

IT a anatomie firmy

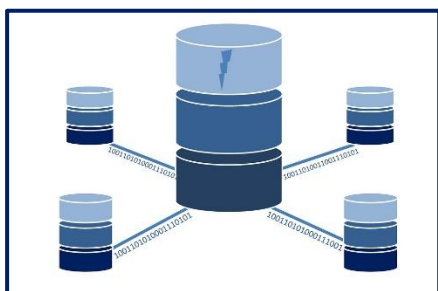
(Řízení datových zdrojů)

(pracovní dokument)



Ondřej Musil, Michal Novotný, Nikola Pátková, Marek, Slezák

VŠE Praha, 2025



Mapa podle částí textu

(A) Musil O.: DAMA-DMBOK, Data Governance, Datové slovníky
(B) Novotný M: Data Mesh
(C) Pátková, N.: Datové katalogy
(D) Slezák, M.: Metodika datové migrace

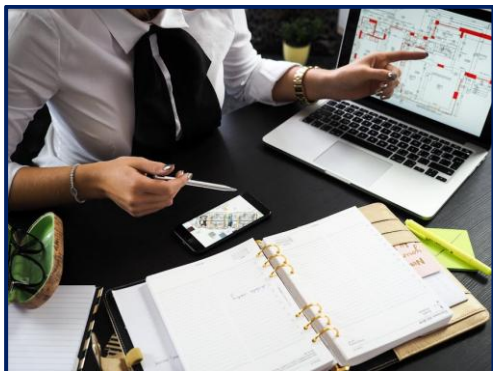
Obsah

A. DAMA-DMBOK, Data Governance, Datové slovníky	6
1. Úvod	7
2. Řízení dat, DAMA-DMBOK2.....	8
2.1 DAMA-DMBOK2 Data Management Framework	8
2.2 Datová architektura.....	10
2.2.1 Podnikový datový model (EDM)	10
2.2.2 Návrh datových toků	12
2.3 Datové modelování a design	14
2.4 Metadata	16
2.4.1 Byznys metadata	16
2.4.2 Technická metadata.....	17
2.4.3 Provozní metadata	17
2.4.4 Data lineage.....	17
3. Data Governance	20

3.1	Komponenty správy dat	20
3.2	Lidé.....	22
3.2.1	Komunikace a spolupráce	22
3.2.2	Znalosti.....	23
3.2.3	Kultura	25
3.3	Procesy	25
3.3.1	Datové principy	25
3.3.2	Datové politiky.....	26
3.3.3	Guidelines	26
3.3.4	Procesy	26
3.3.5	Pravidla a standardy	27
3.4	Nástroje a technologie	27
4.	<i>Byznys a datové slovníky</i>	29
4.1	Business Glossary	29
4.2	Data Dictionary	30
4.3	Lidské výzvy	31
4.3.1	Znalosti.....	32
4.3.2	Motivace a odpor ke změnám	33
4.3.3	Kultura	33
4.4	Organizační a procesní výzvy	33
4.4.1	Nedostatečná podpora zúčastněných stran	33
4.4.2	Nerealistická očekávání.....	34
4.4.3	Procesy a řízení změn.....	34
4.4.4	Obtížnost kvantifikace přínosů.....	35
4.5	Technologické výzvy	35
5.	<i>Zdroje.....</i>	36
B.	<i>Data Mesh.....</i>	40
Úvod		41
6.	<i>Principy.....</i>	41
6.1	Federated computational governance	41
6.2	Self-serve datové platforma	41
6.3	Datové domény	41
6.4	Datový produkt	41
7.	<i>Přínosy a rizika Data Mesh.....</i>	44
7.1	Přínosy	44
7.2	Rizika	44
8.	<i>Role – vlastník datového produktu</i>	44
9.	<i>Závěry a doporučení.....</i>	44
10.	<i>Zdroje.....</i>	45
C.	<i>Datový katalog</i>	46
Úvod		47
1.	<i>Analýza současných přístupů k řešení a užití datových katalogů, jejich možnosti a omezení</i>	47

1.1	Analýza současného trhu.....	47
1.1.1	SaaS Cloud-based.....	47
1.1.2	Open source.....	48
1.1.3	Commercial Data Catalogs.....	48
1.1.4	On-premises Data Catalogs.....	48
1.2	Současné přístupy k užití datových katalogů	49
1.2.1	Proč implementovat	49
2.	<i>Možnosti datových katalogů.....</i>	51
3.	<i>Omezení datových katalogů</i>	52
4.	<i>Přehled vybraných dostupných nástrojů datových katalogů</i>	52
4.1	DataHub.....	53
4.2	MS Purview	54
4.3	Informatica	56
4.4	Collibra Data Intelligence Cloud.....	58
5.	<i>Závěr</i>	59
6.	<i>Zdroje</i>	59
D.	<i>Metodika datové migrace</i>	62
1.	<i>Metodika a postup datové migrace</i>	64
1.1	Používejte formální metodiku, která byla vyzkoušená a otestovaná [1].....	64
1.2	Metodika migrace dat by měla postupovat od posouzení, přes plánování až po přesun a validaci [12]	64
1.3	Mějte plán [12] [15].....	64
1.4	Mějte dostatek času na plánování projektu [15][16]	65
1.5	Zkontrolujte, zda jsou k dispozici potřebné zdroje [16] [19]	65
1.6	Zdokumentujte proces migrace [12][15].....	65
1.7	Vyvíjejte interní kompetence ohledně migrace dat a nespolehejte zcela na externí zdroje [1] 65	65
1.8	Podnik musí být zapojen ve všech fázích projektu, od zahájení až po testování [1] [6].	66
1.9	Získejte písemný závazek vendora [16]	66
1.10	Odstraňte komunikační bariéry [16].....	66
1.11	Automatizujte, pokud je to možné [15]	67
2.	<i>Nástroje pro datovou migraci.....</i>	67
2.1	Používejte nástroje pro profilování dat před zahájením projektu i v jeho průběhu [1]	67
2.2	Pro čištění dat využívejte nástroje pro čištění dat. Nepokoušejte se o čištění ručně [1].	67
2.3	Používejte nástroje pro integraci dat. Nepokoušejte se o ruční kódování [1] [13].....	68
2.4	Ujistěte se, že pro migraci dat používáte správný software [12] [15] [17] [19].....	68
2.5	Je důležité zvážit, zda je možné migraci snadno ukončit a znova spustit [12]	68
3.	<i>Práce s daty.....</i>	69
3.1	Pochopte data ve zdrojovém systému ještě předtím, než začnete migrovat [6] [15].....	69

3.2	Zaměřte se na správné mapování dat, aby byl zajištěn kompletní přenos dat [16].....	69
3.3	Rámec pro řízení dat má zásadní vliv na celý migrační projekt [1] [6]	69
3.4	Pravidelně zálohujte data pro případ, že by během procesu migrace došlo k problému [17] [19]	70
3.5	Identifikujte formáty dat, které budete přenášet [17] [19]	70
3.6	Nahrajte data, aby bylo možné zahájit proces migrace [17]	70
3.7	Investujte do následné údržby, abyste se ujistili, že byla zachována integrita dat [17]..	70
3.8	Omezte počet zprostředkovatelů (vrstev abstrakce) [15]	70
3.9	Vytvořte si knihovnu opakovaně použitelného kódu [15]	71
3.10	Opravujte kód, ne data [15]	71
4.	Testování	71
4.1	Proveďte v cílovém systému testování, abyste se ujistili, že byla data úspěšně přenesena [17] [19]	71
4.2	Pravidelně během migračního projektu testujte [16]	72
4.3	Musí být provedeno systémové a uživatelské testování a před spuštěním ostrého provozu je nutné získat podpis [15]	72
4.4	Plánujte prodloužené období systémových a uživatelských testů [15]	72
4.5	Vytvořte si vztah s uživateli, kteří testují [15]	73
4.6	Ujistěte se, že váš kód loguje vše, co dělá [15]	73
4.7	Získejte novou sadu dat pro každé kolo uživatelských testů [15]	73
5.	Rozpočet	73
5.1	Je-li migrace součástí většího projektu, řešte ji zcela samostatně, a to od sestavování rozpočtu až po testování [1]	73
5.2	Zhodnoťte velikost projektu a zjistěte, jaké zdroje bude vyžadovat a kolik na něj potřebujete v rozpočtu [15] [17] [19]	74
6.	Použitá literatura	74



A. DAMA-DMBOK, Data Governance, Datové slovníky

(pracovní dokument)



Ondřej Musil

VŠE Praha, 2025

1. Úvod

V dnešním rychle se rozvíjejícím digitálním prostředí se data stala jedním z nejcennějších aktiv organizací ve všech odvětvích. Spolehlivá data jsou **klíčem k získání konkurenční výhody**, poskytují přehled o zákaznících, produktech, službách nebo mohou pomoci inovovat a dosahovat strategických cílů. Zároveň jsou nezbytně důležitá **pro každodenní provoz**. Kvůli těmto schopnostem jsou někdy **nazývány „měnou“ nebo „novou ropou“ informační ekonomiky**. V současné době se tak data posunula z pozice sekundárního zdroje do pozice centrálního nervového systému organizace. Neschopnost řídit data je podobná neschopnosti řídit kapitál. Vede k plýtvání a ztrátě příležitostí. Z tohoto důvodu je nutné **věnovat datům, jejich integritě a správě velkou péči** a pozornost (Gupta a Cannon 2020; DAMA International 2017).

Přestože si spousta společností tuto důležitost uvědomuje, jen málo organizací spravuje data jako podniková aktiva, z nichž může získávat trvalou hodnotu. Získávání hodnoty se neděje izolovaně a náhodou, **vyžaduje záměr, plánování, koordinaci a odhodlání**. Vyžaduje řízení a vedení. V naplnění těchto požadavků **může pomoci koncept data governance**, který představuje strukturovaný přístup ke správě dat, jako kritického aktiva organizace. Data governance **zahrnuje procesy, zásady a standardy nezbytné k zajištění integrity, bezpečnosti a použitelnosti dat** v celé organizaci. Díky správě dat mohou organizace zavést konzistentní postupy, které usnadňují sdílení dat, dodržování předpisů a provozní efektivitu (DAMA International 2017; Otto 2011).

Protože se jedná o zavedení významné změny do organizace, tak úspěšná správa dat **vyžaduje silného sponzora z řad vedoucího managementu**, který bude prosazovat iniciativy a sladit vedení, spolu s jasnou vizí a hlavními zásadami, které se začlení do organizačních cílů. Zásadní je rovněž kulturní posun, který umožní vnímat data jako **cenné aktivum, podpořený proaktivním řízením změn**, které řeší odpor. Klíčové je rovněž zapojení všech zúčastněných stran a jasná, nepřetržitá komunikace zajišťující spolupráci a porozumění napříč týmy. Postupná implementace pomáhá organizacím efektivně se přizpůsobit, zatímco multifunkční týmy s různými dovednostmi a školením budují kapacitu. Udržitelnosti je dosaženo **začleněním řízení do standardních procesů s metrikami** pro měření efektivnosti a řízení neustálého zlepšování (DAMA International 2017). Prvků pro úspěšnou a dlouhodobě efektivní správu dat je mnoho, což z tohoto procesu dělá **velmi komplexní záležitost**, které je potřeba se pečlivě věnovat a postupně vyvíjet a posouvat.

Data governance je **centrálním prvkem řízení dat**, jež se skládá z **několika „znalostních“ oblastí**, které se vzájemně doplňují. Jedná se o oblasti datové architektury, datového modelování a designu, ukládání dat, zabezpečení dat a soukromí, integraci dat a interoperabilitu, správu obsahu a dokumentů, správu referenčních a kmenových dat, data warehousing a business intelligence, metadata management, data quality management a řízení datové architektury (DAMA International 2017). Znalostní oblasti jsou velmi rozsáhlé a definice každé z nich není předmětem tohoto dokumentu. Pozornost bude tedy zaměřena především na oblast datové architektury, datového modelování a metadat ve vztahu k data governance.

V rámci konceptu správy dat hrají **zásadní roli metadata**, která umožňují organizaci porozumět svým datům, jejich původu, účelu, vlastnictví apod. Tím je umožněno, že **data jsou přesně definována, strukturována a snadno přístupná** všem relevantním uživatelům. Zde hrají klíčovou roli **business (dále také „byznys“) a datové slovníky**. Byznys slovníky standardizují terminologii v rámci organizace a poskytují všeobecně srozumitelné definice klíčových pojmů, zatímco datové slovníky dokumentují specifickou strukturu datových objektů a jejich vztahů (Seiner 2014). Tyto nástroje společně podporují data governance tím, že **podporují společný jazyk**, snižují nejednoznačnost a podporují přesnost dat. Zavedením byznys slovníků a datových slovníků mohou organizace **zlepšit srozumitelnost a transparentnost dat**, což je nezbytné pro informované rozhodování a dodržování regulačních požadavků (Kimball a Ross 2013). Samotná oblast metadat však nestačí, protože ta se zabývá především tím, jak je publikovat, spravovat, uchovávat a podobně. Více, **jak a kde metadata držet je předmětem datové architektury**, přičemž prostředkem pro získání strukturovaných dat je **oblast datového modelování**. Mezi architekturou a modelováním je silná vazba, protože modely jsou provázány s datovou architekturou a umožňují vizualizovat její vize (DAMA International 2017).

2. Řízení dat, DAMA-DMBOK2

Data governance (DG) je centrálním prvkem **konceptu zvaného Data management (DM)**, jež se skládá z několika znalostních oblastí, které se vzájemně doplňují. V tuto chvíli je vhodné formálně rozlišit pojmy data governance a data management. V českém jazyce lze oba termíny přeložit jako **správa či řízení dat**, a tak je jednoduché tyto termíny zaměnit. V dalších částech tohoto dokumentu budou použity především anglické alternativy či zkratky uvedeny výše. Při použití českých spojení se „správou dat“ dále v textu rozumí Data Governance a „řízením dat“ se rozumí Data Management. Zaměňování těchto pojmů je však běžnou praxí i v jejich originálním znění, pro lepší porozumění se hodí v krátkosti přiblížit co je to data management a jaký je vztah mezi DM a DG.

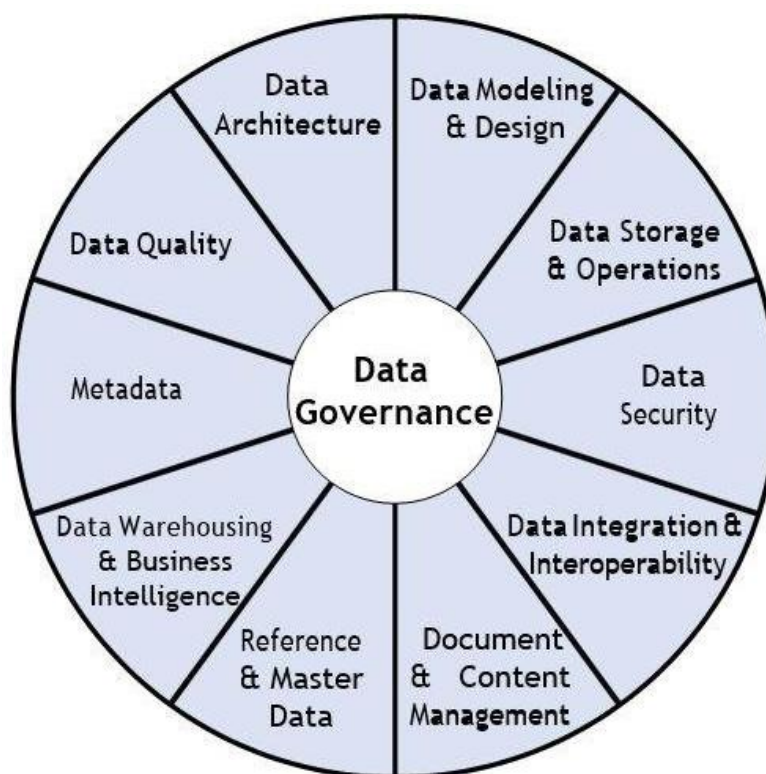
V úvodu práce byla zmíněna **správa dat, jako aktiva, tzv. „data as an asset“**. Podle (O'Sullivan a Sheffrin 2007), **aktivem je jakýkoliv hmotný či nehmotný zdroj**, který podnik nebo ekonomický subjekt vlastní nebo ovládá a který může být použit pro vytvoření kladné ekonomické hodnoty. Data lze považovat rovněž za **podnikové aktivum**, ačkoli chápání toho, co znamená spravovat data jako aktivum se stále vyvíjí (Lake a Crowther 2013). Mnoho podniků, jejichž cílem je udržet si konkurenceschopnost, se v současné době označují jako „**data-driven**“, což znamená, že přijímají rozhodnutí podložená daty a používají analytiku k získání akčních poznatků. Být „data-driven“ zahrnuje **poznání, že data musí být spravována efektivně a profesionálně** prostřednictvím byznys vedení a technických znalostí (Anderson 2015).

Data management má společné rysy s jinými formami řízení aktiv. Zahrnuje **znalost toho, jaká data organizace má a čeho by s nimi mohla dosáhnout, a následné určení, jak nejlépe využít datová aktiva** k dosažení cílů organizace (EWSolutions 2013). Řízení dat lze shrnout jako postup, při kterém se s daty zachází jako s jedinečným a cenným aktivem, což vyžaduje pečlivé plánování, zajištění kvality a spolupráci napříč jednotlivými funkcemi. Na rozdíl od jiných aktiv se data při používání nespotřebovávají, ale **jejich hodnota musí být ekonomicky měřena bilancováním nákladů spojené s nízkou kvalitou dat a naopak přínosů** plynoucích z vysoké kvality dat. Efektivní data management se opírá o metadata, která poskytují kontext a usnadňují porozumění, spolu se správou životního cyklu, která řeší komplexity vytváření, používání a transformace dat. Vyžaduje holistický celopodnikový přístup, který integruje technické i netechnické odborné znalosti a vyvíjí se spolu s měnícími se způsoby vytváření a využívání dat. Kromě toho musí **data management zmírňovat rizika, jako je zneužití, ztráta a etické problémy, a zároveň sladit rozhodnutí v oblasti IT se strategickými potřebami** dat. K zajištění odpovědné a bezpečné správy dat a k podpoře cílů organizace je nezbytné odhodlání vedoucích pracovníků a koordinované úsilí napříč celou organizací (DAMA International 2017).

