správy dat, a to i uživatelům s minimálními zkušenosti z oblasti správy dat. Toho je dosaženo transformací pokynů investorských agentur do dotazníku.

Nástroj má podobu webového formuláře sestávajícího ze sérií otázek seskupených do sekcí. Dotazník je dynamický, tedy typ a počet otázek bude závislý na předchozích odpovědích uživatele. Například šablona Horizont 2020 sestává z celkem sedmdesáti otázek rozdělených do šesti sekcí, ale s největší pravděpodobností bude uživatel zodpovídat méně otázek než tento celkový počet. Otázky jsou předem připravené a snadné na zodpovězení. Většina otázek je koncipována jako jednoduché „ano/ne“ otázky.

3.5.1 Vytváření plánu

easyDMP nabízí vytvoření plánu dle šablony. V době psaní práce byly k dispozici tři šablony:

- Horizon 2020 – zjednodušená šablona založená na pokynech dle Horizon 2020
- Horizon 2020 Expert – zkrácená šablona založená na pokynech dle Horizon 2020, předpokládá znalost správy dat
- Science Europe – šablona založená na pokynech dle Science Europe

Na úvodní stránce se zadá název plánu a nepovinná zkratka pro interní použití. Pod každou z otázek jsou pokyny a rady pro vyplnění.

Dále je tvorba rozdělena do sekcí, které jsou dle šablon odlišné. Jelikož má nástroj easyDMP jen tři šablony, bude popsán průchod každou z nich.

Pod každou z otázek je zobrazena nápověda. Sekce má obvykle více otázek, je ale zobrazena pouze jedna otázka a další se zobrazí až po zodpovězení předchozí. Všechny otázky jsou povinné, není možné některé vynechat (výjimkou je, pokud byla otázka reakcí na odpověď předchozí otázky, v tom případě ani není zobrazena). Není možné použít tlačítko zpět pro návrat k předchozí otázce, dokud není vyplněna současná otázka. Je to vnímáno jako nevyplnění povinné otázky. Lze se libovolně přepínat pouze mezi sekcemi. Po vyplnění

poslední otázky či po kliknutí na přehled plánu se zobrazí stránka zobrazující informace o vyplnění sekcí.

V přehledu jsou viditelné otázky a jejich odpovědi, dá se pohodlněji přepínat na jednotlivé otázky a součástí je i nastavení plánu. Aby plán mohl být publikován, musí být validní. Na kontrolu validity lze použít kontrolu přístupnou z nastavení souhrnu plánu. Je možné pozvat další spolupracovníky, s právy pro editaci nebo pro čtení. Plán lze exportovat, ale pouze generací html stránky.

Obrázek 3.10: Nástroj easyDMP

Horizon 2020

Za většinou otázek je nepovinné pole pro případné další informace zadané formou otevřeného textu.

1. Shrnutí údajů o datech

Rozděleno do osmi otázek, zabývajících se důvody produkce dat, souvislostmi dat s cílem projektu, formátem a typem použitých dat. Dále otázky na znovupoužití dat, původ dat, odhadovaný objem dat a výčet komunit, pro která by data projektu mohla být přínosem.

Většina otázek je formou výběru možnosti s případným dopsáním více informací. Dvě otázky jsou formou zadání textu a následného vybrání odpovídající navržené možnosti, podobně jako typ Integrace v nástroji DS Wizard. Lze zadat více možností, případně označit checkbox, že odpověď není v seznamu zahrnuta.

3. ANALÝZA NÁSTROJŮ

2. FAIR data

Sekce zobrazí oznámení, že plán by měl být FAIR. To je následované krátkým vysvětlením a odkazy s více informacemi. Po potvrzení následuje 24 otázek.

3. Přidělení zdrojů

Volí se, zda bude projekt využívat národní, instituční či ještě jinou infrastrukturu, kdo bude zodpovědný za správu dat a jak bude zajištěno znovupoužití dat po skončení projektu.

4. Bezpečnost dat

Informace o naložení s daty, která mají omezené využití a nebudou již potřeba. Kde budou uložena, po jakou dobu, kdy budou vymazána. Takovými daty jsou typicky raw data, ze kterých již byly získány potřebné informace a vyvozeny závěry.

5. Etické aspekty

Zadávají se informace o případných etických aspektech, které nebyly zmíněny v předchozích sekcích.

6. Ostatní

Poslední otázka se týká možného využití dalších procedur pro správu dat.

Horizon 2020 Expert

Šablona vyžaduje téměř stejné informace jako šablona Horizon 2020, ale místo mnoha navazujících otázek vysvětlených detailněji je zde pro každou sekci jen jedna otázka formou otevřeného textu kam by uživatel měl vypsát všechny informace, které obvykle pokrývají širokou škálu otázek. Druhá sekce, FAIR data, obsahuje čtyři navazující otázky, kde každá vyžaduje jako odpověď popis podstoupených kroků pro splnitelnost jednoho z bodů FAIR.

Science Europe

Všechny odpovědi jsou formou otevřeného textu pokrývající širokou škálu otázek, podobně jako předchozí šablona Horizon 2020 Expert.

1. Administrativní informace

Zadává se název projektu, financující programy do kterých projekt patří a číslo projektu (to je nepovinné).

2. Popis a sběr dat či využití existujících dat

Popis způsobu produkce dat či využití existujících dat. Měl by být popsán i používaný software a důvod, proč případná existující data nejsou dostatečná. Dále by měl být popsán typ dat a jejich objem.

3. Dokumentace a kvalita dat

Informace o organizaci a dokumentaci dat, jaké standardy budou použity a jak bude kontrolována konzistence a kvalita dat.

4. Úložiště a zálohování

Mělo by být popsáno, kde budou data uložena a kolik bude záloh. Pokud budou součástí projektu i nějaká citlivá data, tak by mělo být popsáno, jak budou chráněna a jak k nim bude přistupováno.

5. Právní a etické požadavky

Popis způsobu ochrany případných osobních dat. Jak bude získán souhlas a jak budou spravována. Zároveň musí být popsáno, kdo je vlastník dat, kdo bude mít přístupová práva a jaká licence se na data vztahuje. Pokud jsou data a jejich správa nějak ovlivněna z etických hledisek, mělo by být uvedeno, jak bude situace řešena.

6. Sdílení dat a dlouhodobé zachování

Informace o tom, jak a kdy budou data sdílena. Jak bude zajištěna přístupnost dat po danou dobu a jak budou datům přiřazeny persistentní identifikátory.

7. Zodpovědnost správy dat

Popsat, kdo bude zodpovědný za správu dat, ideálně poskytnout informace o osobách a jejich konkrétních rolích. Mezi tyto role patří obvykle produkce metadat, správa úložiště, sdílení dat a jejich archivace. V odpovědi by měla být zahrnuta i osoba zodpovědná za implementaci plánu a případné opravy.

3.5.2 Technická stránka a business model

Nástroj používá PostGreSQL databázi. API je v současné době velmi základní a je read-only pohledem na data v modelu, kterým je objektově relační zobrazení. O zpracování dat externích služeb se stará EEStore, ze kterého si poté model načítá data. Předpoklady pro easyDMP je Python verze alespoň 3.5 a SQL databáze kterou podporuje Django.

Nástroj je k použití se všemi funkcionalitami zcela zdarma.

3.6 ezDMP

Vydání: 2018

Provozovatel: Interdisciplinary Earth Data Alliance (IEDA)

Funkčnosti: Vytváření a editování plánů správy dat. Zaměření na strukturované otázky místo otevřeného textu, umožňuje analyzovat a kontrolovat dodržování požadavků ohledně dat.

Zdroj: Nástroj [25], GitHub [26]

Nástroj ezDMP byl vytvořen pro jednoduché a jasné splnění požadavků pro plány dle národní vědecké nadace (The National Science Foundation, NSF). Ta vytvořila pokyny pro tvorbu plánů, ale jejich vytvoření bylo obtížné. Plány byly psány formou volného textu a bylo složité zjistit, co konkrétně NSF požaduje a jak by to mělo být organizováno. Složitě bylo i pro NSF porozumět dodaným plánům. Nástroj zajišťuje, že všechny požadavky budou splněny, a to přehledně a strukturovaně. ezDMP je založeno na současném nástroji svého provozovatele IEDA. Rozhraní umožňuje snadné zadání a propojení, a poskytuje návrhy a informace specifické pro NSF. Tím se zajistí, že uživatelé najdou správné řešení pro dlouhodobou správu dat.

The screenshot displays the ezDMP interface. At the top, there are two main sections: "NSF Submission Target" and "Solicitation Info (optional)".

The "NSF Submission Target" section features a dropdown menu with the following options:

- Select a Directorate -
- Select a Directorate -
- Directorate for Biological Sciences
- Directorate for Geosciences
- Directorate for Mathematical & Physical Sciences
- Directorate for Social, Behavioral & Economic Sciences
- Computer and Information Science and Engineering

The "Solicitation Info (optional)" section contains a text input field with the placeholder text "e.g., GeoPRISMS, LTER". Below this field, there is a link that says "Click here for more information about metadata standards, backups, ethics and privacy".

Below these sections, there is a "Products" section. It shows a list of products, with the first one being "Climate Software". The product details include:

- Icon: A blue circle with a white upward-pointing arrow.
- Name: Lorem
- Description: Morbi scelerisque luctus velit. Donec quis nibh at felis congue commodo.
- Buttons: "Edit" and "Delete"

At the bottom of the "Products" section, there is a blue button labeled "Add a Product".

Obrázek 3.11: Nástroj ezDMP

3.6.1 Vytváření plánu

Na rozdíl od ostatních nástrojů má ezDMP velmi málo otázek a další informace se zadávají přidáním produktu. Mezi základní otázky patří název projektu, hlavní výzkumník a jeho instituce, spolupříspěvníci a celkový popis správy dat. Důležitou otázkou je zvolení oddělení NSF. Výběrem se zvolí konkrétní ředitelství a poté oddělení. Po zvolení je možné si přečíst podmínky plánu pro

vybrané oddělení. Otázky, které se zodpovídají vybráním jedné z odpovědí, jsou označeny zelenou barvou, ostatní, zodpovídané formou otevřeného textu, jsou bílé. Dokud není otázka vyplněna, je v okénku odpovědi zobrazen popis jako nápověda pro vyplnění.

Produkty

Produktů pro přidání je celkem pět typů:

- Software
- Datový produkt
- Osnova
- Fyzický vzorek
- Pracovní postup

Formuláře pro jednotlivé produkty jsou velice podobné, odlišnější je formulář pro Fyzický vzorek. První otázkou je vědecké pole formou výběru. Touto formou jsou následně ještě otázky na repozitář či distribuční platformu, licenci znovupoužití a případné použití superpočítačového centra. Zadává se zodpovědný výzkumník a popis, ostatní otázky se týkají technické stránky – formát, objem, vydání. Na závěr lze zadat vztah produktu k ostatním produktům. Zda spolu souvisí, zda je jeden od druhého odvozen či je vstupem. Ve formuláři Fyzického vzorku je místo otázky na licenci znovupoužití otázka na repozitář fyzického vzorku. Kromě popisu v nevyplněné otázce nejsou u základních otázek ani u produktů žádné další nápovědy ani odkazy.

Nástroj neumožňuje sdílení plánu. Export je možný stažením pdf souboru.

3.6.2 Technická stránka a business model

Nástroj je open-source a je přístupný na GitHubu. ezDMP používá JSON formát, informace jsou uloženy v JSON objektu. Nástroj umožňuje i lokální instalaci, ke které poskytuje návod. Lze implementovat s ezDMP API nebo vlastní API, pokud spravuje objekty dle dokumentace a poskytuje odkazy na určené soubory. ezDMP používá JWT Authorization headery s knihovnou Satellizer AngularJS.

Nástroj je k použití se všemi funkcionalitami zcela zdarma.

3.7 QUT DMP

Vydání: 2014

Provozovatel: Queensland University of Technology

Funkčnosti: Vytváření, editování a sdílení plánů správy dat. Vytvořen pro potřeby Queensland University of Technology.

Zdroj: Nástroj [27], GitHub [15]

QUT DMP je založen na DMPRoadmap a přesto, že uvádí „Powered by DMPOnline“, je od DMPOnline odlišnější než dříve jmenovaný DMPTool či francouzský DMP OPIDoR. Univerzita (QUT) plán správy dat často vyžaduje, a i z tohoto důvodu mají vlastní nástroj na tvorbu plánu, přizpůsobený vlastním požadavkům.

3.7.1 Vytváření plánu

Při vytváření plánu je automaticky vybrána instituce Queensland University of Technology a šablona QUT Data Management Plan. Poté je vytvořen plán a je možné zadat název.

Vyplňování odpovědí na otázky je velmi podobné DMPOnline a v odstínu modré barvy. Také umožňuje formátování textu odpovědi, zobrazuje informace o zodpovězení, ale neumožňuje přidat komentář. Na rozdíl od ostatních nástrojů založených na DMPRoadmap má QUT DMP tímto způsobem i zodpovídání úvodních otázek, kde se vyplňuje název projektu, výzkumníci, příslušnost k univerzitě a další. Tato část má také oproti ostatním nástrojům více otázek.

Plán lze sdílet s jinými uživateli přes možnost přidání spolupracovníka. Ten může mít práva spoluvlastníka, editora, či klasického uživatele, který může plán pouze číst. Plán lze exportovat v několika formátech, mimo jiné v klasickém pdf.





3.7.2 Technická stránka a business model

Stejně jako ostatní z této rodiny nástrojů je QUT DMP open-source a GitHub je přístupný ze seznamu forků z [15]. DMPRoadmap je Ruby on Rails aplikace, což je framework pro vývoj webových aplikací napojených na databázi. Je vyžadováno MySQL nebo PostgreSQL.

Nástroj je vytvořen pro potřeby Queensland University of Technology a je tedy zcela zdarma.

— Documentation and Metadata (1 / 1)

What documentation and metadata will accompany the data?

B *I*    

Donec quis nibh at felis congue commodo. *Aliquam erat volutpat.* Etiam quis quam. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Nulla turpis magna, cursus sit amet, suscipit a, interdum id, felis. Vivamus porttitor turpis ac leo.





Save

Answered just now by machacova.terka@gmail.com

+ Ethics and Legal Compliance (0 / 2)

— Storage and Backup (1 / 2)

How will the data be stored and backed up during the research?

B *I*    

Duis sapien nunc, commodo et, interdum suscipit, sollicitudin et, dolor:

- Maecenas ipsum velit, consectetur eu lobortis ut, dictum at dui.
- Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos.

Vivamus porttitor turpis ac leo. Ut tempus purus at lorem.

Save

Answered less than a minute ago by machacova.terka@gmail.com

Obrázek 3.12: Nástroj QUT DMP

3.8 DMP OPIDoR

Vydání: 2016

Provozovatel: Institut de l'Information Scientifique et Technique (INIST)

Funkčnosti: Vytváření, editování, sdílení a publikování plánů správy dat. Uzpůsobený pro potřeby výzkumu, institucí a investorů ve Francii.

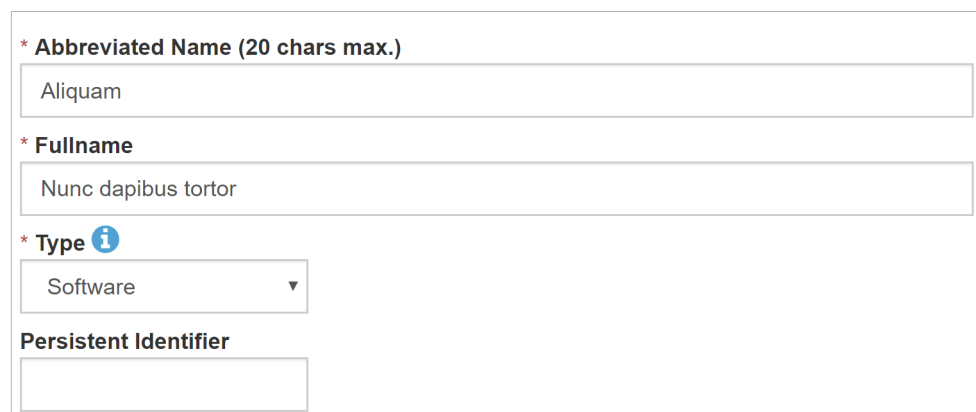
Zdroj: Nástroj [28], GitHub [15]

DMP OPIDoR je založen na DMPRoadmap a byl vytvořen a přizpůsoben francouzským výzkumníkům. Obsahuje mnoho šablon reprezentujících požadavky různých investorů a institucí. Uživatelům jsou položeny tři otázky, na základě kterých je zvolena vhodná šablona, a v průběhu tvorby plánu mají k dispozici instrukce a rady.

3.8.1 Vytváření plánu

Nástroj DMP OPIDoR je, stejně jako DMPTool, velice podobný DMPOnline. Umožňuje výběrem vybrat investora a šablonu, sekce jsou téměř totožné, pouze ve světle modré barvě místo oranžové. Výchozí šablonou je šablona Science Europe.

Navíc je zde sekce Výstupy výzkumu, kde lze zadat více výstupů. U každého se zadává zkrácený a celý název, případný persistentní identifikátor a typ. Ten se vybírá ze seznamu, a součástí této otázky je i odkaz pro více informací o jednotlivých typech.



The screenshot shows a form with the following fields:

- * Abbreviated Name (20 chars max.)**: Text input field containing "Aliquam".
- * Fullname**: Text input field containing "Nunc dapibus tortor".
- * Type**: Dropdown menu with an information icon (i) and the selected option "Software".
- Persistent Identifier**: Empty text input field.

[Add a research output](#)

Obrázek 3.13: Nástroj DMP OPIDoR

Jelikož DMP OPIDoR používá odlišnou výchozí šablonu, otázky jsou odlišné, ale uspořádání skupiny otázek, formátování textu odpovědi, komentáře,

a informace o zodpovězení otázky jsou stejné jako u DMPOnline. Umístění nápovědy, kterou zde poskytuje Science Europe, je taktéž stejné. Je možné přidat další organizace pro pomoc s plánem, mezi nimi i DCC. Odpovědi otázek jsou vázány ke konkrétním výstupům výzkumu. Otázka tak může být vyplněna odlišně pro každý výstup, nebo lze označit checkbox, že odpovědi dané skupiny otázek jsou pro všechny výstupy stejné. Pak se vyplňují odpovědi pouze pro první výstup, a ostatní již nelze označit.

Nástroj má v sekci Sdílení jednu možnost viditelnosti plánu navíc. Soukromá viditelnost DMPOnline, kdy plán vidí jen tvůrce plánu, přidání spolupracovníci a administrátoři, zde byla rozdělena na dvě možnosti. Její verzi v DMP OPIDoR je viditelnost administrátor. Soukromá zde totiž neumožňuje viditelnost administrátorům, pouze tvůrci plánu a přidáním spolupracovníkům. Ve stažení plánu lze zvolit, které výstupy výzkumu mají být součástí a zda chce uživatel výstupy řadit dle skupin otázek či rovnou podle otázek. Součástí nástroje je i přehled šablon s detailními informacemi, a přehled veřejných plánů správy dat. Plány i šablony lze stáhnout.

3.8.2 Technická stránka a business model

Stejně jako ostatní z této rodiny nástrojů je DMP OPIDoR open-source a GitHub je přístupný ze seznamu forků z [15]. DMPRoadmap je Ruby on Rails aplikace, což je framework pro vývoj webových aplikací napojených na databázi. Je vyžadováno MySQL nebo PostgreSQL.

Nástroj je k použití zdarma, ale pro tvorbu vlastních šablon je zapotřebí mít administrátorský účet. Ten lze získat zdarma po zažádání, momentálně se ale uděluje pouze uživatelům z francouzské komunity, kteří si přejí vytvořit šablonu pro svou organizaci.

3.9 CEDAR

Vydání: 2014

Provozovatel: The Board of Trustees of Leland Stanford Junior University

Funkčnosti: Vytváření a sdílení formulářů. Zaměření na metadata a metadatové šablony. Zajištění FAIR dat, zvýšení možnosti opětovného využití dat.

Zdroj: Nástroj [29]

CEDAR není přímo nástrojem na vytváření a správu plánů správy dat jako byly předchozí nástroje, je to spíše alternativa primárně určená na metadata. I zde je ale možné vytvářet formuláře a nástroj má dobrou podporu machine-actionability.

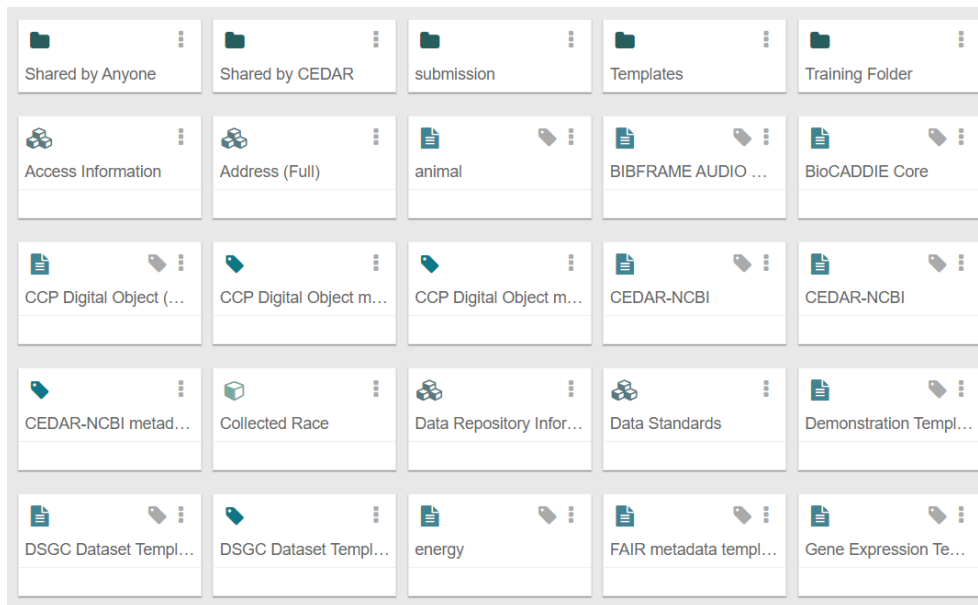
CEDAR (The Center for Expanded Data Annotation and Retrieval) byl založen v roce 2014 jako ekosystém pro vývoj, vyhodnocení a použití biomedicínských dat. CEDAR je podstatnou komponentou open science, jelikož zajišťuje FAIR data a zvyšuje možnost opětovného využití dat. Práce s daty je rychlejší a efektivnější a je tak jednodušší vytvářet a používat kvalitnější metadata. Toho je dosaženo pomocí lepších rozhraní, terminologie, praktik a analytik. CEDAR formuláře vypadají podobně jako známé webové formuláře a uživatelé s nimi tak mohou lépe pracovat. Na rozdíl od nich ale poskytují CEDAR formuláře s machine-actionable metadaty výhody. Mohou díky kontrolovaným slovníkům pomáhat dokončováním rozepsaných slov nebo pomoci s vyplňováním formulářů nabízením souvisejících termínů.

CEDAR se zaměřuje na metadatové šablony datových elementů potřebných k popsání určitých typů biomedicínských experimentů. Šablony obsahují pojmy z kontrolovaných slovníků. Je využívána celá knihovna takovýchto šablon, která slouží k přiřazení popsaných datasetů do vhodného repozitáře.

Činnosti CEDARu:

- Vytváří uživatelsky přívětivé formuláře.
- Vytváří znovupoužitelné formulářové komponenty, které je možné sdílet a použít v knihovnách.
- Umožňuje spolupracovat s uživateli a skupinami na vytvoření formulářů a jejich vyplnění metadat.
- Publikuje formuláře pro zadání metadat.
- Odesílá data a metadata přímo do repozitářů NCBI.
- Verzuje šablony pro uchování historie.

Nástroj umožňuje vytvořit novou metadatovou šablonu, či použít již existující, například některou z veřejně přístupných šablon.



Obrázek 3.14: Nástroj CEDAR

3.9.1 Technická stránka a business model

Metadata jsou založena na JSONu. Instance šablon, elementů a polí jsou reprezentována JSON objekty.

Nástroj je k použití zcela zdarma. CEDAR také pracuje na možnosti provozu nástroj jako vlastní instance přes Docker, ale zatím není připravena pro externí použití.