2.1 DAMA-DMBOK2 Data Management Framework

V úvodu bylo nastíněno, že data management zahrnuje soubor vzájemně provázaných funkcí či oblastí, z nichž každá má své specifické cíle, činnosti a odpovědnosti. V rámci řízení **dat je třeba zohlednit problémy spojené se snahou získat hodnotu z abstraktního podnikového aktiva** a zároveň vyvažovat strategické a provozní cíle, specifické obchodní a technické požadavky, požadavky na rizika, regulační požadavky a protichůdné chápání toho, co data představují a zda jsou kvalitní. Z toho je patrné, že je potřeba **sledovat spoustu faktorů**, a proto pomáhá mít k dispozici „aplikační rámec“, neboli framework, který umožňuje komplexně pochopit tuto problematiku a vidět vztahy mezi jejími dílčími částmi.

Jeden z rámců, který podrobněji popisuje znalostní oblasti, které tvoří celkový rozsah DM je **DAMA-DMBOK framework**. Jednou z možných vizualizací, který tento framework znázorňuje je tzv. DAMA Wheel (viz další obrázek).



Obrázek 2-1: DAMA-DMBOK2 Wheel (DAMA International 2017)

DAMA Wheel umísťuje **DG do centra data management činností**, protože je nutná pro **zajištění konzistence všech dílčích funkcí a rovnováhy** mezi nimi. Ostatní znalostní oblasti jsou umístěny okolo. Kromě data governance (uprostřed) zobrazuje DAMA Wheel následující znalostní oblasti:

- datová architektura,
- datové modelování a design,
- ukládání a operace,
- zabezpečení dat a soukromí,
- integrace dat a interoperabilita,
- správa dokumentů a obsahu,
- referenční a kmenová data,
- data warehousing a business intelligence,
- metadata,
- datová kvalita (DAMA International 2017).

Všechny oblasti jsou **nezbytnými součástmi vyspělého řízení dat**, ale mohou být implementovány **v různém čase** v závislosti na požadavcích organizace. Jak je z výčtu patrné, existuje mnoho znalostních oblastí a vzhledem k jejich širokému rozsahu se nelze věnovat každé z nich.

Tento fakt je zmíněn také v samotném úvodu dokumentu. Důležitým **prvkem v rámci implementace byznys a datových slovníků jsou metadata**. Oblast metadat sama o sobě však řeší především proces publikace, uchovávání a nakládání s metadaty. Data a metadata je potřeba sledovat již při jejich pořízení. O tom, jaká data se budou sbírat, jakým způsobem, jakou budou mít strukturu, jak se budou používat ve společnosti, pojednává **oblast datové architektury společně s datovým modelováním**. V rámci metadat je třeba věnovat zvláštní **pozornost obchodním metadatům**, která poskytují datům kontext a význam, což usnadňuje jejich lepší pochopení a využití v rámci organizace.

V tomto dokumentu tedy bude věnována **pozornost především oblastem datové architektury, datového modelování a metadat**. I v rámci těchto oblastí je však třeba vybrat pouze ty části, které jsou pro řešení relevantní.

2.2 Datová architektura

Návrh a **správa architektury** se provádí **na různých úrovních** v rámci organizace (podniková, doménová, projektová atp.) a **s různými oblastmi zaměření** (infrastruktura, aplikace, data). Lidé, kteří nejsou architekti, a nerozpoznají rozdíly mezi těmito úrovněmi a oblastmi zaměření, mohou být zmateni, co přesně je úlohou architektů. Jedním z důvodů, proč jsou architektonické rámce cenné, je to, že umožňují ne-architektům tyto vztahy pochopit (DAMA International 2017).

Datová architektura je nezbytná pro správné řízení dat, neboť organizace mají typicky více dat, než jsou jednotliví lidé schopni pochopit, je tedy **nutné reprezentovat data organizace na různých úrovních abstrakce**, aby jim bylo možné porozumět. Datová architektura poskytuje **základní rámec** pro řízení a organizaci podnikových datových aktiv. Slouží **jako plán**, který zajišťuje soulad dat s obchodními cíli i technickými systémy a umožňuje organizacím přijímat informovaná rozhodnutí a udržovat konzistenci. Tím, že poskytuje vizi toho, jak jsou data strukturována, uložena a integrována, podporuje datová architektura **efektivní vývoj a aplikaci těchto nástrojů** a překlenuje propast mezi obchodními potřebami a technickou implementací (DAMA International 2017).

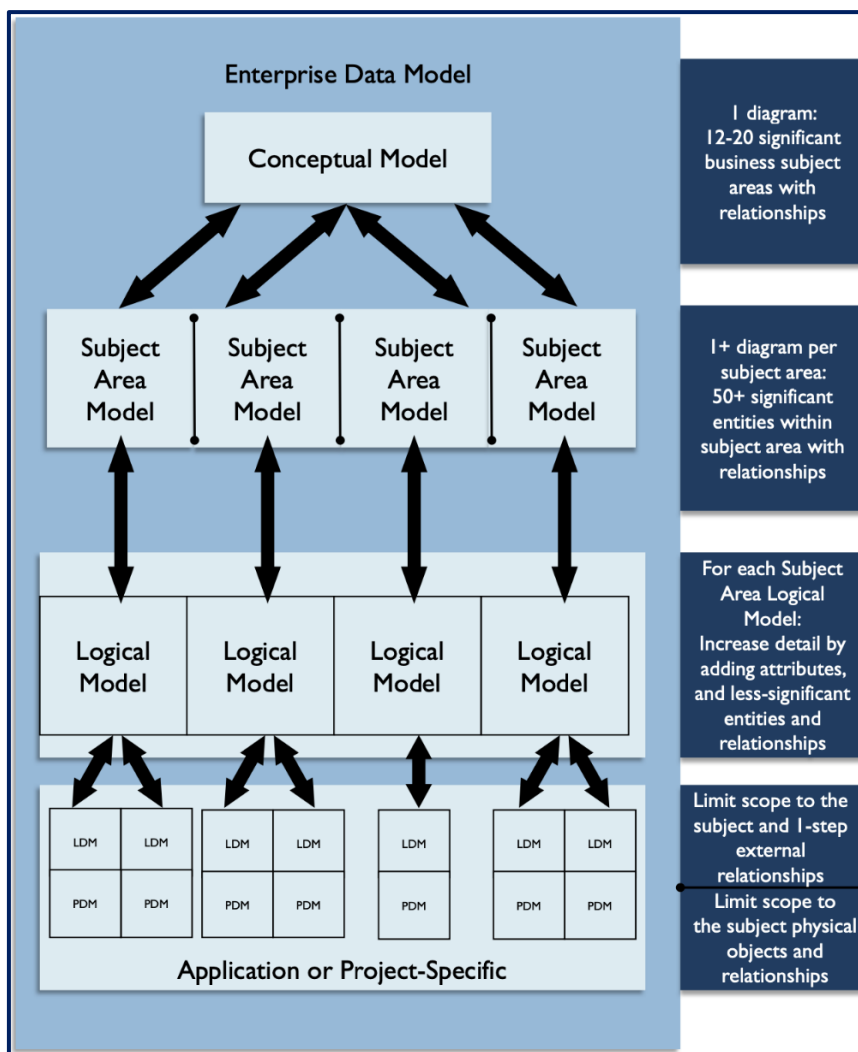
Základní funkcí datové architektury je stanovit **vizi pro nakládání s daty a jejich ukládání**. To zahrnuje vytvoření **plánu, který definuje vztahy mezi obchodními daty** – používanými k popisu procesů, cílů a výsledků – a aplikačními daty, která se nacházejí v základních technických systémech. Propojením obchodních a aplikačních dat mohou organizace zajistit, aby jejich datová aktiva byla nejen technicky spolehlivá, ale také smysluplná a relevantní pro zúčastněné strany v celém podniku (DAMA International 2017).

Jak **data** proudí v organizaci prostřednictvím kanálů nebo rozhraní, **jsou zabezpečena, integrována, ukládána, zaznamenávána, katalogizována, sdílena, vykazována, analyzována a dodávána zúčastněným stranám**. Po cestě mohou **být data ověřována, vylepšována, propojována, certifikována, agregována, anonymizována a používána pro analýzy, dokud nejsou archivována nebo vyčištěna**. Popisy podnikové datové architektury proto musí zahrnovat jak podnikové datové modely (např. datové struktury a specifikace dat), tak **návrh datových toků** (DAMA International 2017).

2.2.1 Podnikový datový model (EDM)

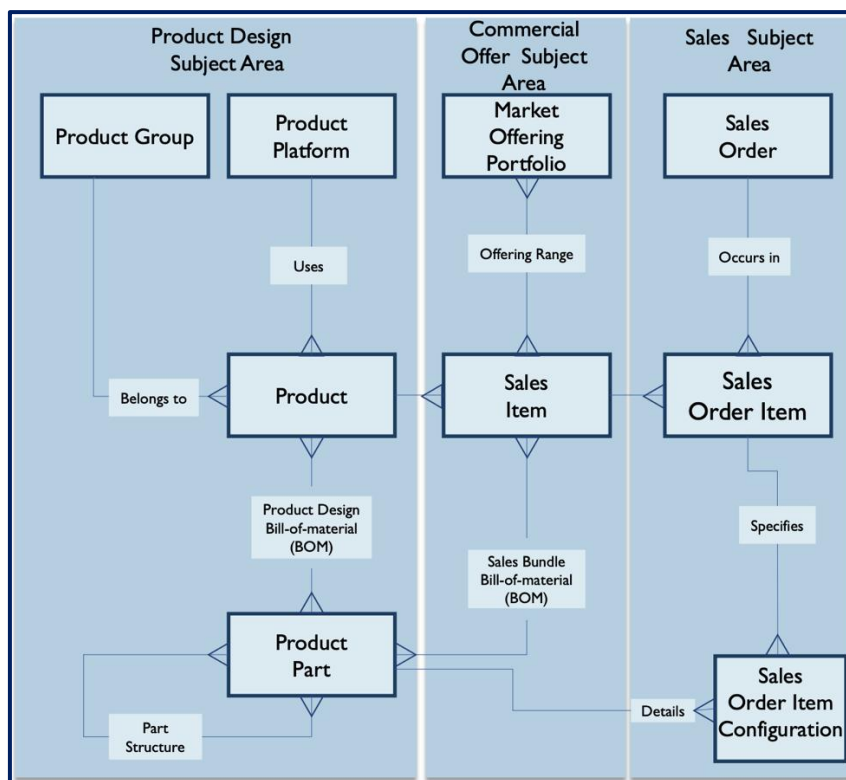
Podnikový datový model (*Enterprise data model; EDM*) je **ucelený konceptuální nebo logický datový model na podnikové úrovni, nezávislý na implementaci**, který poskytuje konzistentní pohled na datové entity organizace a jejich vztahy. **Zahrnuje názvy** a komplexní **definice** dat a metadat, klíčové **datové entity** podniku (tj. obchodní koncepty), jejich **vztahy**, **kritická řídicí obchodní pravidla** a některé **kritické atributy**. Stanovuje **základ pro všechny projekty** týkající se dat. **Každý datový model** na úrovni projektu musí vycházet **z podnikového datového modelu**. EDM pomáhá organizacím standardizovat chápání dat definováním základních konceptů a jejich vzájemných vazeb. To přímo **souvisí s byznys slovníky**, jejichž cílem je standardizovat terminologii napříč odděleními, a datovými slovníky, které definují technické specifikace těchto pojmů. Sladěním EDM s těmito nástroji mohou organizace zajistit, aby jejich obchodní a technické definice byly konzistentní a integrované (DAMA International 2017).

EDM může být vybudován s různou úrovní detailu. Postupem času, jak to potřeby podniku vyžadují, se rozsah a úroveň podrobnosti v podnikovém datovém modelu obvykle rozšiřují. K tvorbě podnikových datových modelů je tedy vhodné **přistupovat iterativně, pomocí vrstev**. Všechny úrovně jsou součástí EDM a **vazby mezi nimi vytvářejí cesty pro sledování entity shora dolů** a mezi modely ve stejné úrovni. Jak spolu různé typy modelů souvisejí a jak jsou konceptuální modely nakonec propojitelné s fyzickými datovými modely aplikací je k vidění na obrázku: Obrázek 2-2.



Obrázek 2-2 Enterprise Data Model (DAMA International 2017)

Horizontální vztahy mezi modely zobrazuje Obrázek 2-3, kde jsou znázorněny **vztahy mezi konceptuálními datovými modely** tří předmětových oblastí. Vztahy mohou překračovat hranice předmětových oblastí; každá entita v podnikovém datovém modelu by měla být umístěna pouze v jedné předmětové oblasti, ale může být spojena s entitami v jakékoli jiné předmětové oblasti. **Konceptuální datový model podniku je tedy vytvořen kombinací modelů předmětových oblastí** (DAMA International 2017).

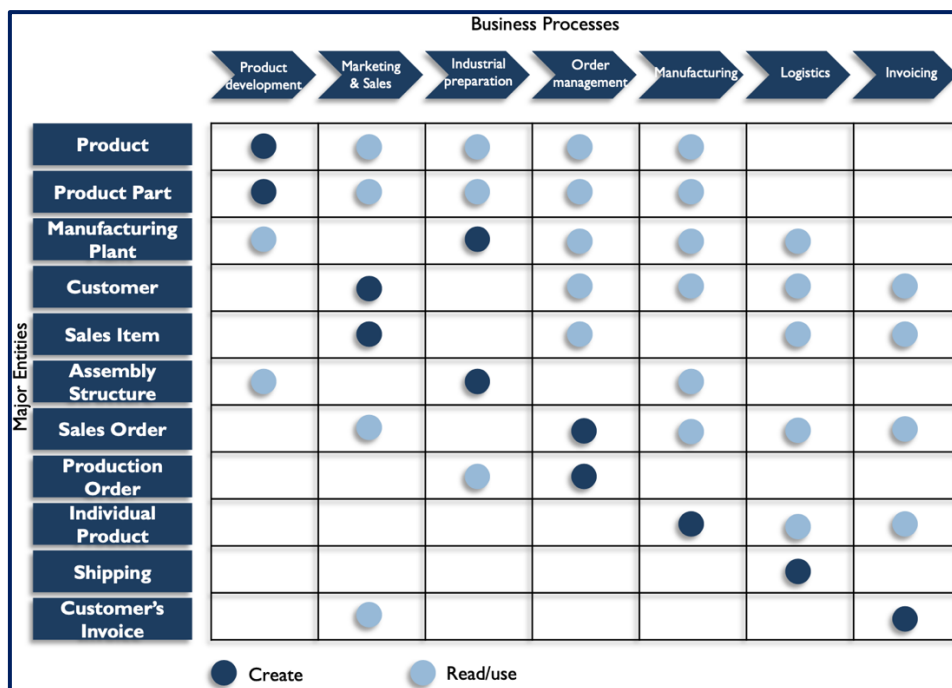


Obrázek 2-3 Horizontální závislost mezi datovými modely (DAMA International 2017)

2.2.2 Návrh datových toků

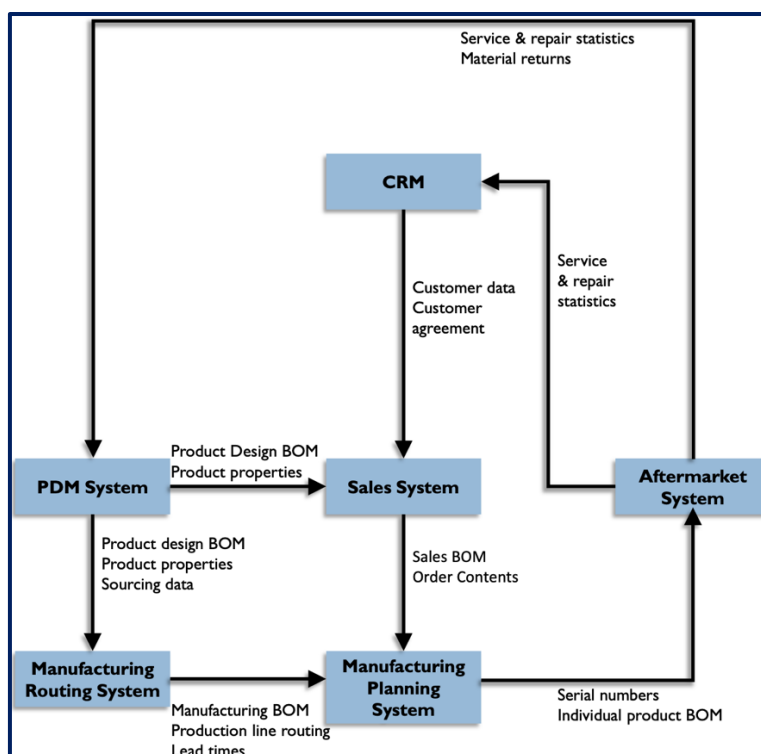
Návrh datových toků **definuje požadavky a hlavní plán pro ukládání a zpracování dat v databázích, aplikacích, platformách a sítích**. Tyto datové toky mapují **pohyb dat k podnikovým procesům**, místům, podnikovým rolím a k technickým komponentám – mezi různými systémy, aplikacemi atp. Tento proces zajišťuje, že data jsou správně chápána a využívána po celou dobu svého životního cyklu. Zdokumentováním původu, transformací a cílů dat **podporuje návrh toku dat možnosti mapování datových cest (data lineage) a sledovatelnosti (traceability)**, které jsou pro obchodní i datové slovníky zásadní. Toto sladění zajišťuje, aby používání dat bylo v souladu s organizačními zásadami a standardy (DAMA International 2017).

Podobně jako v případě podnikových datových modelů mohou být **datové toky dokumentovány na různých úrovních detailu**. Datové toky lze běžně **znázornit** dvěma způsoby – **pomocí dvourozměrné matice** (Obrázek 2-4) **nebo diagramem** datových toků (Obrázek 2-5).



Obrázek 2-4 Matice datových toků (DAMA International 2017)

Výhodou matice je, že poskytuje jasný **přehled o tom, jaká data jsou procesy vytvářena a používána**. Zároveň zohledňuje skutečnost, že data neproudí pouze jedním směrem – výměna dat mezi procesy probíhá **mnohostranně a poměrně složitě**, přičemž jakákoliv data se mohou objevit kdekoli. Kromě toho lze matici použít k objasnění **odpovědnosti procesů za získávání dat a závislosti dat mezi procesy**, což následně zlepšuje dokumentaci procesů (DAMA International 2017).