Zhodnocení nástrojů

Kapitola se věnuje zhodnocení dříve analyzovaných nástrojů dle kritérií, která jsou rozdělena do kategorií:

1. Přizpůsobení

Hodnotí se přizpůsobení nástroje konkrétním potřebám uživatele, jakými jsou výběr šablony, možnosti exportu či funkce zaměřené na jednotlivce nebo organizace.

2. Machine-actionability

Hodnotí se míra machine-actionability, hodnocení vychází z bodů v kapitole 2.3 Machine-actionable DMP.

3. UI/UX dle Jakoba Nielsena

Hodnocení dle desatera Jakoba Nielsena, odborníka na webový design a použitelnost. Dle konkrétních bodů se hodnotí uživatelské rozhraní a reakce uživatele na práci se systémem.

4. Závěrečné hodnocení

Na základě získaných informací budou vypsány silné a slabé stránky jednotlivých nástrojů.

V každé kategorii jsou nejdříve vypsány konkrétní kritéria, dle kterých jsou nástroje hodnoceny. Hodnocení je seřazeno dle kritérií. Nejdříve tedy budou hodnoceny všechny nástroje dle prvního kritéria, a poté dle dalších. Po ohodnocení dle všech kritérií kategorie se přejde k další kategorii hodnocení.

Na základě analýzy nástrojů a zhodnocení ve výše zmíněných kategoriích jsou na závěr této kapitoly určeny silné a slabé stránky nástroje. Hodnocení je provedeno co nejvíce objektivně.

4.1 Přizpůsobení

V této sekci je hodnoceno přizpůsobení nástroje konkrétním potřebám. Jak již bylo zmíněno v analýze nástrojů, plán správy dat není vždy jednotný, liší se v závislosti na investorovi, oblasti výzkumu i konkrétním výzkumu. Není tak možné ani žádoucí, aby existovala nějaká společná šablona pro všechny plány správy dat. Pro společný nástroj je tento scénář pravděpodobnější, ale i přesto je nejlepší mít možnost volby. Existují nástroje, které jsou natolik přizpůsobitelné, že by v nich bylo možné vytvořit a implementovat šablony a funkce ostatních nástrojů a splnit i často náročné požadavky. To však může být zároveň jejich nevýhodou. Pro potřeby některých výzkumů a investorů by zkrátka mohly být příliš složité a detailní. Překážkou by se mohla stát i odlišná terminologie v různých oblastech výzkumu. [3]

Ani splnění většiny kritérií tedy není zárukou, že nástroj je pro danou situaci nejvhodnější. Vždy je potřeba brát v úvahu potřeby konkrétního projektu a jeho okolnosti.

Ke zhodnocení byly vybrány následující funkcionality:

1. Výběr šablony
2. Úprava šablony nebo vlastní šablona
3. Jazyk nástroje
4. Možnosti exportu
5. Správa jednotlivcem či organizací
6. Přidání dalších informací
7. Registrace a přihlašování

4.1.1 Výběr šablony

Výběr šablony není a ani by neměl být nutností, ale když lze při tvorbě plánu vybrat z více šablon, nástroj se může přizpůsobit více požadavkům. Je tak vyšší šance, že bude použit, jelikož mnoho investorů vyžaduje plán dle své šablony. Mezi analyzovanými nástroji jsou i takové, které mají výběr šablon velice omezený. Většinou je tomu tak ale z prostého důvodu, že nástroj nebyl vytvářen se záměrem stát se nejvíce rozšířeným a nejvíce používaným, ale byl vytvářen pro konkrétní potřeby konkrétní skupiny.

1. DMPOnline

Nástroj umožňuje vybrat šablonu na základě investora, k dispozici je přibližně 25 šablon. Před vytvořením plánu jsou uživateli položeny tři otázky, jednou z nich je zvolení investora z širokého seznamu. Po zvolení se může zobrazit čtvrtá otázka, dávající na výběr ze šablon tohoto investora, kterých může být více. Pokud má investor pouze jednu šablonu, je zvolena automaticky.

S interpretací otázek investoři pomáhají, obvykle poskytují nápovědy a instrukce pro vyplnění. Šablony jsou spravovány DCC, je ale přesto doporučováno, aby výzkumníci vždy konzultovali požadavky přímo s investory. V přehledu projektu je více popsáno, jaká šablona je vybrána, jaké specifikace by měl plán dle této šablony splňovat a jaké otázky by měl zodpovídat. V referencích lze dohledat více informací o šablonách.

2. DMPTool

Lze vybrat šablonu, a to téměř totožně, jako DMPOnline. Zvolení šablony také probíhá dle výběru investora, pouze jsou na výběr jiní investoři, a tedy odlišné šablony. Pokud investor nabízí více šablon, lze si konkrétní šablonu zvolit ve čtvrté otázce. Pokud je šablona pouze jedna, je zvolena automaticky. Celkově je na výběr přibližně 35 šablon. Investoři poskytují instrukce k vyplnění plánu nápovědami u otázek. Správu šablon zajišťuje DMPTool, i přesto je ale doporučovaná konzultace šablony s investorem. Více informací o šablonách může uživatel najít v pravém horním rohu pod heslem „Learn“.

3. Data Stewardship Wizard

Šablonu je možné zvolit výběrem konkrétního dotazníku ze seznamu. Probíhá tak, že se zvolí požadovaný dotazník, který je obvykle read-only a tedy do něj není možné zapisovat rovnou. Dotazník se naklonuje a tato kopie je poté vyplňována. Jedná se tak o plán dle zvolené šablony. Výběr šablon může být obrovský, jelikož si je mohou vytvářet i sami uživatelé.

4. DataWiz

DataWiz bohužel neumožňuje zvolit šablonu, dle které by byl plán vytvářen. Nástroj je ale koncipován tak, že je možné plán vyplnit v souladu se třemi různými investory, a to program EU Horizon2020, BMBF (Spolkové ministerstvo pro vzdělání a výzkum v Německu) a DFG (Německá výzkumná společnost). Jednotlivé otázky mohou mít u sebe symbol jednoho či více z těchto investorů, který značí, že otázka je tímto investorem vyžadována. Téměř všechny otázky v tomto nástroji jsou nepovinné, a je tak možné při tvorbě plánu postupovat tak, aby vyhovoval potřebám konkrétního investora.

Nejedná se tedy o vlastní zvolení šablony pro generaci a tvorbu plánu, ale alespoň omezený výběr je uživateli umožněn.

5. easyDMP

V současné době je možné si zvolit ze tří šablon, a to Horizon 2020, Horizon 2020 Expert a Science Europe.

6. ezDMP

Jelikož nástroj ezDMP byl vytvořen pro potřeby nabídek NSF, neobsahuje žádné šablony.

7. QUT DMP

Nástroj obsahuje jedinou šablonu, QUT Data Management Plan, která je automaticky zvolena při vytváření plánu. Jelikož byl nástroj vytvořen pro potřeby Queensland University of Technology, neobsahuje žádné další šablony.

8. DMP OPIDoR

Na výběr je přibližně 30 šablon vyhovujícím potřebám investorů francouzské komunity. Investoři poskytují nápovědy a instrukce pro zodpovězení otázek.

4.1.2 Úprava šablony nebo vlastní šablona

Podobně jako u výběru šablony je tomu i zde. Někdy by uživatel využil možnost si danou šablonu mírně poupravit, případně si vytvořit zcela novou šablonu. Většina nástrojů toto neumožňuje, a možná i z toho důvodu pak vznikají specializované nástroje pro konkrétní potřeby, mající pouze jednu šablonu.

1. DMPOnline

Nástroj umožňuje vytvoření vlastní šablony, ale pouze v rámci předplatného. Úprava šablony či vytvoření nové patří v DMPOnline mezi pokročilejší funkce které jsou dostupné až po zaplacení.

2. **DMPTool**

Nástroj umožňuje vytvoření vlastní šablony uživatelům s admin právy. Administrátoři účtů institucí tedy mají práva k úpravě šablon či vytvoření vlastní.

3. **Data Stewardship Wizard**

Nástroj umožňuje upravovat šablony i vytvářet nové, a to přes editor znalostního modelu. Znalostní model může být vytvořen úplně od začátku nebo rozšířením či změnou stávajícího modelu. Vytváření modelu sestává z tvorby otázek a odpovědí, které mají mnoho nastavitelných funkcí a vlastností. Mezi ty patří například výběr fáze pro zodpovězení otázky a míra splnění FAIR metrik. S použitím tagů pro úroveň detailnosti dotazníku či možností předvyplnění plánu pro ostatní nabízí Data Stewardship Wizard skvělý způsob, jak vytvořit přehlednou šablonu přesně na míru, a to i v několika verzích.

4. **DataWiz**

DataWiz neumožňuje zvolit šablonu, a tedy ji nelze ani upravit či vytvořit vlastní.

5. **easyDMP**

Nástroj umožňuje vytvářet a upravovat šablony. Sám nástroj uvádí, že metoda na tvorbu je „very rough and ready“. Tedy, že tvorba je funkční, ale je poněkud surová. Může být zjednodušená, mít méně možností a funkcionalit. Nemusí tak poskytovat takové možnosti jako jiné nástroje. K tvorbě a úpravě šablon musí mít uživatel admin práva k šablonám, která může získat od superusera.

6. **ezDMP**

Nástroj byl vytvořen pro potřeby NSF a neumožňuje výběr šablony, a tedy ani její úpravu či vytvoření.

7. **QUT DMP**

Jelikož byl nástroj vytvořen pro potřeby Queensland University of Technology, obsahuje pouze jedinou šablonu bez možnosti její úpravy či vytvoření nové šablony.

8. **DMP OPIDoR**

Tvorba vlastních šablon je možná, ale s omezenými možnostmi. Pro tvorbu vlastních šablon je totiž potřeba mít administrátorský účet. Ten lze získat zdarma po zažádání, ale momentálně se uděluje pouze uživatelům z francouzské komunity, kteří si přejí vytvořit šablonu pro svou organizaci.

4.1.3 Jazyk nástroje

Většina nástrojů, minimálně všechny zkoumané v této práci, podporují angličtinu. Je ale výhodou umožnit přepnutí nástroje i do jiného jazyka. Angličtina nebude mezi podporovanými jazyky jmenována.

1. **DMPOnline**

Podporuje němčinu, španělštinu a francouzštinu.

2. **DMPTool**

Podporuje portugalsštinu.

3. **Data Stewardship Wizard**

Nepodporuje žádné další jazyky.

4. **DataWiz**

Podporuje němčinu, která je primárním jazykem.

5. **easyDMP**

Nepodporuje žádné další jazyky.

6. **ezDMP**

Nepodporuje žádné další jazyky.

7. **QUT DMP**

Nepodporuje žádné další jazyky.

8. **DMP OPIDoR**

Podporuje francouzštinu, která je primárním jazykem.

4.1.4 Možnosti exportu

Ve většině situací je vhodné dát uživateli na výběr možnosti exportu plánu. Ať už se jedná o estetickou stránku plánu, jakou je výběr fontu či formátu, nebo nějaké možnosti obsahu plánu. Typicky to bývá zobrazení nevyplněných otázek, nadpisů otázek či nějakých přidaných materiálů.

1. **DMPOnline**

Uživatel si zvolí, zda chce mít vypsaný i text otázek a nadpisy sekcí a zda mají být v souboru i nezodpovězené otázky. První stránku lze vyhradit shrnutí, které obsahuje základní informace jako jméno tvůrce a název šablony. V případě jiné než výchozí šablony si uživatel může zvolit, zda chce mít v plánu obsaženy i doplňkové sekce, které investorem nejsou vyžadovány. Na výběr je pět formátů (pdf, docx a další). U formátu pdf lze navolit další možnosti – velikosti jednotlivých okrajů dokumentu a font a velikost písma.

2. **DMPTool**

Nástroj nabízí stejné možnosti exportu jako DMPOnline. Zvolení obsahu plánů a výběr z pěti formátů s rozšířenými možnostmi exportu pro pdf.

3. **Data Stewardship Wizard**

Export se provádí v záložce Dokumenty. Buď se vybere existující dokument, nebo se vytvoří nový. Při vytváření se vyplní název a zvolí dotazník, šablona a formát. Na výběr je ze sedmi formátů. Po vytvoření je možné dokument stáhnout. Exportovaný soubor vypadá téměř stejně, jako prostředí nástroje, součástí jsou i informace o splnění metrik.

4. **DataWiz**

Nástroj umožňuje vybrat, co vše bude obsahem exportu. Kromě dokumentace projektu a plánu správy dat lze zvolit studie a jejich datasety. U studií a datasetů je na výběr z konkrétních částí, jako jsou popisy a datové matice. Exportovaný soubor je ve formátu zip a obsahuje jeden či více xml souborů v závislosti na vybraných studiích.

5. **easyDMP**

Nástroj umožňuje export pouze formou vygenerování html stránky bez dalších možností.

6. **ezDMP**

Nástroj umožňuje stáhnout plán ve formátu pdf bez dalších možností.

7. **QUT DMP**

Nástroj umožňuje exportovat plán ve čtyřech formátech: csv, html, pdf a xml. Další možnosti exportu nenabízí.

8. **DMP OPIDoR**

Nástroj nabízí velmi podobné možnosti exportu jako DMPOnline. Zvolení obsahu plánu s rozšířenými možnostmi exportu pro pdf, na výběr jsou zde ale tři formáty namísto pěti. Lze zvolit, které výstupy výzkumu mají být součástí a zda chce uživatel výstupy řadit dle sekcí či podle otázek.

4.1.5 Správa jednotlivcem či organizací

Jednotlivec a organizace mohou mít odlišné potřeby, a tedy i požadavky pro tvorbu plánu. Nástroj by měl umožňovat přidat spolupracovníky s nastavením různých přístupových práv. Většina nástrojů zajišťuje sdílení plánu právě přes přidání spolupracovníků s různými právy přístupu a editace.

1. DMPOnline

Nástroj umožňuje kromě klasického přihlášení i přihlášení přes instituci. U jednotlivých plánů umožňuje nastavit viditelnost plánu a přidávat spolupracovníky s různými právy.

V sekci Sdílení lze spravovat spolupracovníky, přidat nové, a nastavit jednu ze tří viditelností plánu. Soukromý plán vidí jen tvůrce plánu, přidání spolupracovníci a administrátoři. Dalšími možnostmi je plán viditelný pro organizaci, a veřejný plán vidí pro kohokoliv.

Spolupracovníky lze přidat pozváním přes email. Pokud již nástroj DMPOnline používají, uvidí daný plán v menu mezi svými plány. Pokud jej nepoužívají, obdrží email s pozvánkou. Při přidání je povinné vybrat práva spolupracovníka vybráním jedné z možností. Možnostmi je spoluvlastník, který má admin práva, poté editor, a nakonec práva jen pro čtení.

Předplatné ještě více rozšiřuje možnosti nástroje, a to převážně pro organizace. Mohou si tak přidat vlastní šablonu či vlastní logo.

2. DMPTool

Stejně jako v DMPOnline, DMPTool nabízí možnost klasického přihlášení a přihlášení přes instituci. Možnosti nastavení viditelnosti a přidání spolupracovníků jsou totožné jako v DMPOnline.

S admin právy umožňuje nástroj vytvořit vlastní šablonu, což využijí hlavně organizace.

3. Data Stewardship Wizard

Nástroj umožňuje sdílet dotazníky i znalostní modely. Dotazníky mohou mít různé úrovně přístupnosti – soukromou, veřejnou pouze pro čtení, a veřejnou.

Nástroj je přizpůsoben pro několik potřeb a rolí uživatelů. Demo s rolí data stewarda slouží pro vyzkoušení a poznání nástroje, není zamýšleno pro seriózní použití. Role výzkumníka nemá viditelný editor znalostního modelu, k použití jsou připraveny hotové znalostní modely poskytnuté nástrojem DS Wizard. Pro organizace je možnost vlastní instance distribuované přes Docker, kde se sama organizace stará o zřízení a správu, nebo možnost cloudu, kdy vlastní instanci spravuje Data Stewardship Wizard.

Díky možnosti vytvářet zcela nové dotazníky a upravovat je dle konkrétních potřeb je nástroj vhodný pro organizace. Ty mohou mít své požadavky na plány a šablony.

4. **DataWiz**

Je možné přidat nové spolupracovníky, a určit jim role a práva. Ty se přiřazují ke konkrétním studiím.

5. **easyDMP**

Nástroj umožňuje přidat spolupracovníky, a to s právem editace či pouze čtení. Tato funkcionality ale nebyla několik dní až týdnů v nástroji dostupná kvůli chybě.

6. **ezDMP**

Nástroj neumožňuje přidat spolupracovníky či plán jakkoliv sdílet, ke každému plánu tedy patří vždy přesně jeden uživatel. Jedinou možností je plán exportovat a odeslat.

7. **QUT DMP**

Nástroj umožňuje přidat spolupracovníky. Spolupracovník může mít práva spoluvlastníka, editora, či klasického uživatele, který může plán pouze číst.

8. **DMP OPIDoR**

Nástroj má opět velmi podobné možnosti jako DMPOnline a umožňuje přihlášení přes instituci. Obsahuje ale jednu možnost viditelnosti plánu navíc. Soukromá viditelnost DMPOnline, kdy plán vidí jen tvůrce plánu, přidání spolupracovníci a administrátoři, zde byla rozdělena na dvě možnosti. Její verzí v DMP OPIDoR je viditelnost administrátor. Soukromá zde totiž neumožňuje viditelnost administrátorům, pouze tvůrci plánu a přidáním spolupracovníkům.

Je možné vytvářet vlastní šablony, ale pouze po zažádání o administrátorský účet, který se uděluje uživatelům z francouzské komunity, kteří si přejí vytvořit šablonu pro svou organizaci.

4.1.6 Přidání dalších informací

Někdy může být vhodné přidat do plánu nějaké další doplňující informace. Těmi není myšleno přidání otázky a úprava šablony, ale například nahrání souboru.

1. **DMPOnline**

Nástroj se skládá z klasického dotazníku a neobsahuje možnost přidání dalších informací mimo odpovědi. V odpovědích ale lze vytvářet tabulky

4. ZHODNOCENÍ NÁSTROJŮ

a seznamy, což může umožnit vložení přehlednějších a detailnějších informací.

2. **DMPTool**

Nástroj se skládá z klasického dotazníku a neobsahuje možnost přidání dalších informací mimo odpovědí. V odpovědích ale lze vytvářet tabulky a seznamy, což může umožnit vložení přehlednějších a detailnějších informací.

3. **Data Stewardship Wizard**

Nástroj nabízí výhody zobrazení splnění FAIR metrik a odhad ceny úložiště. Zobrazeny jsou kromě FAIR i míry splnění tzv. good DMP practice – osvědčených postupů a otevřenost.

4. **DataWiz**

Nástroj umožňuje vytvářet studie a nahrávat materiály. Ty mohou být nahrány ke konkrétním studiím, nebo celkově k plánu. Studie mohou obsahovat datasety a codebooky.

5. **easyDMP**

Nástroj se skládá z klasického dotazníku a neobsahuje možnost přidání dalších informací mimo odpovědí.

6. **ezDMP**

Nástroj nabízí možnost zadat informace o konkrétních produktech, kterých může být několik druhů. Zadávání je formou navazujících otázek, a otázky tohoto typu obsahuje většina ostatních nástrojů, pouze nejsou takto oddělené. Informace o produktech se tedy nedají považovat za možnost přidání dalších informací.

7. **QUT DMP**

Nástroj se skládá z klasického dotazníku a neobsahuje možnost přidání dalších informací mimo odpovědí. V odpovědích ale lze možné vytvářet tabulky a seznamy, což může umožnit vložení přehlednějších a detailnějších informací.

8. **DMP OPIDoR**

Nástroj nabízí možnost zadat informace o výstupech projektu, kterých může být mnoho typů. Zadávání je formou navazujících otázek, a otázky tohoto typu obsahuje většina ostatních nástrojů, pouze nejsou takto oddělené. Informace o výstupech projektu se tedy nedají považovat za možnost přidání dalších informací. V odpovědích ale lze možné vytvářet tabulky a seznamy, což může umožnit vložení přehlednějších a detailnějších informací.

4.1.7 Registrace a přihlašování

Nástroj by měl umožňovat rychlou a pohodlnou registraci i přihlášení. S výjimkou doby ohlášené údržby by měl být nástroj dostupný. Výhodou je, pokud nástroj umožňuje lokální kopii a vlastní instanci nástroje. Jelikož jsou všechny nástroje zdarma a přístupné na githubu, je možné kód použít a nainstalovat si vlastní instanci nástroje. Některé nástroje toto podporují a poskytují návody, je možné se v případě potíží obrátit na podporu.

1. DMPOnline

Nástroj umožňuje přihlášení přes DMPOnline účet, nebo přihlášení přes instituci. Při registraci se zadává jméno, email, heslo a organizace. Pole organizace se zobrazuje jako nepovinné k vyplnění, ale bez jeho vyplnění nelze registraci dokončit. Po zadání údajů je uživatel zaregistrován a může nástroj rovnou používat.

2. DMPTool

Nástroj umožňuje přihlášení přes DMPTool účet, nebo přihlášení přes instituci. Při registraci se zadává jméno, email a heslo. Po zadání údajů je uživatel zaregistrován a může nástroj rovnou používat.

3. Data Stewardship Wizard

Nástroj umožňuje přihlášení přes DS Wizard účet, nebo přes Google či Elixir. Při registraci se zadává jméno, email, heslo a nepovinně organizace. Před použitím nástroje je nejdříve potřeba aktivovat účet odkazem v emailu.

Nástroj je přizpůsoben pro několik potřeb a rolí uživatelů. Demo s rolí data stewarda slouží pro vyzkoušení a poznání nástroje, není zamýšleno pro seriózní použití. Role výzkumníka nemá viditelný editor znalostního modelu, k použití jsou připraveny hotové znalostní modely poskytnuté nástrojem DS Wizard. Pro organizace je možnost vlastní instance distribuované přes Docker, kde se sama organizace stará o zřízení a správu, nebo možnost cloudu, kdy vlastní instanci spravuje Data Stewardship Wizard.

4. DataWiz

Nástroj umožňuje pouze přihlášení přes DataWiz účet. Při registraci se zadává jméno, email a heslo. Před použitím nástroje je nejdříve potřeba aktivovat účet odkazem v emailu.

5. easyDMP

Nástroj umožňuje přihlášení přes B2ACCESS nebo Dataporten. Nevytváří se účet jako u jiných nástrojů, ale přihlašuje se přes jeden z již

4. ZHODNOCENÍ NÁSTROJŮ

existujících účtů Google, Microsoft, Facebook, Twitter a další. Před použitím nástroje je nejdříve potřeba aktivovat účet odkazem v emailu.

Nástroj je bohužel poměrně často nedostupný. Alespoň dle osobní zkušenosti. Během analýzy nástroje v období od 28. 3. 2020 do 28. 4. 2020 byl nedostupný tři týdny.

6. ezDMP

Nástroj umožňuje přihlášení přes Google nebo OrcID. Není tedy nutné (a ani možné) se registrovat a aktivovat účet emailem.

7. QUT DMP

Nástroj umožňuje přihlášení přes účet QUT nebo externí přihlášení. Při registraci se zadává pouze email a heslo. Před použitím nástroje je nejdříve potřeba aktivovat účet odkazem v emailu. Email přišel později než u jiných nástrojů.

8. DMP OPIDoR

Nástroj umožňuje přihlášení přes DMP OPIDoR účet, nebo přihlášení přes instituci. Při registraci se zadává jméno, email, organizace a heslo. Po zadání údajů je uživatel zaregistrován a může nástroj rovnou používat.

Tabulka 4.1: Shrnutí možností přizpůsobení. Výběr šablony, úprava šablony, jazyk nástroje (kromě angličtiny), možnost exportu (množství parametrů a formátů) a přidání dalších informací.

Nástroj	Výběr	Úprava	Jazyk	Export	Přidání
DMPonline	ano	freemium	de, es, fr	vysoká	ne
DMPTool	ano	domluva	pt	vysoká	ne
DS Wizard	ano	ano	-	vysoká	ano
DataWiz	částečně	ne	de	střední	ano
easyDMP	ano	domluva	-	nízká	ne
ezDMP	ne	ne	-	ne	ne
QUT DMP	ne	ne	-	střední	ne
DMP OPIDoR	ano	domluva	fr	vysoká	ne

4.2 Machine-actionability

V této sekci je hodnocena míra machine-actionable. Které funkce podporují machine-actionability a které jsou naopak jeho porušením. Hodnocení vychází z bodů v kapitole 2.3 Machine-actionable DMP.