Obrázek 2-5 Diagram datových toků (DAMA International 2017)

Datová architektura se skládá ze tří dimenzí – artefaktů (modely, definice, datové toky, road mapy), **činností** (formování, zavádění a naplňování záměrů datové architektury) a **chování** (spolupráce,

mindset, dovednosti mezi různými rolemi. **Implementace** podnikové datové architektury se tedy týká **především vytvoření**:

- organizace týmů podnikové datové architektury,
- vytvoření počátečních verzí artefaktů datové architektury,
- vytvoření a zavedení způsobu práce datové architektury ve vývojových projektech,
- vytváření povědomí v celé organizaci o hodnotě úsilí v oblasti datové architektury (DAMA International 2017).

Budování a udržování podnikové architektury je komplexní proces, který je nutné sledovat a vyvíjet společně s tím, jak se vyvíjí celá společnost. S růstem organizace se datové prostředí stává složitější, proto se **datová architektura musí přizpůsobit tak, aby podporovala další zdroje dat a případy použití**. Aby bylo možné takovou architekturu škálovat, je potřeba při jejím budování myslet i na budoucnost (DAMA International 2017).

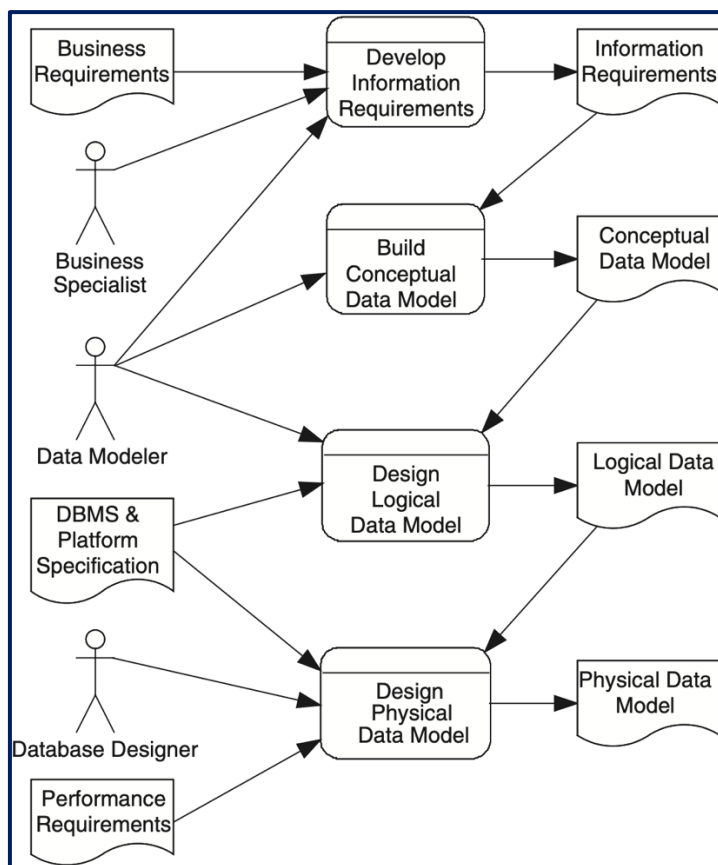
2.3 Datové modelování a design

V předchozí části byla představena znalostní oblast datové architektury, jejímž jedním z hlavních artefaktů jsou datové modely. Tvorbou těchto modelů se zabývá oblast datového modelování a designu. **Datové modelování** je procesem **zjišťování, analyzování a stanovení datových požadavků a následné reprezentace a komunikace** těchto požadavků formou datových modelů, což umožňuje společnosti porozumět svým datovým aktivům (Simsion a Witt 2004; DAMA International 2017).

Podobně jako dům nelze postavit bez plánu, ani systém či databázi nelze vybudovat bez alespoň implicitního modelu. Zároveň jde o designovou činnost s jistou formou volby a kreativity. Pro daný problém bude obvykle existovat mnoho možných modelů, které splňují byznys požadavky a odpovídají pravidlům správného návrhu. **Výběr vhodného modelu je dán celou řadou kritérií**, které se v jednotlivých případech mohou lišit. Tvůrci datových modelů často čerpají z předchozích zkušeností, používají **standardizované vzory** a různé tvůrčí techniky k navržení datových modelů (Simsion a Witt 2004).

Existuje celá **řada schémat** používaných k reprezentaci dat (např relační, dimenzionální, objektově orientované atp.). Modely těchto schémat existují **na třech úrovních abstrakce – konceptuální, logické a fyzické**. Každý model obsahuje **sadu komponent – entity, vztahy, fakta, klíče, atributy** atd. (DAMA International 2017).

Obrázek 2-6 zobrazuje **přehled úkolů a výsledků návrhu databáze** od obchodních požadavků až po kompletní specifikaci databáze.



Obrázek 2-6 Přehled úkolů a výsledků návrhu databáze (Simsion a Witt 2004)

Konceptuální datový model je (relativně) **technologicky nezávislá specifikace dat**, která mají být uložena v databázi. Je výsledkem komunikace mezi tvůrcem datového modelu a zainteresovanými stranami z oblasti byznysu a obvykle se prezentuje **ve formě diagramu s podpůrnou dokumentací**. **Logický datový model** je převodem konceptuálního modelu do struktur, které lze implementovat pomocí systému pro správu databází (DBMS). Tím se obvykle rozumí, že tento model specifikuje tabulky a sloupce, které budou použity. **Fyzický datový model** zahrnuje veškeré změny nutné k dosažení odpovídajícího výkonu a je rovněž prezentován v podobě tabulek a sloupců spolu se specifikací fyzického úložiště (které může zahrnovat distribuci dat) a přístupových mechanismů (Simsion a Witt 2004).

Úloha datového modelování však přesahuje rámec technického návrhu schématu; je **zásadní pro podporu konzistence mezi konceptuální a logickou reprezentací dat a obchodní sémantikou** definovanou ve slovnících. Například konceptuální modely a byznys slovníky: Na konceptuální úrovni datové modely často definují „**high-level entity a vztahy**“, které odpovídají termínům v byznys slovníku. Toto mapování zajišťuje, že slovník odráží nejen obchodní definice, ale také jejich kontextové vztahy v rámci organizace. **Logické datové modely** zpřesňují strukturu definovanou v konceptuálním modelu podrobným popisem atributů, klíčů a vztahů. Tyto podrobnosti jsou **v úzkém souladu s metadaty** spravovanými v datových slovnících a poskytují jasnou **specifikaci datových typů, formátů a omezení**, které posilují datovou kvalitu a data governance (DAMA International 2017).

Mezi **klíčové principy**, které se prolínají s **implementací byznys a datových slovníků**, patří:

- **Formalizace:** Přesné **definování datových struktur a vztahů** snižuje nejednoznačnost sémantiky dat a umožňuje posoudit, jak jsou data ovlivněna implementovanými obchodními pravidly. Vysoce kvalitní definice z datových modelů mohou být **přímo navázány na byznys slovníky**, čímž se zajistí jasnost a soulad s obchodním jazykem.
- **Definice rozsahu:** Datový model může pomoci vysvětlit **hranice datového kontextu a implementace** zakoupených aplikačních balíčků, projektů, iniciativ nebo stávajících systémů.

- **Uchovávání znalostí:** Datové modely fungují jako **dokumentační artefakty**, které uchovávají **porozumění organizace jejím datovým strukturám**. To doplňuje úlohu datových slovníků, které uchovávají podrobná metadata pro průběžnou referenci a správu (DAMA International 2017).

Datové modelování je **prostředkem k získání strukturovaných dat**, ale pouze v případě, že je zajištěn soulad s podnikovou datovou architekturou. **Sladěním datových modelů s architektonickými standardy** organizace zajišťují **konzistenci**, vyhýbají se redundanci a udržují **bezproblémové spojení mezi obchodními požadavky a technickými implementacemi**. Toto sladění nejen usnadňuje interoperabilitu a správu dat, ale také umožňuje přizpůsobivost v závislosti na vývoji obchodních potřeb a technologií. Bez tohoto propojení hrozí, že se datové modely stanou izolovanými a nekonzistentními, což podkopává jejich schopnost poskytovat spolehlivá a použitelná data pro rozhodování (Simsion a Witt 2004).

Významnou výzvou při využívání datového modelování je **udržení souladu mezi obchodní sémantikou a podrobnými technickými specifikacemi**. Konceptuální modely musí přesně odrážet obchodní potřeby, zatímco logické a fyzické modely musí tyto potřeby převádět do realizovatelných technických struktur. Důležitá je tedy spolupráce mezi obchodními uživateli a technickými týmy, která zajistí, že oba pohledy budou zachyceny a sladěny (DAMA International 2017).

Datové modely **obsahují metadata**, která jsou pro datové spotřebitele zásadní. Velká část těchto metadat odhalených během procesu modelování dat je nezbytná pro další funkce řízení dat – např. definice pro data governance a lineage pro datové sklady a analytiku (DAMA International 2017). Datové modely jsou rovněž zdrojem metadat pro datové slovníky.

2.4 Metadata

Aby mohla organizace zhodnotit kvalitu svých dat musí mít nejprve schopnost porozumět svým datům, systémům a pracovním postupům. K tomu pomáhají metadata, která jsou nedílnou součástí správy databází a dalších aplikací a výrazně **přispívají ke schopnosti udržovat, integrovat, zabezpečovat, auditovat a řídit** další data (Hopper 2023).

Nejběžnější definice metadat, „data o datech“, je záludně jednoduchá. Rozsah informací, které lze klasifikovat jako metadata, je široký. Metadata zahrnují **informace o technických a obchodních procesech, datových pravidlech a omezeních a logických a fyzických datových strukturách**. Popisují samotná data (např. databáze, datové prvky, datové modely), koncepty, které data reprezentují (např. obchodní procesy, aplikační systémy, softwarový kód, technologickou infrastrukturu), a vazby (vztahy) mezi daty a koncepty. Pro koncové uživatele vypráví příběh o tom, **jak byla data vytvořena, kde skončila, co se s nimi po cestě stalo** (např. obchodní pravidla, transformace atd.) **a jak by se měla a neměla používat** (DAMA International 2017; Hopper 2023).

Bez spolehlivých metadat organizace neví, jaká data má, co data představují, odkud pocházejí, jak se pohybují v systémech, kdo k nim má přístup ani co znamená, že data jsou kvalitní. Bez metadat nemůže organizace spravovat svá data jako aktivum. Ve skutečnosti bez metadat nemusí být organizace schopna spravovat svá data vůbec. Pokud chce být organizace **„data-driven“ musí být rovněž „metadata-driven“** (DAMA International 2017; Ladley 2012).

Metadata lze rozdělit **do tří kategorií: byznys** (nebo také „obchodní“), **technická a provozní** (nebo „operační“). Nad kategoriemi je však vhodné uvažovat z pohledu původu metadat, nikoliv z pohledu, komu jsou určena. Ačkoli každá kategorie podporuje konkrétní obchodní nebo technickou potřebu, celý soubor metadat **poskytuje hodnotu všem uživatelům metadat**. Obchodní uživatelé najdou hodnotu v technických a provozních metadatach stejně jako techničtí uživatelé v obchodních metadatach (Hopper 2023).

2.4.1 Byznys metadata

Obchodní metadata popisují **obsah a stav dat a zaměřují se na netechnické detaily** související se správou dat. Zahrnují definice, popisy a vlastnosti datových prvků, jako jsou **koncepty, entity, atributy a obchodní pravidla**. Mezi příklady obchodních metadat patří definice data setů, obchodních pravidel, datových modelů, standardů kvality dat, časové posloupnosti dat, úrovně zabezpečení a kontaktní informace na správce nebo vlastníky dat (DAMA International 2017). Obecně lze říci, že byznys metadata **poskytují odpověď na tyto otázky:**

- Jaká data máme k dispozici?

- Jaká je jejich definice?
- Kde je lze získat?
- Odkud pochází?
- Kdy byla nahrána?
- Byla změněna?
- Kde je mohu použít?
- Kdo je může vidět?
- Jak je smím nazývat?
- Koho mám kontaktovat, když mám dotaz? (Hopper 2023)

Obchodní metadata jsou **klíčovým druhem metadat**. Umožňují data **pochopit, kategorizovat, lokalizovat a rozhodnout**, jak je v organizaci využít, čímž pomáhají uživatelům pochopit význam dat a důsledky z obchodního hlediska. Aby však mohla být přínosná, vyžadují pečlivou správu. Jejich údržba a aktualizace může být **náročná na zdroje**, zejména v organizacích s rozsáhlými a dynamickými systémy a databázemi. Pokud nejsou spravována správně, může to vést k nekonzistenci a nepřesnostem, což snižuje jejich spolehlivost a efektivitu (Dremio 2023).

Obchodní metadata jsou mimo jiné významnou součástí tohoto dokumentu. Jsou **součástí byznys slovníků**, které lze považovat za specifickou sbírku metadat, která definuje klíčové termíny, koncepty a metriky používané v organizaci. Tato sbírka pak slouží jako **referenční bod pro všechny zúčastněné strany**, který zajišťuje společné porozumění podnikové terminologii. Některá byznys metadata jsou zároveň součástí i datových slovníků, například ve formě definic datových prvků, obchodních pravidel o použití dat v obchodních procesech nebo pro porozumění, jak specifické datové elementy souvisejí s obchodními funkcemi či cíli.

2.4.2 Technická metadata

Technická metadata poskytují podrobné **informace o technických aspektech dat, systémech, v nichž jsou uložena, a procesech**, které řídí jejich pohyb. Zahrnují prvky, jako jsou **názvy databázových tabulek a sloupců, vlastnosti sloupců, fyzické datové modely, přístupová oprávnění, pravidla CRUD** a zdokumentované vztahy mezi modely a aktivy. Zahrnují rovněž podrobnosti o ETL procesech, **source-to-target mapování, data lineage**, názvy programů a aplikací včetně jejich popisu, pravidla obnovy a zálohování, přístupová práva k datům a další. Tato metadata jsou nezbytná **pro správu technické infrastruktury** a pochopení toku a dopadu změn v datových systémech (DAMA International 2017).

2.4.3 Provozní metadata

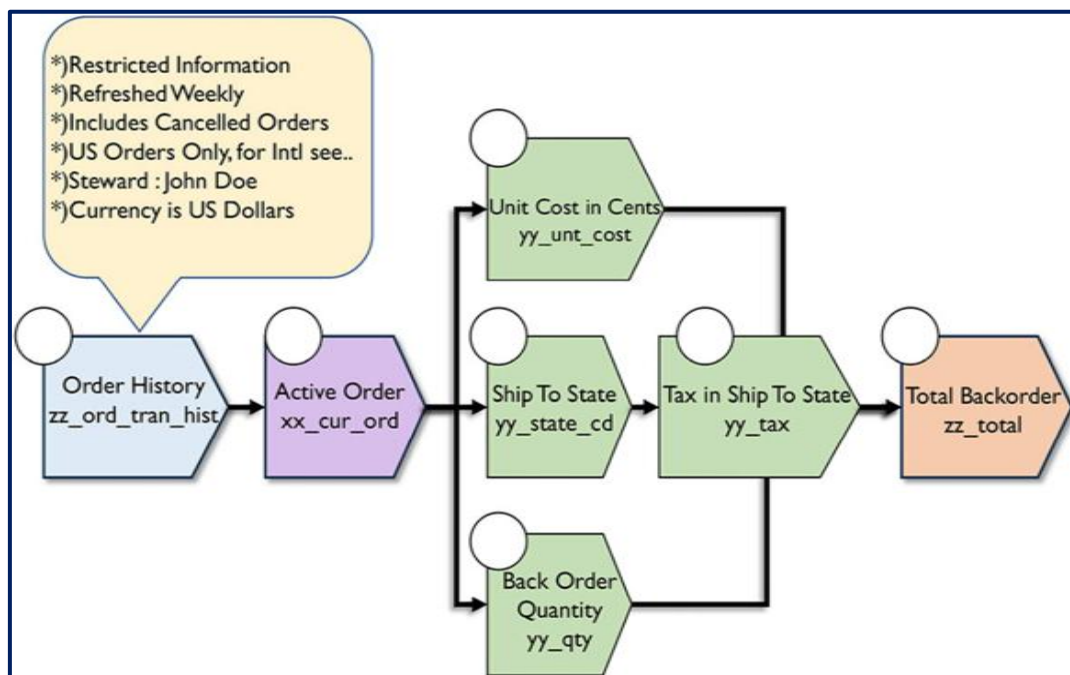
Provozní metadata popisují **charakteristiky rutinních operací s daty a souvisejícími statistikami**. Patří sem scheduling, popisy datových úloh, statusy přesunu a transformace dat, statistiky read/update výkonu, statistiky objemu dat, statistiky dostupnosti a informace o zálohování a archivaci. **Denní informace o procesech extrakce, transformace a načítání dat (ETL)** jsou důležité pro pochopení aktuálního stavu dat a jejich dostupnosti (Hopper 2023).

Data governance tým by měl být zodpovědný za **definování standardů a řízení změn stavu metadat**. Také může být zodpovědný za propagační aktivity a rozvoj školení nebo skutečné školení v celé organizaci. Vyspělejší správa metadat často vyžaduje, aby obchodní termíny a definice procházely různými změnami stavu; například od kandidátského termínu, přes schválený, zveřejněný až po konečný bod životního cyklu – nahrazený nebo vyřazený. **Tým správy dat může také spravovat asociace obchodních termínů**, jako jsou související termíny, synonyma, a také kategorizaci a seskupování termínů (DAMA International 2017).

2.4.4 Data lineage

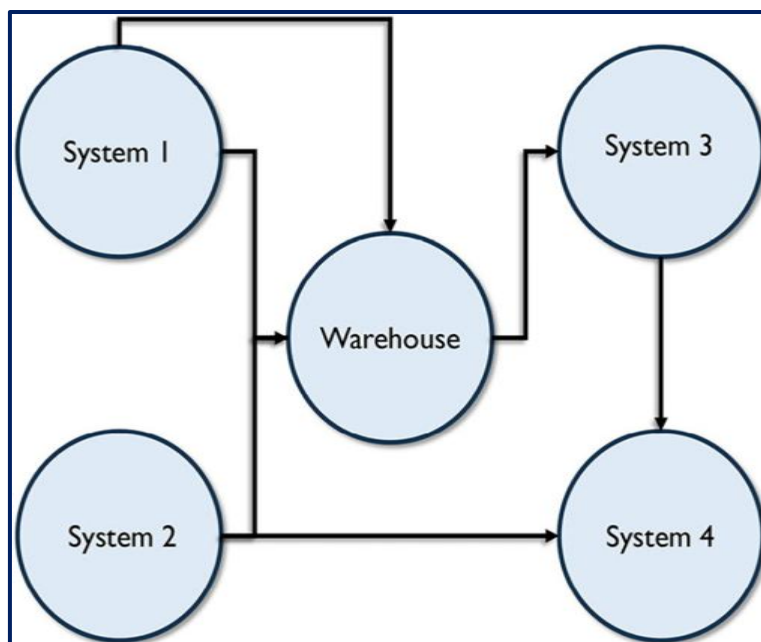
Významným přínosem zjišťování a dokumentování metadat o fyzických aktivech je poskytování informací o tom, **jak se data při pohybu mezi systémy transformují – tzv. data lineage** (někdy také mapování datových toků). Výsledkem je ucelená vizualizace dat při jejich pohybu z původních míst (oficiální zdroj nebo systém záznamu) až do okamžiku, kdy se ocitnou ve svém konečném cíli (DAMA International 2017).