Machine-actionable plán správy dat by tak dle těchto bodů měl mít například tyto vlastnosti:

- Počet otázek formou otevřeného textu by měl být co nejnižší, odpovědi by měly být spíše formou výběru z možností
- Jasně popsaná metadata s možností pohodlného vyhledání zdroje.
- Plán by měl být přizpůsoben postupnému vyplňování otázek a změnám v průběhu času.
- Na plánu by mělo spolupracovat více osob.

Tato kategorie nebude hodnotit nástroje dle konkrétních bodů jako ostatní kategorie, ale budou jmenovány funkcionality podporující vizi machine-actionable plánu, a i funkcionality, které této vizi odporují.

1. DMPOnline

Nástroj podporuje spolupráci přidáním spolupracovníků a při výběru organizací doplňuje odpovídající názvy. I přesto, že se obsah a rozdělení plánu velmi liší v závislosti na zvolené šabloně, naprostá většina odpovědí je formou otevřeného textu, což odporuje machine-actionable vizi. Odpovědi tak mohou být nejasné, neúplné a není možné s nimi dále pracovat bez účasti člověka.

2. DMPTool

Stejně jako DMPOnline podporuje spolupráci přidáním spolupracovníků a při výběru organizací doplňuje odpovídající názvy. Obsah a rozdělení plánu se velmi liší v závislosti na zvolené šabloně, ale naprostá většina odpovědí je formou otevřeného textu, což odporuje machine-actionable vizi.

3. Data Stewardship Wizard

Nástroj podporuje spolupráci přidáním spolupracovníků, a navíc umožňuje dotazníky předvyplnit pro ostatní uživatele. Díky fázím, ve kterých se projekt může nacházet, se podporuje návrat k plánu. Uživatelé nemohou mít všechny informace na začátku, a proto bývají plány a jeho otázky rozděleny do třech fází – před odesláním žádosti, před odesláním plánu a před dokončením projektu.

Velkou výhodou podporující machine-actionability je velice nízký počet odpovědí formou otevřeného textu. Většina odpovědí se zadává výběrem

4. ZHODNOCENÍ NÁSTROJŮ

z možností. Tím je zaručeno, že bude zodpovězeno vše potřebné a přesně. S tím souvisí i FAIR metriky, jelikož každá odpověď může mít přiřazenou jednu z metrik, váhu a míru. Na základě odpovědí tak lze odhadnout, nakolik plán splňuje FAIR principy.

Nástroj při zodpovídání některých otázek napovídá a zobrazuje vyhovující varianty, například při zadávání zdrojů informací odkud projekt čerpá. Nástroj nejen doplňuje odpovídající názvy, ale i stránky kde lze najít další informace. To zajišťuje propojení s FAIRsharing.org a zaručuje přehledná metadata.

V Data Stewardship Wizard lze také využít možnosti odhadu ceny potřebného úložiště, jehož parametry mohou být nastaveny ve třech úrovních detailů. Je viditelné, že při vývoji nástroje byl kladen důraz na body FAIR a vizi machine-actionable.

4. DataWiz

Nástroj podporuje spolupráci přidáním spolupracovníků a umožňuje přiřadit práva ke konkrétním studiím. Vytváření a spravování studií se může považovat za machine-actionable prvek, jelikož i díky němu plán podporuje k průběžnému vracení uživatele. Použití studií tak může zamezit tomu, aby byl plán použit pouze jednorázově a zůstal statickým dále nepoužívaným dokumentem. Ve studiích jsou zároveň symboly, rozdělující otázky do dvou fází vyplnění, podobně jako v nástroji Data Stewardship Wizard, ale jednodušeji. Opět prvek podporující návrat uživatele a práci na plánu.

Plán nenabízí šablony, pouze označuje, které otázky jsou vyžadovány některým za tří investorů. I přesto, že otázky jsou tedy neměnné a vždy stejné, je poměrně vysoký počet otázek zodpovídaných formou otevřeného textu. V nástroji je ale viditelná snaha počet těchto odpovědí snížit a nabídnout odpovědi formou výběru.

5. easyDMP

Nástroj podporuje spolupráci přidáním spolupracovníků. Míra machine-actionability velmi závisí na zvolené šabloně. Šablony Horizon 2020 Expert a Science Europe obsahují pouze otázky formou otevřeného textu, což je s machine-actionable vizí v rozporu. Pouze šablona Horizon 2020 je vizi blíže – podstatná většina otázek je formou výběru, pouze případné komentáře jsou formou otevřeného textu. Nástroj zde u některých otázek i napovídá při psaní odpovědi, hlavně u otázek na vyjmenování použitých formátů a standardů.

6. ezDMP

Jako jediný ze zkoumaných nástrojů nepodporuje spolupráci a není možné přidat spolupracovníky ani plán sdílet. Jedinou možností je plán expor-

tovat. Plán není rozsáhlý, obsahuje málo otázek, ale i přesto je jich mnoho formou otevřeného textu. I vzhledem k nemožnosti přidání spolupracovníků plán příliš nepodporuje vracení. Nástroj ezDMP je tedy poměrně v rozporu s machine-actionable.

7. QUT DMP

Stejně jako DMPOnline podporuje spolupráci přidáním spolupracovníků. I přesto, že plán obsahuje jedinou šablonu, je naprostá většina odpovědí formou otevřeného textu, což odporuje machine-actionable vizi. Součástí je i otázka, kdy bude další review plánu, a doporučení, že by mělo být alespoň jednou za rok. Nástroj tak podporuje k vracení uživatele.

8. DMP OPIDoR

Nástroj podporuje spolupráci přidáním spolupracovníků a při výběru organizace doplňuje odpovídající názvy. I přesto, že se obsah a rozdělení plánu velmi liší v závislosti na zvolené šabloně, naprostá většina odpovědí je formou otevřeného textu, což odporuje machine-actionable vizi.

4.3 Zhodnocení UI/UX dle desatera Jakoba Nielsena

Jakob Nielsen je odborník na webový design a použitelnost. Drží i spoustu patentů, převážně na způsoby zjednodušení použití internetu. Bývá označován jako světový odborník na webovou použitelnost. Vymyslel i metodu heuristického vyhodnocení, která je použita v této práci. Ta umožňuje posoudit design uživatelského rozhraní na základě deseti ověřených principů pro komunikaci mezi člověkem a počítačem. Tyto informace i pravidla vychází z [30].

Nielsenovo desatero se skládá z pravidel:

1. Viditelnost stavu systému
2. Shoda mezi systémem a realitou
3. Uživatelská svoboda
4. Konzistence a standardy
5. Prevence chyb
6. Rozpoznání před vzpomnutím
7. Flexibilita a efektivita použití
8. Estetický a minimalistický design
9. Pomoc s pochopením chyb a jejich vyřešením
10. Pomoc a dokumentace

4.3.1 Viditelnost stavu systému

Uživatel by měl být vždy informován o tom, co se v systému právě děje, a to vhodnou zpětnou vazbou a v rozumném čase.

1. **DMPOnline**

Nástroj je přehledný a je jasné vidět, která sekce či otázka je momentálně vybrána.

2. **DMPTool**

Nástroj je přehledný a je jasné vidět, která sekce či otázka je momentálně vybrána.

3. **Data Stewardship Wizard**

Nástroj přehledně zobrazuje, která sekce, model či dotazník je vybrán.

4. **DataWiz**

Nástroj by mohl být trochu lépe uzpůsoben pro vyšší přehlednost. Převážně v sekci Plánování správy dat je více sekcí a skupin, které se ztrácejí mezi textem. Uživatel tak může chvíli trvat, než se zorientuje, která sekce či skupina je momentálně vybrána. Při práci na studii je mnoho zanořování, ve kterém může být těžší se zorientovat.

5. **easyDMP**

Nástroj je pomalejší a nemusí tak být jasné, co se v systému právě děje. Problematická je i orientace v otázkách, jelikož otázky je nutné zodpovídat v daném pořadí. Bez zodpovězení současné otázky se uživatel nedostane k následující ani předchozí, ale není mu vždy jasné oznámeno. Práce s tlačítky další a předchozí je tedy nepohodlná a matoucí. Součástí možností plánu je i tlačítko „Check“, ověřující validitu plánu. Uživatel se ale nezobrazí žádný závěr, jež by z kontroly vyplynul, zobrazí se hlavní obrazovka plánu – přehled plánu.

6. **ezDMP**

Nástroj je jednoduchý, neobsahuje dlouhé nápovědy, tvorba plánu probíhá na jediné stránce (při přidání produktu na dvou stránkách). Uživatel se nedostává do situace, kdy by nevěděl, co se v systému děje, či na něco čekal.

7. **QUT DMP**

Nástroj je poměrně pomalý a uživatel tak někdy musí čekat, než se akce provede. V některých momentech tak nemusí být jasné, co se v nástroji děje, zda se stránka načítá, či tlačítko nezareagovalo.

8. DMP OPIDoR

Nástroj je přehledný a je jasné vidět, která sekce či otázka je momentálně vybrána.

4.3.2 Shoda mezi systémem a realitou

Systém by měl používat jazyk, slova a fráze srozumitelné uživateli, a nikoliv systémově zaměřené výrazy.

1. DMPOnline

Nástroj nepoužívá systémově zaměřené výrazy, kterým by uživatel nemusel rozumět. Téměř všechny otázky jsou doplněny nápovědou, popisující, jak otázky vhodně a jasně zodpovědět a vysvětlující zmíněné pojmy. Nápověda není psána odborně a obsahuje mnoho příkladů.

2. DMPTool

Nástroj nepoužívá systémově zaměřené výrazy, kterým by uživatel nemusel rozumět. Stejně jako u DMPOnline, jsou téměř všechny otázky doplněny vhodnou nápovědou vysvětlující zmíněné pojmy. Nápověda není psána odborně a obsahuje mnoho příkladů.

3. Data Stewardship Wizard

Vzhledem k možnosti vytvářet vlastní dotazníky a vlastní znalostní modely obsahuje nástroj více výrazů, kterým by uživatel nemusel rozumět. Pojmenování funkcionalit a prvků je od ostatních nástrojů velmi odlišné. Pro neznalého uživatele tak může být náročnější i provedení základních akcí, které zná dobře z jiných nástrojů.

Je ovšem možné, že při skutečném použití se tento problém vyskytne v menším měřítku, či se nevyskytne vůbec. Výzkumníci totiž nemají přístup k editoru znalostního modelu a spousta pokročilých funkcí ve kterých by se nemuseli plně orientovat je tak pro ně skryta.

4. DataWiz

Nástroj je srozumitelný, pouze některé části studií obsahují odborněji zaměřené fráze. Otázky jsou doplněny nápovědami.

5. easyDMP

Nástroj nepoužívá systémově zaměřené výrazy, kterým by uživatel nemusel rozumět. Stejně jako u DMPOnline, jsou téměř všechny otázky doplněny vhodnou nápovědou vysvětlující zmíněné pojmy. Klasická nápověda jako v ostatních nástrojích zde není, ale otázky v některých šablonách mohou být detailnější, a sloužit tak jako nápověda.

6. ezDMP

Nástroj neobsahuje systémově zaměřené výrazy, pouze výrazy související s plánováním správy dat. Náповědy jsou stručné, zobrazované v polích pro odpovědi.

7. QUT DMP

Nástroj nepoužívá systémově zaměřené výrazy, kterým by uživatel nemusel rozumět. Všechny otázky jsou doplněny náповědou, obsahující pomoc se zodpovězením otázky a příklady. Jelikož nástroj obsahuje jedinou šablonu, náповěda obsahuje i kontakty na konkrétní osoby či informace zaměřené na různé skupiny studentů.

8. DMP OPIDoR

Nástroj nepoužívá systémově zaměřené výrazy, kterým by uživatel nemusel rozumět. Stejně jako u DMPOnline, jsou téměř všechny otázky doplněny vhodnou náповědou vysvětlující zmíněné pojmy. Náповěda není tak rozsáhlá jako u DMPOnline a DMPTool.

4.3.3 Uživatelská svoboda

Uživatelé často použijí nějakou funkci omylem a potřebují jasně viditelnou možnost, jak se vrátit do stavu před touto akcí, aniž by museli procházet mnoha dialogy. Vhodná jsou tlačítka zpět a vpřed.