Lineage lze dokumentovat na různých úrovních. Obrázek 2-7 obsahuje ukázkou datového toku na úrovni datového prvku, která popisuje, co se děje s vybraným datovým prvkem (tj. na úrovni sloupců). (DAMA International 2017).



Obrázek 2-7 Ukázka toku na úrovni datového prvku (DAMA International 2017)

Vyšší úrovně lineage (např. systémová lineage) shrnují **pohyb na úrovni systému nebo aplikace**. Obrázek 2-8 ukazuje ukázkou systémového toku, kde je vizualizován obecný pohyb dat na úrovni systému nebo aplikace (DAMA International 2017).



Obrázek 2-8 Ukázka toku na úrovni systémů (DAMA International 2017)

S rostoucím počtem datových prvků v systému se datová lineage stává složitým a obtížně zvládnutelným procesem. Proto je vhodné **začít dokumentovat především ty prvky, které jsou z obchodního hlediska klíčové** a jejich tok zdokumentovat podrobně od zdrojových systému až k cílovému prvku.

Tento přístup umožní technologickým a obchodním uživatelům používat efektivně data a odpovědět na **otázky týkající se datových prvků v celém podniku**, například „Kde je uloženo číslo sociálního pojištění?“, nebo generovat dopadové reporty, například „Jaké systémy budou ovlivněny, pokud se změní šířka určitého sloupce?“ (DAMA International 2017).

Metadata jsou nezbytná pro efektivní řízení dat a musí být dobře spravována, aby byla zajištěna jejich spolehlivost a užitečnost. Správně spravovaná metadata **poskytují kontext, zvyšují kvalitu dat a zlepšují efektivitu organizace** tím, že umožňují maximální využití dat, snižují redundanci a minimalizují náklady na výzkum a školení. Podporují **dodržování regulatorních předpisů**, zefektivňují pracovní postupy a usnadňují komunikaci mezi uživateli dat a IT odborníky. Kromě toho snižují rizika projektů díky přesné analýze dopadů a urychlují vývoj systémů. Špatně spravovaná metadata mají za následek redundantní data, nekonzistentní definice, protichůdné zdroje a sníženou důvěru ve spolehlivost dat a metadat (DAMA International 2017).

Vzhledem k tomu, že řízení metadat je stěžejní pro určení jaká data organizace má, jde často **spolu s řízením datové kvality o výchozí body při definování data governance ve společnosti** (Hopper 2023).

3. Data Governance

Data Governance není pojmem novým, první zmínky o tomto konceptu lze zaznamenat již v osmdesátých letech dvacátého století a od té doby na toto téma vyšlo nespočet knižních publikací, článků v časopisech a internetových zdrojů. Existence těchto zdrojů je bezesporu velmi důležitá pro rozvoj a výzkum v této oblasti a jako vzdělávací materiál pro organizace zavádějící principy správy dat. Na druhou stranu toto může vést přesvědčení, že jde o jednoduchý proces, spočívající v dodržení souboru logických kroků týkajících se lidí, procesů a technologií. Ve skutečnosti však **jde o velmi komplexní a dlouhodobý proces, který je třeba neustále vyvíjet** a vyžaduje dlouhodobou vytrvalost a kreativitu (Gupta a Cannon 2020; Dremio 2023).

Pojem Data Governance může být obtížné jednoznačně definovat, protože správu dat nelze definovat izolovaně – význam, rozsah a implementace do značné míry závisí na organizaci, odvětví a případu použití a vyžaduje čas a velké úsilí. Jinými slovy, **platnost jakékoliv data governance definice je úzce a složitě spjata s řadou charakteristik organizace**, jako jsou vize, hodnoty a strategické priority organizace, portfolio rizik, lidé a kultura, organizační struktura, požadavky zainteresovaných stran, regulační požadavky a požadavky na dodržování předpisů, směrnice a protokoly pro správu dat, stupeň vyspělosti správy dat v organizaci, dynamika trhu a další. Díky těmto faktorům je nezbytné, aby vedoucí **pracovníci správy dat a vrcholový management vypracovali definici data governance na míru**. Proces může začít obecnější převzatou definicí, ale neměl by tím skončit. Data governance však není pouze zodpovědností C-level managementu nebo IT oddělení, **každý zaměstnanec, který užívá data pro výkon svých činností, přímo či nepřímo přispívá k data governance**, byť malou mírou. Každý zaměstnanec má tedy formální povinnost zajistit, že data governance programy a iniciativy ve společnosti uspějí (Gupta a Cannon 2020).

Níže je **několik populárních definic**, od kterých se lze odrazit při definování data governance v organizaci.

- Data Governance je soubor postupů a procesů, které pomáhají zajistit formální správu datových aktiv v organizaci (Knight 2019).
- Data Governance je definována jako výkon pravomoci a kontroly (např. plánování, monitorování a prosazování) nad správou datových aktiv (DAMA International 2017).
- Data Governance je disciplína, která stanoví jasné zásady, postupy, standardy, role, odpovědnosti a zodpovědnosti, aby bylo zajištěno, že data jsou dobře spravována jako podnikový zdroj (DGPO [b.r.]).

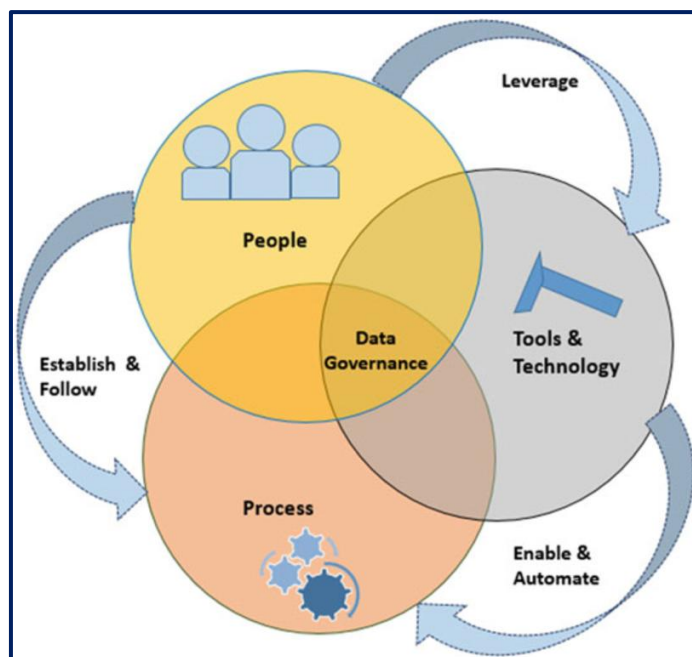
3.1 Komponenty správy dat

Správa dat je velmi komplexní a rozsáhlou funkcí řízení dat. Zajišťuje **zásady, procesy, pokyny, standardy, pravidla, role, odpovědnosti, kontroly a rozhodovací práva**. Zajišťuje také, aby byly zavedeny **vhodné metriky** pro dohled nad efektivním zachycováním a správou dat v rámci celého podniku, a podporuje vhodné chování při využívání dat. V průběhu celého životního cyklu dat se zaměřuje na **zpřístupnění dat všem zúčastněným stranám**, aby k nim měly snadný přístup a aby je mohli využívat způsobem, který generuje požadované obchodní výsledky a **zajistit soulad s regulačními normami** a vnitřními politikami (Eryurek et al. 2021).

Vidíme tedy, že provádění správy dat má několik vzájemně se ovlivňujících složek, které lze rozdělit do **tří obecných komponent** kterými jsou:

- lidé,
- procesy,
- nástroje a technologie (Mahanti 2021).

Lidé zřizují a dodržují procesy a zároveň **využívají nástroje a technologie** pro umožnění a automatizaci těchto procesů (Obrázek 3-1).

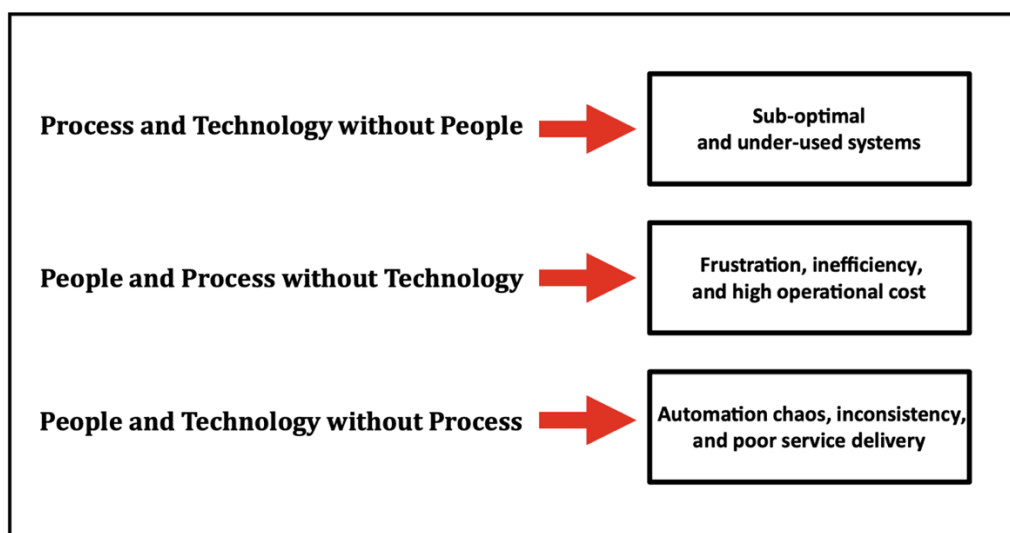


Obrázek 3-1 Komponenty správy dat a jejich vzájemné vztahy (Mahanti 2021)

Aby bylo možné čerpat benefitů ze správy dat, **musí všechny komponenty vzájemně spolupracovat**. Pokud pouze jedna z komponent chybí, může to znamenat neúspěch při implementaci DG programu.

- **Procesy a technologie bez lidí** vedou k systémům, které nesplňují požadavky a zůstávají jako nedostatečně využívané systémy. Je to proto, že systémy používají lidé a pokud by se lidé nezapojili, systémy by nebyly správně vybudovány a nebyly by používány.
- **Lidé a procesy bez technologií** vedou k frustraci, neefektivitě a vyšším provozním nákladům. Zatímco například profilování velkého množství dat pomocí nástrojů a technologií může správně trvat několik minut, manuální provedení by vyžadovalo mnoho lidských zdrojů a bylo by velmi časově náročné, neefektivní a náchylné k chybám.
- **Lidé a technologie bez procesů** vedou k automatizačnímu chaosu, nekonzistenci a poskytování nekvalitních služeb. Je to proto, že proces určuje, jaké činnosti je třeba provést, kdy a kdo je má provádět, aby byly výsledky včasné, předvídatelné a konzistentní. Zavádí tedy určitou úroveň kontroly. Pokud není proces zaveden, mohou lidé využít technologie k automatizaci požadavků, to by však mělo za následek mnoho zmatků, špatné poskytování služeb (například dlouhé dodací lhůty a zpoždění) a zvýšené náklady (Mahanti 2021).

Výše uvedené skutečnosti jsou shrnuty v obrázku níže (Obrázek 3-2).



Obrázek 3-2 Dopady chybějících Data Governance komponent (Mahanti 2021)

Vzájemná spolupráce všech komponent je tedy skutečně velmi důležitá. Zároveň však lze říci, že některé komponenty jsou více důležité než jiné. Tato skutečnost vyplývá i z popisu dopadů chybějících komponent, z jehož je patrné, že v případě že chybí lidé či procesy, znamená to špatně vybudované systémy, chaos, nekonzistenci nebo nekvalitní služby. V případě že chybí technologie, jde především o snížení efektivity. Z toho vyplývá, že **nejdůležitějšími složkami jsou jednoznačně lidé a jejich znalosti** spolu s procesy. Nástroje a technologie jsou sice nezbytné pro podporu iniciativ v oblasti správy dat, ale slouží především ke zlepšení efektivity, nikoliv jako hnací síla. **Nástroje** pomáhají automatizovat úkoly, usnadňují komunikaci a poskytují analytické funkce, které zlepšují rozhodování. Efektivní nástroje mohou také motivovat lidi tím, že zjednoduší jejich úkoly a umožní lepší spolupráci.

Důraz na lidskou a procesní komponentu je obzvlášť důležitý především protože spousta společností, i v současné době, věnuje spoustu času výběrem různých specializovaných nástrojů, na které vynaloží významné finanční prostředky s vidinou, že nástroj vyřeší všechny jejich potíže s daty. Zároveň si neuvědomují, či nechtějí uvědomit, že je rovněž potřeba změnit přístup lidí, nastavit procesy, zavést pravidla atp. Organizace procházejí jednotlivými kroky, najímají si konzultanty a nakupují správné nástroje, avšak v momentě kdy nastane čas změnit každodenní zacházení s daty, setkají se s odporem a selhávají (Ladley 2012).

3.2 Lidé

Lidská komponenta je jádrem data governance programu. Lidé spolupracují na **vytváření formálních, standardních a konzistentních zásad, procesů, pravidel a standardů** v rámci celé organizace. Aby toho dosáhli, je třeba definovat **provozní model správy dat, organizační struktury a orgány, definovat a přiřadit jasné role, odpovědnosti a rozhodovací práva**. Tato část je velmi důležitá, neboť přesvědčit zaměstnance, že data governance je i jejich zodpovědnost, a ne pouze záležitostí vedoucího managementu je jedním z nejtěžších úkolů, kterým je třeba čelit a zároveň jde o klíčovou složku pro zaručení úspěchu data governance programu (Mahanti 2021).

Podle (Ladley 2012) je data governance v zásadě o **úpravě chování a obchodních procesů** s cílem zlepšit způsob správy a využívání dat. Zvládnout **změnu v organizační kultuře** je jedním z klíčových faktorů úspěchu, protože data governance se zpravidla dělá proto, že organizace něco nedělá správně a proto je třeba něco změnit. Je třeba se **orientovat, školit, vzdělávat, komunikovat, vést, povzbuzovat**.

3.2.1 Komunikace a spolupráce

Úspěch správy dat do značné míry **závisí na efektivní komunikaci**. Rozdíly v terminologii a interpretaci dat, informací a rozhodování v rámci organizace často vytvářejí překážky. Dosažení společného porozumění vyžaduje dlouholeté úsilí. Navíc komunikační strategie, které fungují v jedné organizaci nebo kontextu, nemusí fungovat v jiné a neexistuje žádný univerzální přístup nebo návod, který by zaručil úspěch. Aby byla zajištěna účinnost iniciativ správy dat, **musí být komunikační strategie**

přizpůsobeny a upraveny tak, aby odpovídaly jedinečným potřebám dané organizace a jejích jednotlivých útvarů (Gupta a Cannon 2020).

Klíčovým aspektem je **přířazení odpovědnosti za výsledky související s daty**, což často zahrnuje **zavedení rolí**, jako jsou správci dat a vedoucí správy. Tyto osoby mají za úkol dohlížet na kvalitu dat, zajišťovat dodržování zásad a fungovat jako most mezi obchodními a IT týmy. Komunikace hraje klíčovou roli při sjednocování těchto zúčastněných stran a vyžaduje jasné sdělení a opakované posilování, aby se překonal odpor a podpořilo přijetí (Ladley 2012).

3.2.2 Znalosti

Práce s daty je komplexní záležitostí, která vyžaduje kombinaci technických, obchodních a mezilidských znalostí a dovedností. V předchozí kapitole byly představeny znalostní oblasti, které jsou v rámci implementace byznys a datových slovníků důležité. Znalosti v rámci těchto oblastí budou nyní blíže rozebrány.

Datová architektura

Ve světě datové architektury je základním kamenem profesionálního úspěchu mít komplexní soubor dovedností. **Datový architekt musí být mistrem přesnosti** a musí umět splétat obrovské množství informací do ucelených struktur, které organizacím umožní plně využít potenciál jejich datových zdrojů. Rychle se rozšiřující digitální prostředí vyžaduje, aby **datoví architekti nejen pevně ovládali základní dovednosti, ale také aby zůstali agilní a přizpůsobovali se rychlému vývoji** datových technologií a metodik. (Teal 2023a).

Autoři (Teal 2023a; Kimball a Ross 2013; DAMA International 2017) se shodují na těchto dovednostech, jakožto klíčových pro roli datového architekta:

Datoví architekti **musí mít strategickou podnikovou perspektivu a zajistit, aby jejich návrhy odpovídaly obchodním a IT cílům organizace**. To zahrnuje přípravu na budoucí vývoj produktů, služeb a technologií, sladění datových strategií s cíli organizace a zajištění souladu se standardy správy dat. Klíčové je, **aby datové systémy podporovaly dlouhodobý růst a reflektovaly celopodnikové strategie**.

Schopnost komunikace a spolupráce je zásadní, protože architekti **propojují obchodní a technické týmy**. Pracují úzce s vývojáři, správci a koncovými uživateli, aby zajistili, že technická řešení odpovídají provozním i analytickým potřebám, jinými slovy, převádějí obchodní požadavky do funkčních technických návrhů.

Technické znalosti jsou pilířem dovedností datového architekta. To zahrnuje tvorbu datových modelů, správu datových toků, **návrh úložišť a integraci architektonických rámců**. Důležitá je také schopnost vizualizace a dokumentace prostřednictvím jasných diagramů a plánů, které usnadňují komunikaci a porozumění mezi zúčastněnými stranami. **Návrh škálovatelných systémů, optimalizace výkonu a správa metadat** jsou klíčové pro efektivní architektury.

Řízení změn, řešení problémů a zmírňování rizik patří mezi základní úkoly architekta. Je nezbytné, aby podporoval inovace a přizpůsobil architektury dlouhodobým cílům organizace. Schopnost předvídat rizika, jako jsou problémy se škálovatelností nebo datovou nekonzistencí, a navrhovat strategie jejich řešení je zásadní pro zajištění spolehlivých systémů.

Vedle technických dovedností jsou podceňované, ale důležité, schopnosti, jako je **obchodní prozíravost a empatie k datům**. Architekti musí chápat strategické cíle organizace a potřeby uživatelů, aby mohli vytvářet systémy podporující růst a rozhodování. Vyžaduje to **vyvážení technických možností s obchodními prioritami** a schopnost efektivně obhajovat zdroje.

Datové modelování

Podle (Simsion a Witt 2004; Teal 2023b) lze **znalosti v oblasti datového modelování** shrnout takto:

Datové modelování vyžaduje **kombinaci celé řady dovedností**, aby bylo možné vytvořit přesné a praktické modely v souladu s cíli organizace. Ve své podstatě zahrnuje konceptualizaci a abstrakci, kdy se k **převedení požadavků do funkčního datového modelu** používá abstraktní myšlení a kreativita. Schopnost abstrakce zjednodušuje složité scénáře na zvládnutelné modely a vyvažuje efektivitu s přesností. Zkušenosti modelářů často zvyšují efektivitu tím, že přizpůsobují vzory a řešení z předchozích projektů, místo aby začínali od nuly. Znalost správy dat navíc pomáhá zajistit soulad s etickými směrnicemi a regulačními normami, což zvyšuje spolehlivost a důvěryhodnost datových modelů.

Hluboké **porozumění obchodní doméně** je stejně důležité, zejména u modelů na podnikové úrovni. To zahrnuje schopnost efektivně **komunikovat s vrcholovým vedením**, aby byl zajištěn soulad s obchodními cíli. Jejich schopnost srozumitelně formulovat složité koncepty a zapracovávat zpětnou vazbu podporuje soulad mezi různými týmy. Čerpání **poznatků z jiných oborů**, například architektury, vybavuje datové modeláře **interdisciplinárními návrhářskými dovednostmi**, které jim pomáhají formulovat problémy a posuzovat různá řešení.

Úspěšné datové modelování se také opírá o spolupráci a komunikaci, která často zahrnuje **workshopy a rozhovory se zúčastněnými stranami** s cílem shromáždit požadavky a ověřit modely. Kromě toho jsou nezbytné silné analytické a hodnotící dovednosti, které umožňují posoudit různé možnosti modelů a zdokonalit je na základě zpětné vazby od uživatelů či zúčastněných stran. V neposlední řadě je nezbytná **znalost oborových standardů a osvědčených postupů**, což zajišťuje, že modely jsou konzistentní, kompatibilní a udržovatelné.

Metadata, byznys a datové slovníky

V případě architektury a datového modelování je jednodušší hovořit o nutných znalostech dané role, protože role jsou jasně dané - tj. architekt či modelář. Z pohledu **metadata managementu** to může být o něco složitější, neboť řízení metadat vyžaduje závazek celé organizace a odpovědnosti za jejich řízení a správu jsou často **rozděleny mezi různé role**. Různé zúčastněné strany mohou převzít konkrétní aspekty správy (definování, údržba atd.) metadat na základě svých znalostí a odpovědností (DAMA International 2017).

V souvislosti s byznys a datovými slovníky, které jsou důležitými nástroji metadata managementu, je důležité zmínit **dokumentační dovednosti**. Metadata slouží jako základ pro dokumentaci datových elementů tím, že poskytují strukturované a popisné informace o datech. Dokumentační dovednosti jsou jemné dovednosti, které pomáhají při **přípravě různých typů dokumentace** (např. datová, projektová, procesní, uživatelská atp.). Osoba, která je tvorbou dokumentace pověřená by měla mít znalosti, které odpovídají znalostem knihovníka či encyklopedisty. Jedná se o kombinaci komunikačních, analytických a organizačních schopností (Indeed Editorial Team 2024a).

Dobré **komunikační dovednosti** jsou opakovaným prvkem, který je zmiňován napříč znalostními oblastmi. V případě dokumentace je tato dovednost velmi důležitá, protože umožňuje **srozumitelně sdílet informace s konzumenty daného zdroje**. Písemné komunikační dovednosti umožňují psát definice, které jsou stručné, informativní a relevantní. Psaní jasných a stručných definic nebo pokynů je znakem dobré dokumentace. Encyklopedisté se specializují na shrnutí složitých témat do srozumitelných hesel, což je dovednost, která je stejně důležitá při vytváření stručných a konzistentních definic. (Indeed Editorial Team 2024a).

Pozornost věnovaná detailům umožňuje zajistit, aby **všechny informace byly přesné a dobře prezentované**. Tato schopnost pomáhá provést korekturu, hledat interpunkční, pravopisné nebo gramatické chyby. Dobrá pozornost věnovaná detailům také pomůže zajistit, aby vytvořená **dokumentace splňovala firemní standardy** (Indeed Editorial Team 2024a).

Důležité jsou rovněž **kvalitní rešeršní dovednosti**, aby bylo možné se rychle zorientovat v dané problematice a **rychle najít relevantní informace**. Tato dovednost také pomáhá efektivně využívat více zdrojů dat a najít přesně ty informace, které jsou třeba. Psaní kvalitní dokumentace často vyžaduje předchozí **průzkum a znalost dané doménové oblasti**.

Aby bylo zajištěno, že dokumentace je napsaná v jazyce, který je smysluplný a známý napříč organizací, je nezbytná **znalost firemní kultury**. Jasné pochopení firemní kultury zajistí, že tón, styl a jazyk používaný v dokumentaci bude **rezonovat s hodnotami organizace** a zohledňovat potřeby a preference publika, kterému je určena. Zároveň pokud dokumentace odráží firemní kulturu, je pravděpodobnější, že ji zaměstnanci přijmou. Znalost oblasti a firemního jazyka rovněž umožňuje **vytváření křížových odkazů**, které uživatelům pomáhají orientovat se v souvisejících tématech. Dobrá dokumentace by také měla obsahovat odkazy nebo propojení na související dokumenty nebo sekce pro lepší kontext (Indeed Editorial Team 2024a; Gupta a Cannon 2020).

Při tvorbě dokumentace je důležité mít organizaci a pořádek. Knihovníci vynikají v systematickém uspořádání informací, což je pro tvorbu strukturované dokumentace klíčové. To zahrnuje **kategorizaci dat, definici pojmů a zajištění snadné dostupnosti informací**. Organizovanost umožňuje spravovat rozsáhlé katalogy a slovníky, přidávat nové položky, aktualizovat stávající nebo vyloučit zastaralé části (Indeed Editorial Team 2024b).

Analytické dovednosti jsou důležité k analýze informací a k určení způsobu jejich začlenění do dokumentace. Díky těmto dovednostem lze také analyzovat vyplněnou dokumentaci a hledat v ní chyby nebo opomenutí a následně opravit a zajistit tak profesionální úroveň dokumentů. Analytické dovednosti lze rovněž využít **k posouzení požadavků uživatele a jejich uplatnění v dokumentaci** (Indeed Editorial Team 2024a).

3.2.3 Kultura

Pro úspěch iniciativ v oblasti správy dat je rozhodující dobré **porozumění firemní kultuře a kultuře dat** v organizaci v kombinaci s účinnými strategiemi řízení změn. Firemní kultura označuje **sdílené hodnoty** a psychologické prostředí organizace. Zahrnuje poslání, vizi a hodnoty organizace a ovlivňuje způsob, jakým zaměstnanci vzájemně komunikují a přistupují ke své práci. Firemní kultura se může projevovat **různými způsoby, včetně stylů komunikace, rozhodovacích procesů a celkové pracovní atmosféry**. Bez kulturního vhledu se nelze účinně zabývat řízením změn a iniciativy budou s větší pravděpodobností neúspěšné. Přílišný důraz na procesy a technologie při zanedbávání „lidského“ aspektu správy dat často vede k neudržitelným snahám. Kulturní složka je sice složitá a neustále se vyvíjející, ale **má zásadní význam** pro to, jak organizace vytváří, spravuje a využívá data. Liší se napříč týmy, vedoucími pracovníky a funkcemi, což ovlivňuje zapojení zainteresovaných stran a dlouhodobou životaschopnost zacházení s daty jako s cenným zdrojem. Mnozí zaměstnanci však význam správy dat nevnímají a často ji odmítají jako záležitost IT, což podtrhuje potřebu kulturního sladění, které by podpořilo smysluplnou účast a porozumění napříč organizací (Gupta a Cannon 2020; Kumar a Ansari 2016).

Kultura a řízení změn jsou vzájemně propojené, ale každá z nich vyžaduje vlastní plánování, provádění a nástroje. Řízení změn, přestože je v DG programech často opomíjeno, je pro zajištění úspěšného přijetí iniciativ zásadní. Ačkoli je často spojováno s technologickými projekty zahrnujícími nové platformy nebo procesy, **řízení změn přesahuje rámec technologií**. Zahrnuje přípravu, vybavení a podporu jednotlivců, aby přijali změnu a dosáhli tak organizačního úspěchu. Uvědomění si, že změna je ze své podstaty náročná, je prvním krokem k jejímu efektivnímu řešení, podpořenému rámci a modely přizpůsobenými potřebám organizace (Gupta a Cannon 2020).

3.3 Procesy

V rámci této komponenty se hovoří o **procesní složce data governance**, což je ale vcelku široký pojem. Podle (Mahanti 2021) lze tuto komponentu **rozdělit na několik elementů**:

- principy nebo datové principy,
- zásady nebo datové politiky,
- guidelines,
- procesy
- pravidla a standardy.

3.3.1 Datové principy

Datové principy jsou tzv. **high-level výroky o tom, čeho chce organizace dosáhnout zavedením správy dat**. Neměly by být zaměňovány s cíli nebo zásadami. Cíli se rozumí požadované výsledky, zatímco principy vedou organizaci po celou dobu její existence (Mahanti 2021). (DAMA International 2017) definuje **principy jako základní zákon, doktrínu, předpoklad, pravidlo či kodex chování**.

Organizační principy jsou jádrem efektivní správy dat, jedná se prohlášení o filozofii, které si lze představit jako listinu práv – základní přesvědčení, která tvoří kotvu pro všechny zásady týkající se správy informačních aktiv. Jsou to přesvědčení, která je třeba uplatňovat každý den jako vodítko pro postupy a rozhodovací úsilí (Ladley 2012).

Některé **příklady zásad řízení dat** jsou následující:

- s daty je třeba zacházet jako se strategickým podnikovým aktivem,
- kvalita kritických datových aktiv by měla být definována, hodnocena, řízena a monitorována v průběhu celého životního cyklu dat,
- data musí být spravována tak, aby byla v souladu s interními i externími zákony a předpisy,
- v souvislosti s daty by měla být stanovena jasná odpovědnost,
- podniková data by měla být standardizována (Mahanti 2021).

Podrobnější pohled na datové principy poskytuje (Ladley 2012), který **definuje sadu tzv. obecně přijímaných informačních principů**, které představuje Obrázek 3-3.

GAIP™ - Generally Accepted Information Principles	
Principle	Description
Content as Asset	Data and content of all types are assets with all the characteristics of any other asset. Therefore, they should be managed, secured, and accounted for as other material or financial assets.
Real Value	There is value in all data and content, based on their contribution to an organization's business/operational objectives, their intrinsic marketability, and/or their contribution to the organization's Goodwill (balance sheet) valuation.
Going Concern	Data and content are not viewed as temporary means to achieve results (or merely as a business by-product), but are critical to successful, ongoing business operations and management.
Risk	There is risk associated with data and content. This risk must be formally recognized, either as a liability or through incurring costs to manage and reduce the inherent risk.
Due Diligence	If a risk is known, it must be reported. If a risk is possible, it must be confirmed.
Quality	The relevance, meaning, accuracy, and life cycle of data and content can affect the financial status of an organization.
Audit	The accuracy of data and content is subject to periodic audit by an independent body.
Accountability	An organization must identify parties which are ultimately responsible for data and content assets.
Liability	The risks in information means there is a financial liability inherent in all data or content that is based on regulatory and ethical misuse or mismanagement.

Obrázek 3-3 Obecně přijímané informační principy (Ladley 2012)

3.3.2 Datové politiky

Politiky nebo zásady mohou být pro nově vznikající DG program užitečné, ale také destruktivní. (DAMA International 2017) definuje politiky jako „...*prohlášení o zvoleném postupu a popis požadovaného chování na vysoké úrovni k dosažení souboru cílů.*“ Pro nový DG program je však snadné chrlit zásady či politiky bez jakéhokoli obsahu. Skutečná **podstata politiky** spočívá v tom, že **jde o kodifikaci principů**. Politiky jsou vynutitelné procesy. Principy bývají příliš „vznešené“ na to, aby se daly přímo prosazovat. Zásady musí být opakovatelné a snadno naučitelné. **Datové zásady** jsou dále **definovány datovými standardy, pravidly a procesy**. Standardy, které jsou pro správu dat důležité, jsou typem politiky nebo dokonce charakteristikou určité politiky, například standardy pojmenování dat nebo standardy kvality dat (Ladley 2012). Konkrétní zásady se mohou soustředit například na přístup, užití, uchovávání dat nebo například datovou kvalitu a integraci.

3.3.3 Guidelines

Guidelines nebo také „**pokyny**“ **se skládají z doporučení a osvědčených postupů** určených k dosažení cílů politiky tím, že poskytují podklady pro navrhování standardů a provádění procesů. Na rozdíl od politik, které jsou závazné, pokyny **nejsou povinné, ale doporučené** a nemusí se uplatnit ve všech situacích (Mahanti 2021).

3.3.4 Procesy

Procesy jsou **metody nebo soubor kroků na podporu různých činností potřebných ke správě dat a implementaci datových politik v organizaci**. Zatímco zásady popisují, co by se mělo dělat, procesy definují, jak je efektivně provádět. Doporučuje se, mít **formální, zdokumentované, konzistentní a**

opakovatelné procesy pro správu dat, které zahrnují stanovení procesů pro implementaci datových politik, identifikaci a stanovení priorit kritických datových prvků, pojmenování a definice dat, dokumentaci obchodních pravidel, změny datového modelu, hodnocení datových rizik, podávání zpráv, řízení problémů s kvalitou dat), stanovení vlastnictví dat a správy dat, rozhodovacích práv a řešení konfliktů, definování změn členství v radě pro správu dat, začlenění kontrolních mechanismů a monitorování metrik (Mahanti 2021).

Aby organizace zmírnily rizika v průběhu celého životního cyklu dat, zavádějí **kontrolní mechanismy**, které mohou být manuální nebo automatizované. Kontrolní mechanismy **pomáhají předcházet chybám v datech**, například ověřovat ruční zadávání dat do polí nebo zajišťovat integritu dat při přenosu do systému. **Automatizované kontroly** mohou zahrnovat porovnávání kontrolních souborů se skutečnými datovými soubory za účelem zjištění nesrovnalostí, zatímco **manuální kontroly** vyžadují lidský zásah, například kontrolu počtu záznamů. Pokud jsou kontroly neúčinné nebo neproveditelné, je třeba zavést nápravné činnosti k odstranění problémů se správou dat (Mahanti 2021).

Procesy správy dat lze dle (Mahanti 2021) volně **rozdělit do čtyř fází**:

- **Objevení** – posouzení současného stavu vyspělosti správy dat, identifikace nedostatků a stanovení plánu pro zlepšení.
- **Definování** – stanovení datových standardů, konvencí pojmenování, zásad a kritérií měření.
- **Aplikování** – implementace zásad správy, obchodních pravidel a provozních pracovních postupů.
- **Měření a monitorování** – vyhodnocení efektivity správy a sledování pokroků při dosahování cílů a v případě potřeby úprav poskytování poznatků zpět do fáze zjišťování.

Ačkoliv procesy poskytují **strukturovaný přístup ke správě dat**, v některých případech mohou být vyžadovány podrobné postupy, které uživatele povedou krok za krokem. Takové procesy by měly být **pravidelně revidovány a aktualizovány**, aby odpovídaly organizačním změnám, aktualizacím předpisů a vyvíjejícím se potřebám správy. Změny v organizační struktuře, revize stávajících politik v oblasti dat, nové zákony a nové předpisy nebo úpravy stávajících zákonů a předpisů jsou některé důvody, které mohou vyžadovat změnu procesů (Mahanti 2021).

3.3.5 Pravidla a standardy

Pravidla zahrnují **obchodní pravidla a pravidla kvality dat**. Standardy odkazují na **datové standardy**. Pravidla a standardy **zajišťují konzistentní výsledky lidí a procesů**, kteří je používají (Mahanti 2021).

Datové standardy definují referenční rámec pro reprezentaci, formát, význam, strukturování, přenos a správu datových prvků. Příkladem datových standardů jsou standardy pojmenování dat, standardy modelování dat, standardy přenosu dat nebo standardy zabezpečení dat. Stejně jako datové politiky a procesy musí být i **datové standardy dodržovány v celém podniku** a měl by být definován proces revize, který zajistí aktuálnost datových standardů (Mahanti 2021).

Je důležité, aby jednotlivci napříč organizací měli **společné chápání podnikových dat**, toho, co tato data představují, a obchodních termínů, které tato data reprezentují. V tomto mohou pomoci právě **byznys slovníky**, které uchovávají definice termínů napříč podnikovými oblastmi, specifické pro dané odvětví a organizaci, které jsou sdíleny napříč organizací. Obchodní slovník však není jen seznam pojmů a definic. Zahrnuje také další cenná **metadata**, jako jsou **transformační pravidla, vlastníci dat, správci dat, popis případných výjimek a variací** (Mahanti 2021).

3.4 Nástroje a technologie

Zatímco správa dat je především o lidech a procesech, nástroje a technologie usnadňují a umožňují správu dat **prostřednictvím automatizace, škálování, rozšiřování a auditovatelnosti**. Organizace uchovávají velké objemy dat a činnosti správy a řízení dat by měly být automatizovány, kdekoli je to možné, aby se dosáhlo rychlosti a přesnosti (Mahanti 2021).

Jeden aspekt nástrojů, který je třeba v tomto bodě pochopit, je, že není nutně vhodné kupovat nástroje pro správu dat jen proto, že se organizace správou zabývá. Podle definice nástroj existuje proto, aby zlepšil něco, co se už v současnosti provádí. Pokud ještě společnost neprovádí formální správu dat nebo ji provádí špatně, pak je shánění nástroje, který pomůže zavést DG, ztrátou času. To je v rozporu s typickou filozofií IT, kdy se nástroj obvykle pořizuje jako první. To je běžně známý postup, který však není nutně správný. Je snadné **koupit nástroj a nainstalovat ho**. Většinou jsme však svědky toho, že

nové nástroje pro správu dat leží nevyužité nebo špatně nasazené. Je to proto, že nikdo nezvládl proces, který nástroj podporuje (Ladley 2012).