1. DMPOnline

Nástroj neobsahuje situace, kdy by uživatel váhal, jak svou akci vrátit. Text lze po uložení pohodlně smazat, komentář odstranit. Po uložení neprobíhá přesměrování na jinou stránku, a uživatel se tak nepotýká s problémy vrátit se na stránku, ze které by chtěl akci vracet do předchozího stavu.

2. DMPTool

Nástroj neobsahuje situace, kdy by uživatel váhal, jak svou akci vrátit. Postupy jsou stejné jako u DMPOnline. Text lze po uložení pohodlně smazat, komentář odstranit. Po uložení neprobíhá přesměrování na jinou stránku, a uživatel se tak nepotýká s problémy vrátit se na stránku, ze které by chtěl akci vracet do předchozího stavu.

3. Data Stewardship Wizard

Nástroj obsahuje u každé otázky tlačítko pro smazání odpovědi. Prvky v editoru jsou také viditelně označené. Při otevření modelu či dotazníku je stále viditelná možnost uložení či zahození změn.

4. **DataWiz**

Text v odpovědi lze po napsání pohodlně smazat, ukládá se vždy celá stránka otázek. Pravděpodobně jediným problémem by zde mohla být práce na studii, kde je mnoho zanořování, ve kterém může být těžší se zorientovat. Uživatel zde případně může použít šipku zpět v prohlížeči.

5. **easyDMP**

Nutnost zodpovídat otázky v daném pořadí je velmi problematické, jelikož uživateli není bez zodpovězení současné otázky umožněno přejít k následující ani předchozí otázce. Ne vždy je to ale jasně oznámeno a nástroj je v tomto ohledu matoucí. Uživatel může smazat text odpovědi, ale bez opětovného vyplnění bude mít velice ztížený přechod k dalším otázkám. Pokud vyplní otázku s formou povinného výběru a chce odpověď smazat, není mu to umožněno, a je nucen ponechat otázku vyplněnou onou nežádoucí odpovědí.

6. **ezDMP**

Text v odpovědi lze po napsání pohodlně smazat, ukládá se vždy celá stránka otázek. Produkty i jejich vazby mají jasně označeny tlačítka pro odstranění.

7. **QUT DMP**

Pokud nebyly vyplněny všechny povinné otázky z dané sekce, při změně či zavření sekce se zobrazí varování, že všechny otázky nebyly zodpovězeny. Na výběr je se k otázkám vrátit a vyplnit je, či je zatím ponechat nevyplněné. To může být nepohodlné, pokud uživatel vybere jinou sekci, než chtěl, nebo si chce jednotlivé sekce pouze prohlédnout. Tyto dialogy tedy v některých případech mohou být příliš časté a obtěžující.

8. **DMP OPIDoR**

Nástroj neobsahuje situace, kdy by uživatel váhal, jak svou akci vrátit. Postupy jsou stejné jako u DMPOnline. Text lze po uložení pohodlně smazat, komentář odstranit. Po uložení neprobíhá přesměrování na jinou stránku, a uživatel se tak nepotýká s problémy vrátit se na stránku, ze které by chtěl akci vracet do předchozího stavu.

4.3.4 Konzistence a standardy

Měly by být následovány konvence platformy. Neměl by nastat moment, kdy uživatel váhá, zda odlišné pojmenování znamená opravdu něco jiného nebo stejnou věc.

1. DMPOnline

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný není pro uživatele náročný.

2. DMPTool

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný není pro uživatele náročný. DMPOnline a DMPTool jsou téměř totožné.

3. Data Stewardship Wizard

Nástroj používá na všech místech stejná pojmenování, jsou ale velmi odlišná od pojmenování ostatních nástrojů. Přechod tak může být složitější, jelikož většina funkcí a postupů se zde liší od ostatních nástrojů, které povětšinou sdílí alespoň základní myšlenku rozdělení nástroje a vytvoření plánu.

Je ovšem možné, že při skutečném použití se tento problém vyskytne v menším měřítku, či se nevyskytne vůbec. Výzkumníci nemají přístup k editoru znalostního modelu a spousta pokročilých funkcí ve kterých by se nemuseli plně orientovat je tak pro ně skryta.

4. DataWiz

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný by pro uživatele nebyl příliš náročný. DataWiz má ale v některých částech poněkud nepřehlednou strukturu a na rozdíl od ostatních nástrojů má možnost vytvořit a spravovat studie či nahrávat soubory, což může přechod zkomplikovat.

5. easyDMP

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný není pro uživatele náročný. Nepříjemná ale může být poněkud pomalejší odezva, a hlavně nucené pořadí otázek.

6. ezDMP

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou poněkud odlišná od ostatních nástrojů, jelikož ezDMP je vytvořeno pro konkrétní potřeby a obsahuje tak zcela jiné otázky.

7. QUT DMP

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný není pro uživatele, co se týče pojmenování a konvencí, náročný. Horší může být přechod v ohledu odezvy systému, jediné šablony, či častých dialogových oken.

8. DMP OPIDoR

Nástroj používá na všech místech stejná pojmenování. Pojmenování jsou shodná či velmi podobná jako u ostatních nástrojů, tedy ani přechod z jednoho nástroje na jiný není pro uživatele náročný.

4.3.5 Prevence chyb

Mělo by být zabráněno, aby problém způsobující chybu nastal. Podmínky náchylné k chybám by měly být kontrolovány, podmínky ošetřeny a uživatel by měl akci potvrdit.

1. DMPOnline

Nástroj není náchylný k chybám a případné nevyplnění povinného pole oznámí chybovou hláškou.

2. DMPTool

Nástroj není náchylný k chybám a případné nevyplnění povinného pole oznámí chybovou hláškou.

3. Data Stewardship Wizard

Nástroj neobsahuje povinné otázky a je tak možné plán uložit i bez vyplnění otázek. Nebyl nalezen postup, který by způsobil chybu.

4. DataWiz

Povinná pole nástroji nejsou vždy označena. Při nevyplnění nedovolí uložení, ale chyba není příliš viditelná, a ne vždy se zobrazuje stejně. Vyžadované pole má pouze červeně zbarvené okraje a text chyby se zobrazí až po najetí kurzorem na pole. Jindy se zobrazí chyba, že otázka musí být zodpovězena, ale vyžadované pole není červeně označeno. Na jiném místě se zobrazí text chyby i červeně označené pole.

5. **easyDMP**

Bez zodpovězení současné otázky se uživatel nedostane k následující ani předchozí, ale není mu to vždy jasně oznámeno. Někdy se zobrazí přehledný a srozumitelný dialog chyby, někdy je to jen řádka textu, která je velikostí, fontem i barvou shodná s otázkou a je umístěna uprostřed textu. Uživatel si jí tedy ani nemusí všimnout.

V průběhu analýzy byly nalezeny další chyby, kdy nastala chyba při přechodu k určité otázce či při pozvání spolupracovníků. Tyto chyby byly již odstraněny. V období kdy byl nástroj nedostupný nebylo jasné, proč tomu tak je a kdy bude opět funkční. Chyb je tedy bohužel poměrně mnoho a často i špatně popsanych.

6. **ezDMP**

Nástroj neobsahuje povinné otázky a je tak možné plán uložit i bez vyplnění otázek. Nebyl nalezen postup, který by způsobil chybu.

7. **QUT DMP**

Pokud nebyly vyplněny všechny povinné otázky z dané sekce, zobrazí se varování, že všechny otázky nebyly zodpovězeny. Na výběr je se k otázkám vrátit a vyplnit je, či je zatím ponechat nevyplněné.

8. **DMP OPIDoR**

Nástroj není náchylný k chybám a případné nevyplnění povinného pole oznámí chybovou hláškou.

4.3.6 Rozpoznání před vzpomenutím

Všechny důležité informace, akce a objekty by měly být viditelné, uživatel by neměl být nucen si informace pamatovat. Instrukce k použití systému by měly být viditelné či snadno nalezitelné.

1. **DMPOnline**

Nástroj umožňuje kdykoliv změnit vybranou sekci, otázky jsou rozděleny do skupin, které lze rozbalit. U každé otázky je nápověda a možnost komentáře, v místě odpovědi je stále přítomný řádek pro možnost formátování textu. V horní liště, kde je mimo jiné účet uživatele, je dostupný návod k použití nástroje.

2. **DMPTool**

Nástroj umožňuje kdykoliv změnit vybranou sekci, otázky jsou rozděleny do skupin, které lze rozbalit. U každé otázky je nápověda a možnost komentáře, v místě odpovědi je stále přítomný řádek pro možnost formátování textu. Vedle záložky účtu uživatele je dostupný návod k použití nástroje.

3. Data Stewardship Wizard

Nástroj umožňuje kdykoliv změnit vybranou kapitolu otázek dle seznamu kapitol. U každé otázky je stručná nápověda a případný odkaz na podrobnou nápovědu. Nápověda k použití nástroje ale není dostupná a viditelná přímo z nástroje.

4. DataWiz

Změna sekce či skupiny otázek může být někdy složitější kvůli horší přehlednosti. K většině otázek je poskytnuta nápověda, a to pod polem pro odpověď. Text nápovědy je ale často dlouhý a odborný. Jelikož nejde skrýt a není umístěn bokem jako například v DMPOnline, může také zhoršovat orientaci v nástroji. Přes odkazy v nápovědách se lze dostat k rozsáhlejší nápovědám.

Nástroj obsahuje několik symbolů pro investory, fáze a nápovědu, které vzbouzí pocit, že po najetí kurzorem poskytují nějaký text s významem symbolu. Žádné informace ale neposkytují, což může být matoucí a pokud si uživatel nepamätuje význam symbolů, musí se podívat do hlavičky.

5. easyDMP

Otázky na jednotlivé otázky nejsou stále tak dobře a přehledně viditelné, jako u ostatních nástrojů. Ani ostatní informace, jako například spolupracovníci, nejsou tak rychle přístupné. V horní liště je dobře viditelná záložka pomoci, odkud lze zjistit více informací o jednotlivých šablonách, přečíst si návod k nástroji, či napsat dotaz na podporu nástroje.

6. ezDMP

Nástroj je velmi jednoduchý a obsahuje málo otázek. Nápovědy jsou stručné a v polích pro odpovědi. Všechny informace jsou tedy dobře viditelné.

7. QUT DMP

Nástroj umožňuje po projití dialogem změnit vybranou sekci, otázky jsou rozděleny do skupin, které lze rozbalit. U každé otázky je nápověda, v místě odpovědi je stále přítomný řádek pro možnost formátování textu. Nástroj je přehledný, ale kvůli pomalé odezvě systému a častým dialogům může uživatele nutit k pamatování si informací. Může například odradit od toho, aby se uživatel vrácením k předchozí sekci ujistil, jak otázku vyplnil.

8. DMP OPIDoR

Stejně jako u DMPOnline či DMPTool nástroj umožňuje kdykoliv změnit vybranou sekci a otázky jsou rozděleny do skupin, které lze rozbalit.

U každé otázky je nápověda a možnost komentáře, v místě odpovědi je stále přítomný řádek pro možnost formátování textu. V horní liště, kde je mimo jiné účet uživatele, je dostupný návod k použití nástroje.

4.3.7 Flexibilita a efektivita použití

Pokud uživatel již systém zná, mělo by být možné průchody systémem urychlit. To je možné pomocí klávesových zkratk nebo třeba nabízení dříve zadaných možností při vyplňování formuláře.

1. **DMPOnline**

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit rozsáhlé odpovědi formou otevřeného textu, což nelze urychlit. Samotná práce s nástrojem je rychlá a jednoduchá.

2. **DMPTool**

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit rozsáhlé odpovědi formou otevřeného textu, což nelze urychlit. Samotná práce s nástrojem je rychlá a jednoduchá.

3. **Data Stewardship Wizard**

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit mnoho otázek. Práce s nástrojem je rychlá, ale kvůli velkému množství možností může být pomalá. Pokud již uživatel nástroj zná, může s nástrojem pracovat rychleji. Převážně pokud se jedná o data stewarda, který se stará o znalostní modely a vytváří nové.

4. **DataWiz**

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit rozsáhlé odpovědi formou otevřeného textu, což nelze urychlit. Pokud uživatel nástroj již zná, bude se přehledněji orientovat v méně přehledných sekcích, ale nástroj samotný žádné urychlení nenabízí.