Jakmile je DG zavedeno a uživatelé začnou rozumět různým aspektům konkrétního programu správy dat, okamžitě by **mělo být zřejmé, kde je potřeba nástroj, který zlepší dané procesy**. Některé **funkce** nástrojů DG, které lze zvážit, jsou **následující**:

- správa principů a zásad,
- správa obchodních pravidel a standardů,
- správa organizace,
- workflow pro záležitosti a audity,
- datový slovník,
- podnikový vyhledávač,
- správa dokumentů,
- shromažďování, syntéza a prezentace dat z ukazatelů metrik,
- rozhraní pro další pracovní postupy a metodiky,
- školení a zařízení pro spolupráci (Ladley 2012).

Jak již bylo zmíněno, samotný nákup nástroje tedy není dostačující. Důležité jsou především lidé a rovněž procesy. Jedním z nástrojů efektivní správy dat jsou rovněž byznys a datové slovníky, které jsou blíže popsány v následující kapitole.

4. Byznys a datové slovníky

Byznys a datové slovníky jsou významnými **součástmi dobře strukturovaného data governance rámce**. Ačkoliv se koncepce těchto nástrojů může zdát jednoduchá, jejich implementace může být komplexní záležitostí (Desmet 2021). V této kapitole budou nejprve představeny byznys a datové slovníky, poté budou shrnuty výzvy, na které mohou organizace narazit při jejich nasazování.

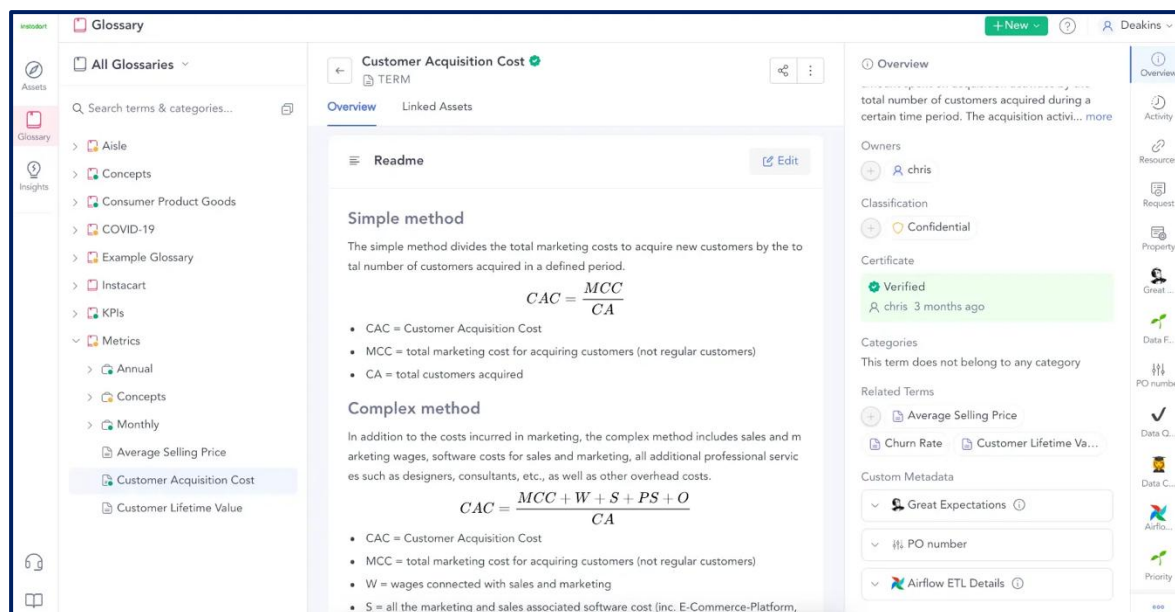
4.1 Business Glossary

Byznys slovník (také „business glossary“, „obchodní slovník“ nebo „slovník obchodních pojmů“) je zásadním **data governance nástrojem**, jehož účelem je dokumentovat a uchovávat obchodní pojmy a terminologii organizace, definice a vztahy mezi těmito pojmy – zabývá se tedy obchodními metadaty. Důležitou charakteristikou byznys slovníků je, že **tato dokumentace je sdílena napříč všemi odděleními** a je tedy nutné **zajistit shodu** pro celou organizaci. Slovník obchodních pojmů vytváří společný jazyk v rámci organizace definováním a standardizací termínů a konceptů a zároveň poskytuje informace o kontextu a vlastnictví datových aktiv. Tato **standardizace zajišťuje konzistenci definic a popisů dat, podporuje společné porozumění** a zlepšuje spolupráci mezi členy týmu. Zlepšuje viditelnost dat, stanovuje standardy kvality a pomáhá při mapování toku dat pro lepší pochopení vztahů mezi datovými aktivy. Tyto prvky společně **budují důvěru v data** a umožňují jejich konzistentní interpretaci a používání v rámci celé organizace (DAMA International 2017; Atlan 2023b; 2023a; Knight 2019).

Byznys slovníky **nemají přesně definovanou formu** nebo povinné atributy, které by měly být dokumentovány, je tedy na každé organizaci, určit si jejich rozsah. Čím jednodušší a snazší je vyhledávání ve slovníku, tím větší je pravděpodobnost, že bude jeho adopce úspěšná a jeho obsah používán. Nej důležitější vlastností slovníku je však to, **aby jeho obsah byl robustní**. Dle (DAMA International 2017) by slovníky obchodních pojmů měli **obsahovat následující atributy**:

- Název termínu, definice, akronym nebo zkratka, případná synonyma.
- Obchodní jednotka a/nebo aplikace odpovědná za správu dat spojených s terminologií.
- Jméno osoby, která termín identifikovala, a datum aktualizace.
- Kategorizační nebo taxonomické spojení pro termín (funkční spojení podniku).
- Rozporné definice, které je třeba vyřešit, povaha problému, časový plán opatření.
- Běžná nedorozumění v termínech.
- Algoritmy a kalkulace týkající se pojmů.
- Mapování datového toku (data lineage).
- Oficiální nebo autoritativní zdroj údajů podporujících termín.

Obchodní slovník by měl být strukturován tak, **aby splňoval funkční požadavky tří hlavních skupin uživatelů – byznys uživatelé, správci dat (data stewardi), techničtí uživatelé**. Byznys uživatelé, jako jsou analytici a vedoucí pracovníci, využijí slovník k pochopení terminologie a vztahům s daty. **Správci dat řídí životní cyklus termínů**, propojují termíny s podnikovými datovými zdroji a **řeší terminologické problémy a kolize** s cílem zlepšit znalosti organizace. Techničtí uživatelé využívají slovník k rozhodování o architektuře, návrhu systémů, vývoji a k provádění dopadových analýz (DAMA International 2017).



Obrázek 4-1 Termín v byznys slovníku (Atlan 2023a)

Byznys slovníky **by neměly být tištěny nebo uchovávány v souborech**, protože jejich obsah není statický. Správci dat jsou obecně zodpovědní za vývoj, používání, provoz a reporting. **Reporting zahrnuje sledování nových termínů a definic**, které ještě nebyly zkontrolovány, termínů, které čekají na schválení a termínů, u nichž chybí definice nebo jiné atributy. Ukázka termínu zaznamenaného v byznys slovníku představuje Obrázek 4-1.

4.2 Data Dictionary

Data dictionary (také „**datový slovník**“) slouží jako **centrální úložiště technických metadat**, která poskytují detailní technické informace o datových elementech organizace v databázových systémech. Datový slovník poskytuje **jasnou představu o datových prvcích, jejich definicích, formátech, vztazích a použití** v rámci organizace (DAMA International 2017).

Vhodně strukturovaný datový slovník obvykle obsahuje **následující prvky**:

- název datového objektu,
- definice,
- datový typ,
- maximální délka nebo rozsah datového prvku,
- pravidla validace,
- vztahy s dalšími datovými elementy,
- vazba na prvky v byznys slovníku (CastorDoc 2024).

Datové slovníky jsou cenným **nástrojem pro IT týmy**, neboť jim umožňují **převést obchodní požadavky do technických specifikací** a rychle identifikovat potřebné vstupy pro projekty, jako jsou typy atributů, požadavky na pole a výchozí hodnoty nebo například původ dat. Data dictionary funguje jako **jediný zdroj pravdy**, čímž zajišťuje sdílené chápání datových aktiv mezi zúčastněnými stranami, podporuje prosazování data governance zásad, **umožňuje data discovery a lineage**. Kromě toho zvyšuje ochranu soukromí a bezpečnost dat tím, že **identifikuje citlivá data a umožňuje řízení přístupu**, a zároveň pomáhá při správě životního cyklu prostřednictvím správné klasifikace, ukládání a archivace dat (DAMA International 2017; CastorDoc 2024; Drori 2019).

V jedné z předchozích kapitol byla představena oblast datového modelování a architektury. **Modelovací nástroje jsou často zdrojem obsahu datových slovníků**. Během vývoje datového modelu je popsáno mnoho klíčových obchodních procesů, vztahů a terminologie. Tyto informace zachycené v logickém datovém modelu se často ztratí při nasazení fyzických struktur do produkčního prostředí. Datový slovník může pomoci zajistit, aby se tyto informace pro organizaci zcela neztratily a aby logický a fyzický model zůstaly ve shodě i po nasazení do produkce (DAMA International 2017).

Přestože se datové slovníky primárně soustředí na technická metadata, **mohou také obchodní terminologií popisovat**, jaké datové prvky jsou k dispozici, poskytovány, s jakými bezpečnostními omezeními, jakým způsobem probíhají jejich transformace nebo jak jsou uplatňovány v obchodních procesech. Vytvořením **vazeb mezi byznys a datovými slovníky** lze překlenout propast mezi IT a podnikovými uživateli (DAMA International 2017).

Rozsah datového slovníku se liší od slovníku obchodních dat. Zatímco datový slovník může být vytvořen **pro každý datový zdroj zvlášť** s cílem porozumět datovým zdrojům a databázím, **rozsah obchodního slovníku** je typicky jeden na celou organizaci s cílem nastolit jednu verzi pravdy (Lim 2021). Datové slovníky jsou často prezentovány ve formátu tabulky s řádky a sloupci definujícími jednotlivé atributy nebo kategorie metadat, které je třeba v systému řešit. Datové slovníky mohou být rovněž součástí modelovacích či databázových systémů. Ukázkou jednoduchého datového slovníku ukazuje Obrázek 4-2.

Database Name	Field Name	Field Label	Description	Field Size (Max number of characters permitted)	Data Type/ Format (e.g., numeric, date, currency, string or free-form text)	Data Codes (for numeric data that represent categories)
HRMS	EmpID	Employee Identification Number	Identification number assigned to employee at time of hire	8	String	N/A
HRMS	SepReas	Separation Reason	Reason an employee has separated from the agency	2	Numeric	1-Abandonment of Position 2-Death 3-Disability – Involuntary 4-Disability – Voluntary 5-Dismissal 6-End of Appointment 7-Layoff 8-Resign 9-Retirement 10-Seasonal

Obrázek 4-2 Jednoduchý datový slovník (QIC WD 2020)

Pojmy business glossary a data dictionary se v praxi často zaměňují, každý z nich se však zabývá jinou oblastí dat. Přestože jde o rozdílné nástroje, panuje mezi nimi komplementární vztah. **Termíny v obchodním slovníku jsou obvykle přímo mapovány na elementy v datovém slovníku**, čímž byznys slovníky dodávají význam datovým slovníkům. **Spojením obchodního a datového slovníku vzniká tzv. datový katalog** (Alation 2024). Jedná se o organizovaný soupis datových aktiv organizace, který informuje uživatele – jak obchodní, tak technické – o dostupných datových sadách k danému tématu a pomáhá jim je rychle vyhledat. **Vztah mezi těmito dvěma prvky je klíčový**, protože uživatelé mají jasný a dostupný přehled o tom, jaká data organizace má, odkud pocházejí, kde se nyní nacházejí, kdo k nim má přístup a jaká rizika nebo citlivé informace s nimi mohou souviset (Lim 2021; Salmi 2021).

Navzdory jasným přínosům plynoucím z těchto nástrojů se organizace často potýkají s problémy při jejich efektivním zavádění. Některé překážky lze zobecnit na celý data governance program, jiné se váží přímo ke slovníkům. Pochopením těchto problémů mohou organizace vyvinout strategie k jejich překonání a **maximalizovat hodnotu svého úsilí v oblasti správy metadat**.

Organizace mají často problém si uvědomit důležitost a význam investic do DG, dokud neutrpí velký tlak ze strany regulačních orgánů nebo nedojde k narušení bezpečnosti dat či výrazné finanční újmě způsobené špatnou manipulací s daty (Bassi a Alves-Souza 2023). Vyvíjející se **regulační prostředí** je zároveň jeden z hlavních hnacích sil a motivací **pro zřízení robustní správy dat**. Dodržování předpisů je neoddiskutovatelným aspektem obchodních operací, což vede k přijetí efektivní správy dat jako způsobu, jak tyto požadavky splnit a udržet si legitimitu (Redman et al. 2025).

V předchozí kapitole byly představeny hlavní komponenty DG programu a shrnuty následky plynoucí z jejich absence. **Pro úspěch data governance programu je tedy nutná součinnost všech komponent** – lidí, procesů a technologií. Výzvy lze rovněž volně rozdělit do těchto kategorií.

4.3 Lidské výzvy

Lidská složka je bezesporu **nejdůležitější komponentou správy dat**, protože úspěšná správa dat je v zásadě o chování organizace a spolupráci, nikoli pouze o technologii nebo procesech. Zatímco procesy definují pracovní postupy správy a nástroje pomáhají automatizovat úkoly správy, ani jedna z těchto složek nemůže být úspěšná bez angažovaných a odpovědných osob. Správa dat je tak účinná, **jak účinní jsou lidé odpovědní za její implementaci, prosazování a vývoj**. Pokud lidé nedůvěřují

snahám o správu, nerozumí jim nebo se jich aktivně neúčastní, ani ty nejlepší nástroje a procesy nedosáhnou zamýšlených výsledků (Ladley 2012).

Vzhledem k tomu, že je lidská složka klíčová, je **zároveň největším možným kamenem úrazu** v rámci DG programu, a proto je třeba ji věnovat speciální pozornost.

4.3.1 Znalosti

V rámci lidské složky hrají významnou roli znalosti. Zaměstnanci a vedoucí pracovníci často nerozumí základním zásadám správy dat. Tato mezera ve znalostech **se rozšiřuje i na pochopení hodnoty kontrolních mechanismů** a jejich přínosu k výkonnosti podniku. Bez komplexního školení chybí mnoha odborníkům základní znalosti potřebné k zavedení účinných postupů správy, což vede k odporu a skepsi (Redman et al. 2025).

Jedním z cílů byznys slovníků je dokumentovat a spravovat organizační znalosti obchodní terminologie související s daty a **vytvořit tak společný jazyk**, díky kterému lidé rozumí obsahu dat a mohou je s jistotou používat (DAMA International 2017). To může být nelehký úkol, protože je třeba **zvolit vhodný, dostatečně exaktní jazyk**, který je srozumitelný napříč organizací a zároveň být schopný použít tento jazyk v praxi. To vyžaduje schopnost dostat jazyk „na papír“ a zároveň znalost firemní kultury a doménové oblasti, které se jazyk týká.

Definice obchodních i technických prvků jsou významnou **součástí byznys a datových slovníků**. Vytvořit kvalitní definici s sebou přináší celou řadu komplexit. V první řadě je samozřejmě nutné zajistit, aby byl popis fakticky správně, kromě toho však existuje celá řada dalších prvků, které je nutné zvážit. **Definice by se měly řídit předem definovanými pravidly.**

- **Publikum** – znalost publika je klíčová, definice musí být přizpůsobena úrovni porozumění jejich konzumentů.
- **Jasnost** – popis musí být jasný a stručný, aby účinně vyjadřoval význam pojmu nebo konceptu. Vhodné je použití jednoduchého jazyka a vyhnutí se žargonu nebo doménově specifických termínů.
- **Rozsah** – definování pojmu vyžaduje zvážení jeho rozsahu a kontextu, aby byla zajištěna jeho přesnost a relevance. Určení vhodné úrovně podrobnosti a zachycení všech relevantních nuancí může být náročné, zejména u termínů s více možnými výklady.
- **Jednoduchost** – definice by měly zahrnovat pouze podstatné aspekty pojmu a zároveň se vyhnout zbytečné složitosti nebo referencím, které s pojmem přímo nesouvisí. Naopak je vhodné použít reference, kde to dává smysl, aby se termíny neopakovaly.
- **Konzistence** – zachování konzistence a dodržování oborových standardů nebo zavedených konvencí je zásadní pro zajištění srozumitelnosti a usnadnění komunikace
- **Revize a aktualizace** – Termíny ve slovníku by se měli vyvíjet spolu s organizací. Je nezbytné pravidelně kontrolovat a aktualizovat definované pojmy (Allemann 2023; Walery 2020).

Kromě toho je nutné mít **hlubokou znalost dané oblasti**. Při vytváření nových definic je možné zpětně získat znalosti o datech z existujících systémů nebo slovníků, existuje však riziko, že nemusí být na první pohled zřejmé, kolik péče bylo věnováno vývoji původních definic. Pokud jsou definice nedostatečně propracované nebo nejednoznačné, neposkytnou spotřebitelům dat informace, které potřebují k pochopení dat, která používají. Na druhou stranu je vhodné nečekat na dokonalost, definice podobně jako celá správa je proces, který je **potřeba neustále vědomě rozvíjet** (DAMA International 2017; Walery 2020).

Lidí s těmito znalostmi a dovednostmi typicky ve firmě **není mnoho**. Kromě toho je tato dokumentace často vložena do rukou technickým autorům, kteří nemusí nutně znát všechny podrobnosti nebo uživatelům, kteří neznají význam těchto činností. Konzumenti pak dokumentaci považují za špatně uspořádanou, špatně zpracovanou a nespolehlivou; nechtějí ji používat a zároveň roste jejich nedůvěra, která může bránit použití v budoucnu (Parnas 2011).

Vytvoření a udržování byznys a datových slovníků, a celého data governance obecně, **vyžaduje specializovaný tým se správnými dovednostmi a schopnostmi**. V případě slovníků jde především o kombinaci jazykových dovedností, odborných znalostí v daném oboru a pochopení potřeb a očekávání cílového publika. Organizace musí investovat do školení, nábory a retenci kvalifikovaných odborníků, aby mohly účinně podporovat svá úsilí v oblasti správy dat (Almeida 2024).