5. **easyDMP**

V jistých případech je možné průchod poněkud urychlit. Nástroj totiž obsahuje dvě šablony Horizon 2020. První je určena pro uživatele kteří s plánováním správy dat nemají příliš zkušeností, a obsahuje tak spoustu otázek s vysvětlením. Druhá šablona je stručná, obsahuje jen několik polí pro odpovědi formou otevřeného textu, kam by měl uživatel zapsat veškeré informace. Ty už by měl znát z předchozí detailnější šablony.

6. ezDMP

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit odpovědi, což nelze urychlit. Samotná práce s nástrojem je rychlá a jednoduchá.

7. QUT DMP

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit rozsáhlé odpovědi formou otevřeného textu, což nelze urychlit. Samotná práce s nástrojem je jednoduchá, ale poněkud zdlouhavá.

8. DMP OPIDoR

Není znám žádný způsob zefektivnění, jelikož se jedná o nástroj, kde je potřeba vyplnit rozsáhlé odpovědi formou otevřeného textu, což nelze urychlit. Samotná práce s nástrojem je rychlá a jednoduchá.

4.3.8 Estetický a minimalistický design

Dialogy by neměly obsahovat informace, které nejsou relevantní nebo jsou potřebné velmi zřídka. Každá informace navíc snižuje viditelnost ostatních důležitých informací.

1. DMPOnline

Nástroj je jednoduchý a přehledný, sekce a skupiny jsou zřetelně označeny. Nápowěda ke každá otázce je zobrazena vedle pole pro odpověď. Nebývá zobrazována celá, pouze její hlavní část. Další nápowědu či podrobnosti si uživatel zobrazí pouze pokud bude mít zájem. Nápowědy v některých sekcích se zobrazí až po najetí kurzorem na pole, aby uživatele nezahlcovaly.

2. DMPTool

Nástroj je jednoduchý a přehledný, sekce a skupiny jsou zřetelně označeny. Nápowěda ke každá otázce je zobrazena vedle pole pro odpověď. Nebývá zobrazována celá, pouze její hlavní část. Další nápowědu či podrobnosti si uživatel zobrazí pouze pokud bude mít zájem. Nápowědy v některých sekcích se zobrazí až po najetí kurzorem na pole, aby uživatele nezahlcovaly.

3. Data Stewardship Wizard

Nástroj je přehledný, používá přiměřené množství textu s možností zobrazení podrobnějších nápowěd přes odkaz. Skupiny otázek jsou dobře viditelné, dobrá čitelnost a viditelnost důležitých informací u otázek je zajištěna i vhodným použitím barev.

4. DataWiz

Nástroj není příliš přehledný. Převážně v sekci Plánování správy dat je více sekcí a skupin, které se ztrácejí mezi textem. Uživatel tak může chvíli trvat, než se zorientuje, která sekce či skupina je momentálně vybrána. Při práci na studii je mnoho zanořování, ve kterém může být těžší se zorientovat. Text nápovědy je často dlouhý. Jelikož nejde skrýt a není umístěn bokem jako například v DMPOnline, může zhoršovat orientaci v nástroji.

Plán tak obsahuje příliš otázek i textu pohromadě a stává se nepřehledným. Důvodem může být i příliš malá velikost písma a snaha zobrazit co nejvíce informací.

5. easyDMP

Nástroj neobsahuje příliš textu a je přehledný. Jelikož je vždy zobrazena pouze jedna otázka, uživatel není zahlcen a design je tak opravdu minimalistický.

6. ezDMP

Nástroj je jednoduchý a přehledný, i proto, že je velice stručný a nápovědy taktéž. Nástroj tak obsahuje minimum textu. Odpovědi formou výběru jsou barevně odlišeny od odpovědí formou otevřeného textu.

7. QUT DMP

Nástroj je jednoduchý a přehledný, sekce a skupiny jsou zřetelně označeny. Nástroj nemá tolik funkcionalit a možností jako například DMPOnline, a je tak ještě minimalističtější. Nápověda ke každé otázce je zobrazena vedle pole pro odpověď.

8. DMP OPIDoR

Nástroj je jednoduchý, sekce a skupiny jsou zřetelně označeny. Nápověda ke každé otázce je zobrazena vedle pole pro odpověď. Nápovědy v některých sekcích se zobrazí až po najetí kurzorem na pole, aby uživatele nezahlcovaly.

4.3.9 Pomoc s pochopením chyb a jejich vyřešením

Chyby by měly být napsány prostým textem a nikoliv kódem. Měly by vysvětlit, kde nastal problém, a navrhnout řešení.

1. DMPOnline

Nástroj téměř neumožňuje způsobit chybu, jelikož naprostá většina otázek je nepovinná. Pokud není vyplněna povinná otázka a nelze tak pokračovat, je to uživateli srozumitelně oznámeno. Text chyby ale není příliš výrazný, jedná se o světle žluté pole na bílém a oranžovém pozadí.

2. **DMPTool**

Nástroj téměř neumožňuje způsobit chybu, jelikož naprostá většina otázek je nepovinná. Pokud není vyplněna povinná otázka a nelze tak pokračovat, je to uživateli jasně a srozumitelně oznámeno. Text chyby je vidět lépe než v DMPOnline, jelikož pole chyby je červené.

3. **Data Stewardship Wizard**

Nástroj neobsahuje povinné otázky a je tak možné plán uložit i bez vyplnění otázek. Nebyl nalezen postup, který by způsobil chybu.

4. **DataWiz**

Jak bylo uvedeno v dřívějším bodu, zobrazení chyb není jednotné a jasné. Někdy není zobrazeno, kde problém nastal, v jiném případě chybí informace o tom, jaký problém nastal a jak ho opravit.

5. **easyDMP**

Jak již bylo zmíněno, bez zodpovězení současné otázky se uživatel nedostane k následující ani předchozí, ale není mu to vždy jasně oznámeno. Někdy se zobrazí přehledný a srozumitelný dialog chyby, někdy je to jen řádka textu, která je velikostí, fontem i barvou shodná s otázkou a je umístěna uprostřed textu. Uživatel si jí tedy nemusí všimnout.

V průběhu analýzy byly nalezeny další problémy, kdy nastala chyba při přechodu k určité otázce či při pozvání spolupracovníků. Tyto akce vedly k opuštění nástroje a zobrazení bílé obrazovky s informací, že nastala chyba. Ta ale nebyla nijak blíže specifikovaná a nebylo možné se jí vyhnout. Tyto chyby byly již odstraněny. V období kdy byl nástroj nedostupný nebylo taktéž jasné, proč tomu tak je a kdy bude opět funkční.

Chyb je tedy bohužel poměrně mnoho a často i špatně popsanych.

6. **ezDMP**

Nástroj neobsahuje povinné otázky a je tak možné plán uložit i bez vyplnění otázek. Nebyl nalezen postup, který by způsobil chybu.

7. **QUT DMP**

Pokud nebyly vyplněny všechny povinné otázky z dané sekce, při změně či zavření sekce se zobrazí varování, že všechny otázky nebyly zodpovězeny. Součástí dialogu je také seznam nezodpovězených otázek. Nástroj také uživatele při opuštění sekce upozorňuje na vyplněné ale neuložené otázky, které by tak mohly být ztraceny.

8. **DMP OPIDoR**

Nástroj téměř neumožňuje způsobit chybu, jelikož naprostá většina otázek je nepovinná. Pokud není vyplněna povinná otázka a nelze tak pokračovat, zobrazí se chybová hláška. Ta je poněkud hůře viditelná, jelikož

se jedná o světle žluté pole na bílém pozadí. Pokud ale uživatel neovládá francouzštinu, tak se hůře dozví, kde chyba vznikla. Tento text je totiž francouzsky, i když je zvolená angličtina. Nezodpovězená otázka není barevně označena.

4.3.10 Pomoc a dokumentace

Dokumentace by měla být poskytnuta, i když by systém mělo být možné použít i bez ní. Jakákoliv informace týkající se pomoci by měla být snadná na nalezení, ne příliš rozsáhlá a s vypsány hlavními body.

1. DMPOnline

Nástroj je jednoduchý na použití a lze ho tedy využít bez pomoci a dokumentace. Ta je dostupná v horní liště, kde je mimo jiné účet uživatele. Kromě návodu k použití nástroje a zodpovězení otázek obsahuje i seznam šablon a časté dotazy.

U otázek jsou uvedeny nápovědy. Nástroj nabízí podporu, na kterou se lze obrátit.

2. DMPTool

Nástroj je jednoduchý na použití a lze ho tedy využít bez pomoci a dokumentace. Ta je případně dostupná ze záložky Learn, vedle uživatelského účtu. Kromě návodu k použití nástroje a zodpovězení otázek obsahuje i seznam šablon a časté dotazy.

U otázek jsou uvedeny nápovědy. Nástroj nabízí podporu, na kterou se lze obrátit.

3. Data Stewardship Wizard

Nástroj je poměrně složité použít bez jeho znalosti či dokumentace. Nabízí zcela jiné funkcionality a možnosti než ostatní nástroje a zároveň používá poněkud odlišné postupy a názvy.

Nápověda k použití nástroje ale není dostupná a viditelná přímo z nástroje, pouze z hlavní stránky. K nástroji je na youtube umístěn i přehledný návod k použití. Jelikož je ale tento nástroj možnostmi rozsáhlejší než jiné nástroje, jsou i jeho instrukce k použití objemnější, a uživateli tak trvá déle se s nástrojem naučit pracovat.

U každé otázky je stručná nápověda a případný odkaz na podrobnou nápovědu. Nástroj nabízí podporu, na kterou se lze obrátit.

4. DataWiz

Nástroj poskytuje k otázkám nápovědy, které mohou být doplněny o odkazy s více informacemi. Tyto odkazy přesměrovávají na jednotlivé části DataWiz dokumentace. Nástroj nabízí podporu, na kterou se lze obrátit.

4. ZHODNOCENÍ NÁSTROJŮ

5. **easyDMP**

Nástroj je jednoduchý na použití a lze ho tedy využít bez pomoci a dokumentace. V horní liště je dobře viditelná záložka pomoci, odkud lze zjistit více informací o jednotlivých šablonách, přečíst si návod k nástroji, či napsat dotaz na podporu nástroje. Nástroj neobsahuje nápovědy jako ostatní nástroje, ale otázky v některých šablonách mohou být detailnější, a sloužit tak jako nápověda.

6. **ezDMP**

Nástroj je bez problému možné použít bez dokumentace. Ta buď není přítomna, nebo není dobře naležitelná.

V polích otázek jsou uvedeny nápovědy. Nástroj nabízí podporu, na kterou se lze obrátit.

7. **QUT DMP**

Nástroj je jednoduchý na použití a lze ho tedy využít bez pomoci a dokumentace. Ta je dostupná z menu pod hlavičkou nástroje.

U otázek v plánu jsou uvedeny nápovědy. Nástroj nabízí podporu, na kterou se lze obrátit.

8. **DMP OPIDoR**

Nástroj je jednoduchý na použití a lze ho tedy využít bez pomoci a dokumentace. Ta je případně dostupná z horní lišty, kde je mimo jiné i uživatelský účet. Nachází se zde informace o použití nástroje a také seznam šablon.

U otázek v plánu jsou uvedeny nápovědy. Nástroj nabízí podporu, na kterou se lze obrátit.

4.4 Závěrečné hodnocení

Poslední sekci je závěrečné hodnocení vycházející z analýzy a předchozích kategorií hodnocení. Nástrojů na tvorbu plánu správy dat je mnoho. Většina se ale zaměřuje na konkrétní investory či region, a nelze tak příliš vybírat. Například nástroje založené na DMPRoadmap jsou si velice podobné, ale jejich hlavním rozdílem jsou právě regiony, pro které jsou určeny, jelikož se jedná o lokální instalace. I přesto, že byl nástroj hodnocen dobře a měl mnoho funkcionalit, tak není zaručeno, že bude vhodný. Vždy je potřeba zvažovat konkrétní potřeby. V tomto závěrečném hodnocení jsou vypsány silné a slabé stránky jednotlivých nástrojů s informacemi, pro které potřeby mohou a nemusí být vhodné.

1. DMPOnline

Silné stránky:

- Výběr a vytváření šablon
- Podpora organizací
- Jednoduchost a přehlednost

Slabé stránky:

- Freemium model s omezeními
- Otázky formou otevřeného textu
- Převážně pro UK

DMPOnline je jediný z nástrojů v této práci, který není zcela zdarma. Byl vytvořen pro potřeby výzkumníků a investorů ve Velké Británii, pro uživatele mimo UK nemusí být vhodný.