4.3.2 Motivace a odpor ke změnám

Další významnou překážkou, která může bránit adopci byznys a datových slovníků je **odpor ke změnám a problémy s motivací**, které mohou často být způsobeny **špatnou komunikací**.

Lidé si v rámci svého pracoviště vytvářejí pocit identity a sounáležitosti, který je hluboce provázán s jejich nástroji, postupy a procesy. Tyto kulturní prvky slouží jako kotvy, které zajišťují stabilitu, jistotu a pocit kontinuity. Dvě slova, před kterými se lidé se zavedenými pracovními postupy, procesy, a navíc i smyslem a úctou ke své roli třesou, jsou „změna“ a „řízení“. Jednotlivci mohou pociťovat úzkost a nejistotu ohledně své schopnosti přizpůsobit se novým nástrojům a procesům. **Nové procesy a aktivity ve vztahu ke sdílení informací a vytváření společného jazyka napříč organizací** vnímají někteří jedinci jako narušení zavedené sociální struktury a mocenské dynamiky organizace. Zaměstnanci nebo pracovní skupiny se mohou pasivně bránit změnám, protože se obávají ztráty kontroly, postavení nebo vlivu, které mohou tyto změny přinést. Strach z toho, že se vzdají zavedených hierarchií nebo způsobů jednání, může vyvolat odpor a zpomalit tempo implementace (Raden 2023a; 2023b).

Termín, který s tímto chováním souvisí je tzv. „**knowledge hoarding**“ **neboli hromadění znalostí**. O hromadění znalostí hovoříme tehdy, když jednotlivci nebo skupiny v rámci organizace **záměrně zadržují cenné znalosti**, místo aby se o ně podělili s ostatními. Jde o vědomé rozhodnutí omezit dostupnost a šíření znalostí. Mezi hlavní důvody hromadění znalostí patří osobní motivace, jako je strach z konkurence nebo snaha udržet si kontrolu či zajistit si jistotu zaměstnání (Charlson 2025).

Pro zmírnění těchto problémů je zásadní aktivně reagovat na obavy a strachy lidí. **Vzdělávací kampaně** mohou jednotlivcům pomoci pochopit výhody a příležitosti, které mohou byznys a datové slovníky přinést, a zmírnit tak obavy a odpor. Organizace a společnosti musí poskytnout náležitě školení, podporu a zdroje, aby usnadnily přechod a adaptaci (Raden 2023a).

Inkluzivní rozhodovací procesy, v nichž mají lidé možnost vyjádřit svůj názor a podílet se, mohou rovněž podpořit pocit vlastnictví a snížit obavy. **Otevřená komunikace, transparentnost** a jasné vysvětlení cílů a přínosů iniciativ mohou jednotlivcům pomoci pochopit potřebu změny a zmírnit obavy. Uznáním obav jednotlivců a včleněním se do nich mohou organizace a společnosti zavést strategie, které usnadní hladší přechod a umožní lidem přijmout nové technologie a postupy (Raden 2023a).

4.3.3 Kultura

Význam podnikové kultury už byl v tomto dokumentu několikrát nastíněn. DG je především o úpravě chování a o změně způsobu, jakým organizace přemýšlí o datech a jak s nimi zachází. Podle (Ladley 2012) je **správa dat v zásadě iniciativou řízení změn**, nikoliv pouze technickým nebo procesním projektem. Organizace často selhávají v úsilí o správu dat, protože podceňují potřebu kulturní transformace.

Správa dat je výzvou pro stávající způsoby práce, což může vést k odporu. Vyžaduje, aby zaměstnanci vnímali data jako sdílené firemní aktivum, nikoli jako něco, co vlastní jednotlivá oddělení. **Změna organizační kultury vyžaduje jasnou a důslednou komunikaci** o tom, proč je správa důležitá. Zaměstnanci musí chápat, jak řízení zlepšuje jejich každodenní práci, a ne je považovat za překážku. Školení a semináře pomáhají lidem osvojit si zásady governance a budovat kulturu, v níž se s daty zachází jako se strategickým aktivem (Ladley 2012).

DG program by se měl přizpůsobit stávající kultuře organizace, a ne se ji snažit přímo změnit. Modely správy dat by měly být navrženy tak, aby odpovídaly způsobu práce zaměstnanců, a přirozeně integrovaly činnosti správy, nikoli vnucovaly nové rigidní procesy. Úspěšná správa dat není jen o pravidlech a technologiích – je to o změně způsobu, jakým si organizace cení dat a jak je spravuje (Ladley 2012).

Stejně jako jiné snahy o správu dat se iniciativy v oblasti metadat často setkávají s kulturním odporem. Přechod od neřízeného ke spravovanému prostředí metadat **vyžaduje práci a disciplínu**. Není to snadné, ani v případě, kdy většina lidí uznává hodnotu spolehlivých metadat (DAMA International 2017).

4.4 Organizační a procesní výzvy

4.4.1 Nedostatečná podpora zúčastněných stran

Implementace byznys a datových slovníků, jako i dalších iniciativ správy dat, vyžaduje silnou podporu všech zúčastněných stran. Jedním z důvodů neúspěchu mnoha iniciativ v oblasti správy dat je **absence silného sponzoringu ze strany vedení**. Bez viditelné podpory ze strany vrcholového vedení nemají

iniciativy v oblasti správy dat často potřebné zdroje, autoritu a vedení, aby mohly uspět. Nedostatečná podpora je často důsledkem toho, že vedoucí pracovníci nechápou význam dané iniciativy správy dat a roli, kterou hraje při zajišťování integrity, dostupnosti, bezpečnosti a použitelnosti dat. Bez tohoto pochopení považují správu za iniciativu s nízkou prioritou, která neospravedlňuje vynaložení času a zdrojů, což ztěžuje zahájení smysluplného úsilí o správu (DAMA International 2017; Sheldon 2024).

Zastánci governance musí přesvědčit vedení, že absence **komplexní strategie governance** by organizaci mohla potenciálně stát více, než kolik je potřeba na její zavedení a udržení. Kromě vedoucích pracovníků je však nutné zapojit všechny zúčastněné strany, které se podílejí na tvorbě nebo používání dat. Nestačí pouze shrnout rizika, **zaměstnanci musí přímo vidět hodnotu či výhody**, které jim iniciativy přinesou. V případě byznys a datových slovníků je to především sdílené chápání dat, zajištění konzistence, přesnosti a důvěryhodnosti každodenně používaných informací nebo zjednodušení komunikace a spolupráce mezi rozdílnými týmy. **Získání podpory ze strany vedoucích zaměstnanců** je tedy důležité, neméně důležité je rovněž vhodně **komunikovat přínosy** dokumentace byznys a datových elementů všem zúčastněným stranám a nabídnout prostor pro aktivní zapojení do dokumentačního procesu (Moutinho 2024; Redman et al. 2025).

4.4.2 Nerealistická očekávání

Při definování rozsahu implementace byznys a datových slovníků je třeba si uvědomit, že podobně jako celá správa dat i dokumentace obchodních a technických metadat je **dlouhodobý a trvalý proces**, nikoliv jednorázový projekt. Organizace, které očekávají rychlé výsledky mohou mít snahu vyřešit všechny své problémy najednou. Při popisu obchodních a datových elementů je nutný **inkrementální přístup**, kdy správné části budou dodány ve správný okamžik. Vytváření metadat a dokumentace jen proto, aby byla vytvořena málokdy funguje dobře. Proto je důležité hledat ten správný příklad, kdy tyto aktivity přinesou okamžitý přínos pro kvalitu dat ve firmě (DAMA International 2017; Mohsin 2023; Schmidbaur 2022).

Nerealistická očekávání mohou rovněž plynout z mylné představy, že implementace DG a jeho iniciativ je jednodušší, než je tomu ve skutečnosti. To může být způsobeno množstvím všemožných publikací, které nadměrně propagují správu dat jako jednodušší a účinnější, než je ve skutečnosti a společnosti mohou nabýt dojmu, že DG program spočívá pouze v jednoduchém dodržení souboru logických kroků. Iniciativy v oblasti správy dat mohou být často prezentovány jako iniciativy s vysokým transformačním potenciálem a slibují rychlou a jasnou návratnost investic. **Dosažení těchto výsledků je však mnohem složitější** a takovýto nadsázkový prodej vede k nespokojenosti a skepsi, když se tyto sliby nenaplní (Redman et al. 2025; Gupta a Cannon 2020).

4.4.3 Procesy a řízení změn

Vzhledem k tomu, že správa dat je trvalý proces, který se neustále vyvíjí, upravuje, vylepšuje, je nezbytné nastavit správné procesy pro řízení těchto změn. I v případě **řízení a správy metadat** je třeba **řídit změny v jejich životním cyklu**. Termíny procházejí různými stavy od návrhu, přes schválení, zveřejnění až po vyřazení a všechny tyto stavy je nutné řídit a dokumentovat, protože metadata fungují jako jediný zdroj pravdy pro datová aktiva a zajištění jejich aktuality podporuje důvěru v data (Gupta a Cannon 2020).

Na druhou stranu je třeba **zásadní procesy vhodně strukturovat a komunikovat**. Zejména je třeba si dát pozor, aby proces řízení změn nebyl zbytečně robustní a časově náročný, protože to může vyvolat frustraci při snaze aktivně udržovat termíny a tím způsobit, že uživatelé nebudou pravidelné změny provádět, což má za důsledek nekonzistence mezi aktuálním stavem slovníků a realitou.

V mnoha případech procesy a zásady nejsou pečlivě naplánovány, zdokumentovány ani distribuovány lidem, kteří je potřebují vidět a pochopit. Lidé, kteří data vlastní a pracují s nimi, mohou zásady a postupy považovat za matoucí, neúplné či rozporuplné. **Proces definování jasných a stručných zásad a procesů je náročný** úkol, protože musí být hladce integrovány do stávajících podnikových procesů, a proto často vyžadují souhlas různých klíčových zúčastněných stran. Poté, co jsou přijaty, může být náročné přimět všechny, aby je dodržovali, zejména v kultuře, kde je správa dat považována spíše za nepříjemnost než za prioritu (Sheldon 2024).

Pro zajištění úspěchu je zásadní definovat a **správně komunikovat zásady, procesy a postupy**. Klíčové zúčastněné strany by se měly podílet na procesu tvorby programu, který by měli obdržet a důkladně pochopit všichni, kdo s daty pracují. Musí také pochopit, že začlenění zásad do stávajících

operací je průběžné úsilí, které vyžaduje čas. Po implementaci by měly zásady a postupy procházet pravidelnými revizemi a podle potřeby by měly být aktualizovány, aby odpovídaly měnícím se obchodním požadavkům. Zároveň je vhodné **vytvářet procesy a zásady postupně podle potřeby**. Vzhledem ke zmíněné komplexní přípravě a začlenění může být obtížnější vyřadit existující a schválený proces než definovat nový (Sheldon 2024).

4.4.4 Obtížnost kvantifikace přínosů

Už bylo zmíněno, že je třeba **správně komunikovat přínosy data governance a jejich iniciativ**, aby zúčastněné strany viděli benefity plynoucí z jejich účasti v těchto procesech. Společnosti se však někdy mohou dostat do situace, kdy sami nedokáží přesně prokázat dlouhodobé hodnoty plynoucí z tohoto programu, a to i v případě, že je potřeba správy dat akceptována (Redman et al. 2025).

Kvantifikace přínosů správy dat je náročná, protože mnoho jejích výhod je nepřímých, dlouhodobých a obtížně oddělitelných od jiných obchodních zlepšení. Pokud například analytika vede ke zvýšení příjmů, jak tento úspěch přiřadit správě dat? (Redman et al. 2025)

Na rozdíl od okamžitých úspor nákladů jsou výsledky správy dat – jako je lepší kvalita dat, dodržování předpisů a lepší rozhodování – **často nehmotné a těžko měřitelné** ve finančním vyjádření. Dopad se kumuluje v čase, takže je obtížné přiřadit konkrétní obchodní přínosy přímo k úsilí v oblasti správy a řízení. Mnohé zúčastněné strany jsou tímto nesouladem frustrovány a tyto obtíže s kvantifikací a přiřazením hodnoty komplikují zdůvodnění průběžných investic do správy dat (Redman et al. 2025).

4.5 Technologické výzvy

Mnoho organizací se mylně domnívá, že na vyřešení veškerých problémů stačí pouze nákup technologických řešení. Toto přílišné spoléhání často zastiňuje význam lidských a procesních prvků (Redman et al. 2025). IT programy mohou pouze aplikovat pravidla, která jsou předem definována organizací – mohou identifikovat případné odchylky od standardu a poskytnout nástroje pro měření kvality dat (Schmidbaur 2022). Organizace **může investovat spoustu finančních prostředků k zakoupení správných nástrojů**, pokud však nebude disponovat správnými procesy, a především správnými lidmi, nástroje zůstanou nevyužité nebo špatně nasazené (Ladley 2012).

Na druhou stranu je důležité zmínit, že technologie jsou **klíčovým pomocníkem** při dosahování cílů iniciativy správy dat. Užití vhodných nástrojů usnadňuje komunikaci a spolupráci, pomáhá automatizovat úkoly a zvýšit efektivitu a motivovat jedince či týmy tím, že zjednoduší jejich každodenní aktivity. Tento fakt je důležitý, protože **motivace je významnou výzvou z hlediska lidské komponenty** data governance a přespříliš manuálních operací může vyvolat odpor (Mahanti 2021).

Při správě a implementaci byznys a datových slovníků jsou **nástroje a technologie rovněž důležité**, protože užití vhodných nástrojů **umožňuje lepší správu metadat**. Pokud by k aktualizaci metadat docházelo manuálně, hrozilo by, že rychle zastarají. Díky nástrojům lze nastavit různé automatizované kontrolní mechanismy, měřit kvalitu a úplnost dostupných dat, objevovat nová data a řídit změny a verze při dokumentaci obchodních i technických dat. Zároveň je klíčové **zajistit integraci mezi obchodními a technickými metadaty**, aby bylo možné pozorovat vazby mezi nimi a tím umožnit uživatelům slovníky aktivně používat.

V rámci implementace byznys a datových slovníků a konceptu data governance lze narazit na celou řadu výzev či bariér, které mohou zpomalit či bránit úspěšnému nasazení tohoto programu. Výsledkem této kapitoly není exhaustivní výčet všech překážek, nýbrž výběr těch, které se zdají napříč rozdílnou literaturou být nejvíce relevantní a je vhodné na ně upozornit.

5. Zdroje

- ALATION, 2024. *What Is a Data Catalog? Importance, Benefits & Features* | Alation [online] [vid. 2025-02-04]. Dostupné z: <https://www.alation.com/blog/what-is-a-data-catalog/>
- ALLEMANN, Gary, 2023. Defining Business Terms: A Best-Practice Guide. *Data Quality Matters* [online]. [vid. 2025-02-06]. Dostupné z: <https://blog.masterdata.co.za/2023/05/23/defining-business-terms-a-best-practice-guide/>
- ALMEIDA, Jose, 2024. Why does Data Governance fail? *LinkedIn* [online] [vid. 2025-02-09]. Dostupné z: <https://www.linkedin.com/pulse/why-does-data-governance-fail-jose-almeida-zwtkf/>
- AMARESAN, Swetha, 2019. Data Governance (DG): A Straightforward Guide. *HubSpot Blog* [online]. [vid. 2024-11-22]. Dostupné z: <https://blog.hubspot.com/marketing/data-governance>
- ANDERSON, Carl, 2015. *Creating a Data-Driven Organization: Practical Advice from the Trenches*. B.m.: O'Reilly Media, Inc. ISBN 978-1-4919-1688-9.
- ASSOCIATION OF BRITISH INSURERS, [b.r.]. Insurance Terms Glossary. *ABI* [online] [vid. 2025-03-12]. Dostupné z: <https://www.abi.org.uk/data-and-resources/tools-and-resources/glossary/>
- ATLAN, 2013. 8 Best Data Governance Books For Your 2024 Reading List. *Atlan* [online] [vid. 2024-11-22]. Dostupné z: <https://atlan.com/data-governance-books/>
- ATLAN, 2023a. *Data Dictionary vs Business Glossary: The TL;DR Version* [online] [vid. 2024-12-02]. Dostupné z: <https://atlan.com/data-dictionary-vs-business-glossary/>
- ATLAN, 2023b. *Data Governance vs Business Glossary: Are They Connected?* [online] [vid. 2024-12-02]. Dostupné z: <https://atlan.com/data-governance-vs-business-glossary/>
- BASSI, Carlos a Solange ALVES-SOUZA, 2023. Challenges to Implementing Effective Data Governance: A Literature Review: In: *15th International Conference on Knowledge Management and Information Systems: Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* [online]. Rome, Italy: SCITEPRESS - Science and Technology Publications, s. 17–28 [vid. 2024-11-22]. ISBN 978-989-758-671-2. Dostupné z: doi:10.5220/0012185900003598
- BENFELDT, Olivia, John Stouby PERSSON a Sabine MADSEN, 2020. Data Governance as a Collective Action Problem. *Information Systems Frontiers* [online]. 22(2), 299–313. ISSN 1387-3326, 1572-9419. Dostupné z: doi:10.1007/s10796-019-09923-z
- BNP PARIBAS CARDIF, 2023. O BNP Paribas Cardif Pojišťovně. *O nás* [online] [vid. 2024-11-03]. Dostupné z: <https://cardif.cz/cs/web/czech-republic/o-bnp-paribas-cardif>
- BNP PARIBAS CARDIF, [b.r.]. O BNP Paribas Cardif Services. *BNP Paribas Cardif Česká republika* [online] [vid. 2025a-02-23]. Dostupné z: <https://cardif.cz/o-bnp-paribas-cardif-services/>
- BNP PARIBAS CARDIF, [b.r.]. Průvodce pojištěním - BNP Paribas Cardif Pojišťovna. *BNP Paribas Cardif Česká republika* [online] [vid. 2025b-02-23]. Dostupné z: <https://cardif.cz/pruvodce-pojistenim/>
- BREZINA, 2024. Na co potřebuji pojištění schopnosti splácet hypotéku, když se pojistím dostatečně životním pojištěním? *BNP Paribas Cardif Česká republika* [online] [vid. 2025-02-23]. Dostupné z: <https://cardif.cz/poradna-bnp-paribas-cardif-pojistovny-na-co-potrebuji-pojisteni-schopnosti-splacet-hypoteku-kdyz-se-pojistim-dostatecne-zivotnim-pojistenim/>
- CASTORDOC, 2024. *Understanding the Role of Data Dictionary in Data Governance* [online] [vid. 2024-12-02]. Dostupné z: <https://role-of-data-dictionary-in-data-governance>
- ČESKÁ ASOCIACE POJIŠŤOVEN, [b.r.]. Anglicko-český cizojazyčný slovník pojmů. *ČAP* [online] [vid. 2025-03-12]. Dostupné z: <https://cap.cz/slovnikacs>
- DAMA INTERNATIONAL, 2017. *DAMA-DMBOK: Data Management Body of Knowledge*. B.m.: Technics Publications. ISBN 978-1-63462-236-3.
- DGPO, [b.r.]. What is Data Governance? *Data Governance Professionals Organization* [online] [vid. 2024-11-20]. Dostupné z: <https://dgpo.org/page/2/>
- DREMIO, 2023a. Business Metadata. *Dremio* [online] [vid. 2025-01-16]. Dostupné z: <https://www.dremio.com/wiki/business-metadata/>
- DREMIO, 2023b. Data Governance. *Dremio* [online] [vid. 2024-11-20]. Dostupné z: <https://www.dremio.com/wiki/data-governance/>
- DRORI, Amnon, 2019. Data Dictionary vs. Business Glossary. *TDAN.com* [online]. [vid. 2024-12-02]. Dostupné z: <https://tdan.com/data-dictionary-vs-business-glossary/24652>
- EIOPA, 2015. Solvency II. *European Insurance and Occupational Authority* [online] [vid. 2025-02-27].