2. DMPTool

Silné stránky:

- Výběr a vytváření šablon
- Podpora organizací
- Jednoduchost a přehlednost

Slabé stránky:

- Převážně pro USA
- Otázky formou otevřeného textu

DMPTool je velice podobný nástroji DMPOnline a nabízí téměř stejné možnosti, pouze zdarma. Jelikož byl ale vytvořen pro potřeby amerických výzkumníků a investorů, pro uživatele mimo USA nemusí být vhodný.

3. Data Stewardship Wizard

Silné stránky:

- Výběr a vytváření šablon
- Podpora organizací
- Málo otázek formou otevřeného textu
- FAIR metriky
- Odhad ceny úložiště

Slabé stránky:

- Odlišné názvy a postupy od ostatních nástrojů
- Složitější na porozumění a orientaci

Ze všech analyzovaných nástrojů nabízí Data Stewardship Wizard nejvíce funkcí. Je zřejmé, že při tvorbě nástroje se kladl důraz na machine-actionability. Výhodou je i skutečnost, že nástroj není omezen na určitý region či investora jako většina nástrojů, a lze ho tak použít téměř univerzálně. S možností tvorby vlastních dotazníků a modelů, a podporou FAIR a machine-actionable praktik tak může být DS Wizard považován za nejlepší volbu z analyzovaných nástrojů.

Je oproti ostatním nástrojům složitější na použití, ale to by se nemuselo vztahovat k výzkumníkům, kteří budou pouze vyplňovat hotové připravené dotazníky.

4. DataWiz

Silné stránky:

- Studie
- Úložiště

Slabé stránky:

- Nelze vytvořit ani vybrat šablonu

DataWiz nabízí kromě tvorby a správy plánu výhody vytváření studií a nahrávání souborů. Neplní tak pouze funkci plánu správy dat. Bohužel neobsahuje šablony, pouze značení, jak plán vyplnit v souladu se třemi konkrétními investory.

5. easyDMP

Silné stránky:

- Výběr a vytváření šablon

Slabé stránky:

- Obsahuje mnoho chyb
- Nespolehlivost, opakovaně nedostupný
- Všechny otázky povinné
- Otázky formou otevřeného textu

Nástroj easyDMP nemůže být doporučen. A to jak kvůli nízkému plnění machine-actionability, tak kvůli chybám a nedostupnosti nástroje. Práce je poněkud zdlouhavá a nepříjemná.

6. ezDMP

Silné stránky:

- Vytvořeno pro potřeby NSF

Slabé stránky:

- Nelze vybrat ani vytvořit šablonu
- Jediné zaměření
- Nelze sdílet
- Velmi stručné

ezDMP nepodporuje spolupráci ani vracení k plánu a počet otázek je velmi nízký. Důvodem vytvoření nástroje bylo usnadnění a zpřehlednění splnění požadavků pro plány dle NSF, což se jistě podařilo. Pro potřeby NSF je tak tento nástroj velmi vhodný, ale pro jiné investory nebude mít s největší pravděpodobností využití.

7. QUT DMP

Silné stránky:

- Vytvořeno pro potřeby QUT

Slabé stránky:

- Nelze vybrat ani vytvořit šablonu
- Jediné zaměření
- Otázky formou otevřeného textu
- Nástroj je pomalý

QUT DMP byl vytvořen pro potřeby Queensland University of Technology v Austrálii a není tak příliš vhodný pro uživatele mimo tuto komunitu.

8. DMP OPIDoR

Silné stránky:

- Výběr a vytváření šablon
- Podpora organizací
- Jednoduchost a přehlednost

Slabé stránky:

- Převážně pro Francii
- Otázky formou otevřeného textu

DMP OPIDoR nabízí na rozdíl od DMPOnline několik drobností navíc a není zpoplatněn. Jelikož byl ale vytvořen pro potřeby francouzských výzkumníků a investorů, pro uživatele mimo Francii nemusí být vhodný.

Závěr

Cílem práce byla převážně analýza a zhodnocení nástrojů sloužících k tvorbě plánů správy dat. Popsat doménu plánování správy dat, analyzovat nástroje a na základě nabytých znalostí je zhodnotit a porovnat.

Nástrojů je mnoho a stále se vytváří nové, které mohou nabízet nové možnosti. Takovým je i nástroj Argos, který se objevil až při finalizaci práce, a proto se jím práce nezabývá. Většinu z těchto nově vzniklých nástrojů ale ještě nelze považovat za dokončené a připravené pro seriózní použití.

V první části práce je popsáno plánování správy dat. Jsou přiblíženy konkrétní aspekty plánů, jejich využití a časté chyby. Je popsán směr machine-actionable, kterým by se plány měly do budoucna ubírat a který zjednodušuje a zkvalitňuje práci s plány. Práce také pojednává o skupinách a projektech spolupracujících na realizaci těchto machine-actionable plánů.

Následující část se věnuje analýze nástrojů. Bylo zvoleno celkem osm nástrojů, které jsou analyzovány a hodnoceny. Jedním z nich je Data Stewardship Wizard vyvíjený na ČVUT FIT v Praze. V analýze je stručně popsán i nástroj CEDAR, který ale není hodnocen. Pro každý nástroj jsou popsány jeho možnosti, rozdělení a funkcionality. Analýza slouží k tomu, aby se čtenář seznámil s nástroji, a je přípravou k následující kapitole Zhodnocení. Již z analýzy s popsánými průchody každým nástrojem si lze udělat představu o využití nástroje a případných nedostatcích.

Kapitola Zhodnocení se detailně zaměřuje na možnosti jednotlivých nástrojů. Ty jsou hodnoceny dle několika kritérií, kterými jsou jejich funkce, ale také uživatelská přívětivost a míra splnění machine-actionability. Taktéž je bráno v potaz, zda je nástroj vhodný pro organizace či jednotlivce. Závěrem této kapitoly je shrnutí silných a slabých stránek nástrojů, ze kterých je patrné, pro které potřeby je nástroj vhodný. Při výběru nástroje je možné zvažovat i cenu, jelikož ne všechny nástroje jsou zcela zdarma. Také je vhodné zvážit použití nástroje, který byl vytvořen pro potřeby konkrétního investora či státu. Pokud je použit mimo zamýšlené oblasti či organizace, může mít omezenou podporu a nějaké funkcionality nemusí být dostupné.

Literatura

- [1] Jones, S.: How to develop a data management and sharing plan. 2011.
- [2] Miksa, T.; Simms, S.; Mietchen, D.; aj.: Ten principles for machine-actionable data management plans. *PLOS Journals*, Mar 2019, naposledy navštíveno Apr 2020. Dostupné z: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006750>
- [3] Jones, S.; Pergl, R.; Hooft, R.; aj.: Data management planning: How requirements and solutions are beginning to converge. *Data Intelligence*, 2020: s. 208–219.
- [4] Voigts, R. J.: Doba datová si žádá data stewardy i chytrá řešení. *itbiz*, Jun 2019, naposledy navštíveno May 2020. Dostupné z: <https://www.itbiz.cz/clanky/doba-datova-si-zada-data-stewardy-i-chytra-reseni>
- [5] European Comission: Data management. Naposledy navštíveno Apr 2020. Dostupné z: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- [6] Walk, P.; Neish, P.; Miksa, T.: DMP Common Standards WG. *RDA*, naposledy navštíveno Apr 2020. Dostupné z: <https://www.rd-alliance.org/groups/dmp-common-standards-wg>
- [7] Giaretta, D.; Ashley, K.; Jones, S.; aj.: Active Data Management Plans IG. *RDA*, Feb 2015, naposledy navštíveno Apr 2020. Dostupné z: <https://rd-alliance.org/group/active-data-management-plans/case-statement/active-data-management-plans-ig.html>
- [8] GO FAIR: FAIR Principles. *GO FAIR*, 2016, naposledy navštíveno Apr 2020. Dostupné z: <https://www.go-fair.org/fair-principles/>

- [9] Wittenburg, P.; Sustkova, H. P.; Montesanti, A.; aj.: The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data. *arXiv preprint arXiv:1902.11162*, 2019.
- [10] Cramer, J. J.: Why You Need a Data Steward and Best Practices to Do it Right. *Dun and Bradstreet*, Mar 2019, naposledy navštíveno May 2020. Dostupné z: <https://www.dnb.com/perspectives/master-data/6-key-responsibilities-of-data-stewards.html>
- [11] Hájek, P.: Jak „mít pořádek“ v datech. *PROFINIT*, Apr 2017, naposledy navštíveno Apr 2020. Dostupné z: <https://profinit.eu/blog/jak-mit-poradek-v-datech/>
- [12] Scholtens, S.; Anbeek, N.; Brullemans, M.; aj.: Towards a community-endorsed data steward profession description for life science research. *F1000Research*, ročník 8, 2019, naposledy navštíveno Apr 2020.
- [13] Walk, P.; Neish, P.; Miksa, T.: RDA DMP Common Standard for machine-actionable Data Management Plans. *RDA*, Mar 2020, naposledy navštíveno Apr 2020. Dostupné z: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>
- [14] DMPOnline: DMPOnline. 2010, naposledy navštíveno May 2020. Dostupné z: <https://dmponline.dcc.ac.uk/>
- [15] DMPRoadmap: DMPRoadmap/roadmap. Naposledy navštíveno May 2020. Dostupné z: <https://github.com/DMPRoadmap/roadmap/network/members>
- [16] Drafiava, M.; Jones, S.; Rust, S.; aj.: Business models for open source: sustaining the DMPonline service. Feb 2020, naposledy navštíveno May 2020. Dostupné z: <https://zenodo.org/record/3670057#.XqFDniNR2iM>
- [17] DMPTool: DMPTool. 2011, naposledy navštíveno May 2020. Dostupné z: <https://dmptool.org/>
- [18] Data Stewardship Wizard: Data Stewardship Wizard. 2015, naposledy navštíveno May 2020. Dostupné z: <https://ds-wizard.org/>
- [19] Data Stewardship Wizard: ds-wizard/dsw-client. Sep 2019, naposledy navštíveno May 2020. Dostupné z: <https://github.com/ds-wizard/dsw-client>
- [20] DataWiz: DataWiz. 2017, naposledy navštíveno May 2020. Dostupné z: https://datawiz.leibniz-psychology.org/DataWiz/?datawiz_locale=en

-
- [21] ZPID: DataWiz. Naposledy navštíveno May 2020. Dostupné z: <https://github.com/ZPID/DataWiz>
- [22] easyDMP: easyDMP. 2015, naposledy navštíveno May 2020. Dostupné z: <https://easydmp.sigma2.no/>
- [23] Moa, H.: easydmp. Naposledy navštíveno May 2020. Dostupné z: <https://github.com/hmpf/easydmp>
- [24] Moa, H.: EasyDMP Documentation. Naposledy navštíveno May 2020. Dostupné z: <https://buildmedia.readthedocs.org/media/pdf/easydmp/latest/easydmp.pdf>
- [25] ezDMP: ezDMP. 2018, naposledy navštíveno May 2020. Dostupné z: <https://ezdmp.org/index>
- [26] ezDMP: ezdmp/ezDMP-Site. Naposledy navštíveno May 2020. Dostupné z: <https://github.com/ezdmp/ezDMP-Site>
- [27] QUT DMP: QUT DMP. 2014, naposledy navštíveno May 2020. Dostupné z: <https://dmp.qut.edu.au/>
- [28] DMP OPIDoR: DMP OPIDoR. 2016, naposledy navštíveno May 2020. Dostupné z: <https://dmp.opidor.fr/>
- [29] CEDAR: CEDAR. 2014, naposledy navštíveno May 2020. Dostupné z: <https://metadatacenter.org/>
- [30] Nielsen Norman Group: 10 Usability Heuristics for User Interface Design. Apr 1994, naposledy navštíveno May 2020. Dostupné z: <https://www.nngroup.com/articles/ten-usability-heuristics>

Obsah přiloženého média

readme.txt.....	stručný popis obsahu média
Obrázky	složka obsahující diagramy a snímky obrazovky
zadaniBP.pdf.....	text zadání práce ve formátu pdf
BP_Machacova_Tereza_2020.pdf.....	práce ve formátu pdf
BP_Machacova_Tereza_2020.tex	zdrojová forma práce ve formátu L ^A T _E X