Dostupné z: https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en

ERYUREK, Evren, Uri GILAD, Valliappa LAKSHMANAN, Anita KIBUNGUCHY-GRANT a Jessi ASHDOWN, 2021. *Data Governance: The Definitive Guide*. United States: O'Reilly Media Inc. ISBN 978-1-4920-6349-0.

EWSOLUTIONS, 2013. Foundations of Enterprise Data Management Framework. *EWSolutions* [online] [vid. 2024-11-18]. Dostupné z: <https://www.ewsolutions.com/foundations-enterprise-data-management/>

FRÝBA, Michal, 2020. *Data governance v podnikové praxi* [online]. Praha [vid. 2024-11-03]. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/81201_data-governance-v-podnikove-praxi?title=data+governance&type=Diplomov%C3%A1+pr%C3%A1ce&page=1

GOOGLE, [b.r.]. *Google Scholar* [online] [vid. 2024-11-21]. Dostupné z: <https://scholar.google.com/>

GUPTA, Uma a San CANNON, 2020. *A Practitioner's Guide to Data Governance : A Case-Based Approach* [online]. 1. vyd. B.m.: Emerald Publishing Limited [vid. 2024-10-13]. ISBN 978-1-78973-567-3. Dostupné z: <https://ebookcentral.proquest.com/lib/vsep/reader.action?&docID=6238638&ppg=2&pq-origsite=360>

HOPPER, Mary Anne, 2023. *Practitioner's Guide to Operationalizing Data Governance* [online]. 1. vyd. Newark, UNITED STATES: John Wiley & Sons, Incorporated [vid. 2024-10-13]. ISBN 978-1-119-85143-1. Dostupné z: <http://ebookcentral.proquest.com/lib/vsep/detail.action?docID=7243468>

HOZÁKOVÁ, Eliška, 2024. *Data governance assessment ve vybrané společnosti* [online]. Praha [vid. 2024-11-03]. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/94243_data-governance-assessment-ve-vybrane-spolecnosti?title=data+governance&type=Diplomov%C3%A1+pr%C3%A1ce&page=1

CHARLSON, Tori, 2025. 9 Signs Your Organization Has a Knowledge Hoarding Problem. *Coveo* [online] [vid. 2025-02-06]. Dostupné z: <https://www.coveo.com/blog/knowledge-hoarding/>

IFRS, 2024. IFRS - IFRS 17 Insurance Contracts. *International Financial Reporting Standards* [online] [vid. 2025-02-27]. Dostupné z: <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-17-insurance-contracts/>

INDEED EDITORIAL TEAM, 2024a. What Are Documentation Skills? (Definition and Examples). *Indeed Career Guide* [online] [vid. 2025-01-28]. Dostupné z: <https://ca.indeed.com/career-advice/resumes-cover-letters/documentation-skills>

INDEED EDITORIAL TEAM, 2024b. What Are Librarian Skills? Definition And Examples. *Indeed Career Guide* [online] [vid. 2025-01-28]. Dostupné z: <https://in.indeed.com/career-advice/resumes-cover-letters/librarian-skills>

KIMBALL, Ralph a Margy ROSS, 2013. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* [online]. B.m.: John Wiley & Sons. ISBN 978-1-118-73228-1. Dostupné z: https://books.google.cz/books?id=4rFXzk8wAB8C&printsec=frontcover&hl=cs&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

KNIGHT, Michelle, 2019. Developing a Functional Data Governance Framework. *DATAVERSITY* [online] [vid. 2024-11-20]. Dostupné z: <https://www.dataversity.net/developing-a-functional-data-governance-framework/>

KUMAR, Ashok a A. ANSARI, 2016. Redefined and Importance of Organizational Culture. *Global Journal of Management and Business Research* [online]. [vid. 2025-01-29]. Dostupné z: <https://www.semanticscholar.org/paper/Redefined-and-Importance-of-Organizational-Culture-Kumar-Ansari/71a5a7d68f48af666915708f5cac4f503c3f17e3>

LADLEY, John, 2012. *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. 1. vyd. B.m.: Morgan Kaufmann. ISBN 978-0-12-415829-0.

LAKE, Peter a Paul CROWTHER, 2013. Data, an Organisational Asset. In: Peter LAKE a Paul CROWTHER, ed. *Concise Guide to Databases: A Practical Introduction* [online]. London: Springer, s. 3–19 [vid. 2024-11-18]. ISBN 978-1-4471-5601-7. Dostupné z: [doi:10.1007/978-1-4471-5601-7_1](https://doi.org/10.1007/978-1-4471-5601-7_1)

LIM, Jason, 2021. Data Dictionary vs. Business Glossary. *Alation* [online] [vid. 2024-12-03]. Dostupné z: <https://www.alation.com/blog/data-dictionary-vs-business-glossary/>

MAHANTI, Rupa, 2021. *Data Governance Success: Growing and Sustaining Data Governance*. Australia: Springer. ISBN 9789811650857.

MOHSIN, Hashmi Sayed, 2023. Top Reasons Why Data Governance Initiatives Fail to Generate Value for Organizations. *LinkedIn* [online] [vid. 2025-02-09]. Dostupné

- z: [https://www.linkedin.com/pulse/glossary-business-terms-challenges-strategies-adriana-moutinho-5vyrf/](https://www.linkedin.com/pulse/top-reasons-why-data-governance-initiatives-fail-generate-hashmi/MOUTINHO, Adriana, 2024. Glossary of Business Terms - Challenges and strategies for successful documentation. <i>LinkedIn</i> [online] [vid. 2025-02-09]. Dostupné z: <a href=)
- MVČR, [b.r.]. Co je GDPR - Ochrana osobních údajů. *Ministerstvo vnitra České republiky* [online] [vid. 2025-02-27]. Dostupné z: <https://mv.gov.cz/gdpr/clanek/co-je-gdpr.aspx>
- NAEEM, Tehreem, 2024. Data Warehouse Concepts: Kimball vs. Inmon Approach. *Astera* [online]. [vid. 2025-03-30]. Dostupné z: <https://www.astera.com/type/blog/data-warehouse-concepts/>
- NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS, [b.r.]. Glossary of Insurance Terms. *NAIC* [online] [vid. 2025-03-12]. Dostupné z: <https://content.naic.org/glossary-insurance-terms>
- OKOLI, Chitu a Kira SCHABRAM, 2010. *A Guide to Conducting a Systematic Literature Review of Information Systems Research*. 2010. B.m.: Sprouts: Working Papers on Information Systems.
- O'SULLIVAN, Arthur a Steven M. SHEFFRIN, 2007. *Economics Principles in Action* [online]. United States: Washington, DC: Pearson Prentice Hall [vid. 2024-11-16]. ISBN 978-0-13-063085-8. Dostupné z: <https://www.scribd.com/doc/55444148/Economics-Principles-in-Action-Arthur-Sullivan-Steven-M-Sheffrin>
- OTTO, Boris, 2011. A Morphology of the Organisation of Data Governance. In: *19th European Conference on Information Systems, ECIS 2011* [online]. Dostupné z: https://www.researchgate.net/publication/221407900_A_morphology_of_the_organisation_of_data_governance
- PANSARA, Ronak, 2023. Unraveling the Complexities of Data Governance with Strategies, Challenges, and Future Directions. *Transactions on Latest Trends in IoT*. **6**(6), 46–56. ISSN 3246-547X.
- PARNAS, David Lorge, 2011. Precise Documentation: The Key to Better Software. In: Sebastian NANZ, ed. *The Future of Software Engineering* [online]. Berlin, Heidelberg: Springer, s. 125–148 [vid. 2025-01-28]. ISBN 978-3-642-15187-3. Dostupné z: [doi:10.1007/978-3-642-15187-3_8](https://doi.org/10.1007/978-3-642-15187-3_8)
- PLOTNIKOVA, Oleksandra, 2021. *Implementace business slovníku ve finanční instituci - Vysokoškolské kvalifikační práce* - [online]. Praha [vid. 2024-11-03]. Bakalářská práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/83450_implementace-business-slovníku-ve-finanční-instituci?title=business+slovn%C3%ADk&page=1
- QIC WD, 2020. Introduction to Data Dictionaries. *Quality Improvement Center for Workforce Development* [online] [vid. 2025-02-13]. Dostupné z: <https://www.qic-wd.org/blog/introduction-data-dictionaries>
- RADEN, Neil, 2023a. Want to get digital transformation right? Address what's left behind. *Diginomica* [online] [vid. 2025-02-06]. Dostupné z: <https://diginomica.com/want-get-digital-transformation-right-address-whats-left-behind>
- RADEN, Neil, 2023b. Want to solve data governance failures? First, move past nonsensical advice. *Diginomica* [online] [vid. 2025-02-06]. Dostupné z: <https://diginomica.com/solve-data-governance-failures-past-nonsensical-advice>
- REDMAN, Thomas C., Anne Marie SMITH, John LADLEY, Mathias VERCAUTEREN, Malcolm HAWKER a Aaron WILKERSON, 2025. Data Governance Is Failing — Here's Why. *CDO Magazine* [online] [vid. 2025-02-05]. Dostupné z: <https://www.cdomagazine.tech/opinion-analysis/data-governance-is-failing-heres-why>
- SALMI, Christina, 2021. What is the Difference Between A Business Glossary, A Data Dictionary, and A Data Catalog, and How Do They Play A Role In Modern Data Management? *Analytics8* [online] [vid. 2024-12-03]. Dostupné z: <https://www.analytics8.com/blog/data-dictionary-vs-data-catalog-vs-business-glossary/>
- SEINER, Robert S., 2014. *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success* [online]. B.m.: Technics Publications. ISBN 978-1-63462-045-1. Dostupné z: https://books.google.cz/books?id=Xo3XBgAAQBAJ&printsec=frontcover&hl=cs&source=gbg_ge_summary_r&cad=0#v=onepage&q&f=false
- SHELDON, Robert, 2024. 10 data governance challenges that can sink data operations. *TechTarget* [online] [vid. 2025-02-09]. Dostupné z: <https://www.techtarget.com/searchdatamanagement/tip/Data-governance-challenges-that-can-sink-data-operations>
- SCHMIDBAUR, Sabine, 2022. 5 Reasons Why Data Governance Initiatives Fail. *Stibo Systems* [online] [vid. 2025-02-09]. Dostupné z: <https://www.stibosystems.com/blog/five-reasons-your-data-governance-initiative-could-fail>
- SIMSION, Graeme C. a Graham C. WITT, 2004. *Data modeling essentials*. 3. ed., transferred to digital

print. Amsterdam: Morgan Kaufmann. ISBN 978-0-12-644551-0.

SLOUKOVÁ, Anna, 2008. *Postup zavádění Data Governance* [online]. Praha [vid. 2024-11-03]. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/14528_postup-zavadeni-data-governance?title=data+governance&type=Diplomov%C3%A1+pr%C3%A1ce&page=1

TEAL, 2023a. *Data Architect Skills* [online] [vid. 2025-01-22]. Dostupné z: <https://www.tealhq.com/skills/data-architect>

TEAL, 2023b. *Data Modeling Skills* [online] [vid. 2025-01-22]. Dostupné z: <https://www.tealhq.com/skills/data-modeling>

VÁCLAVÍK, David, 2024. *Analýza přínosu nasazení business data governance* [online]. Praha [vid. 2024-11-03]. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/92554_analyza-prinosu-nasazeni-business-data-governance?title=data+governance&type=Diplomov%C3%A1+pr%C3%A1ce&page=1

VŠE, [b.r.]. *E-zdroje – Knihovna VŠE – Vysoká škola ekonomická v Praze* [online] [vid. 2024a-11-21]. Dostupné z: <https://knihovna.vse.cz/zdroje/>

VŠE, [b.r.]. *Vysokoškolské kvalifikační práce - Vysoká škola ekonomická v Praze* [online] [vid. 2024b-11-21]. Dostupné z: <https://vskp.vse.cz/>

WALERY, Jim, 2020. *Key Principles and Thoughts of Writing Good Business Glossary Definitions* [online] [vid. 2025-02-06]. Dostupné z: <https://blog.idatainc.com/principles-writing-business-glossary-definitions>

ZOSINČUK, Dominik, 2014. *Zavádění projektu data governance - Vysokoškolské kvalifikační práce* - [online]. Praha [vid. 2024-11-03]. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/40907_zavadeni-projektu-data-governance?title=data+governance&type=Diplomov%C3%A1+pr%C3%A1ce&page=1



B. Data Mesh

(pracovní dokument)



Michal Novotný

VŠE Praha, 2023

Úvod

Data Mesh je podle jeho autorky (Dehghani 2022) nový **přístup k získávání, správě a přístupu k datům pro analytické účely**, který se řeší rostoucí komplexitu správy dat v organizacích a problémy, se kterými se datové architektury v organizacích potýkají. Těmito problémy jsou například centralizované vlastnictví, škálovatelnost či omezená dostupnost. Koncept Data Mesh **je o zásadním posunu v našich předpokladech, architektuře, technických řešeních a sociální struktuře** našich organizací, ve způsobu, jakým spravujeme, používáme a vlastníme analytická data.

6. Principy

Koncept Data Mesh má dle jeho autorky (Dehghani 2022) **čtyři principy**:

6.1 Federated computational governance

Princip federovaného řízení **zajišťuje interoperabilitu všech datových produktů prostřednictvím standardizace** (Jochen et al. 2023). Ve své podstatě jde o vytvoření rámce a zajistit vzájemnou kompatibilitu datových produktů.

6.2 Self-serve datové platforma

Self-serve datová platforma má za **cíl poskytovat funkce a služby, jako je úložiště nebo výpočetní výkon**. Centrální IT poskytuje platformu, která umožňuje to, že se doménové týmy mohou zaměřit pouze na práci, která je pouze doménově orientovaná, tedy že **se mohou zaměřit pouze na vytváření byznysové hodnoty z dat**, místo toho, aby se také staraly o údržbu nástrojů či infrastruktury (Tartow a Mott 2021b).

6.3 Datové domény

Podle (Tartow a Mott 2021a) **datová doména představuje skupinu lidí v organizaci, kteří jsou organizováni kolem společného obchodního účelu či byznysové entity**. Datové domény obvykle odrážejí organizaci z pohledu byznysu. Tedy například pro e-commerce stránku mohou být datové domény uživatelé, obchodníci, produkty nebo marketing. Princip doménového vlastnictví stanovuje, že **doménové týmy mají převzít zodpovědnost za svá data**. Analytická data jsou tedy strukturována kolem domén, do kterých byznysově (doménově) spadají.

6.4 Datový produkt

Posledním principem je chápání dat jako produktu a aplikování produktového myšlení v rámci organizace, tedy tvorba tzv. datových produktů.

Datový produkt lze definovat takto:

„Datový produkt je kombinace datasetu a produktových metadat, která náleží právě jedné datové doméně, má konkrétního vlastníka a vzniká za určitým účelem s cílem naplňovat potřeby svých konzumentů.“

Datový produkt musí splňovat určité charakteristiky, které jsou vypsány v tabulce 1.

Tabulka 1: Charakteristiky datového produktu (zpracováno autorem podle zdrojů (Dehghani 2022; Majchrzak et al. 2023; Jochen et al. 2023))

Podmínka	Popis
Dohledatelnost a snadná dostupnost	Datový produkt musí být jednoznačně pojmenovaný, musí být vy publikován na stanoveném místě a uživatelé musí vědět, kam se v případě potřeby podívat a kde datový produkt hledat

Srozumitelnost a pochopitelnost	Datový produkt musí být pro jeho konzumenty srozumitelný a pochopitelný, respektive dostatečně (nejen byznysově) popsán
Adresovatelnost	Datový produkt musí adresovatelný
Dostatečná zabezpečení	K datovým produktům musí mít přístup jen ti uživatelé, kteří k němu přístup mít mají
Interoperabilita	Datový produkt musí mít schopnost spolupracovat s dalšími datovými produkty (jeden datový produkt může být zdrojem dalšího datového produktu)
Důvěryhodnost	Datový produkt musí být pro uživatele důvěryhodný, tedy že data jsou správná a že jim může věřit
Hodnota	Datový produkt musí mít hodnotu sám o sobě, nesmí být nutné jeho propojení s dalšími datovými produkty, aby teprve poté dostal hodnotu a smysl
Unikátní doména	Datový produkt patří právě do jedné datové domény, není součástí žádné jiné
Vlastník a vývojový tým	Datový produkt musí mít stanoveného svého vlastníka a vývojový tým, který se stará o vývoj a údržbu datového produktu
Druh	Musí spadat do jednoho z druhů datových produktů
Metadata	O datovém produktu se musí sbírat a uchovávat metadata, která mohou mít různou podobu
Účel	Datový produkt musí vzniknout za jasným účelem a cílem (nemůže se stát, že by datový produkt vznikl jako vedlejší produkt činnosti)
Podmínky konzumace a SLO	Datový produkt musí mít jasně definované podmínky jeho využívání (kdo ho může využívat, kdy, jak)
Vstupní a výstupní porty	Datový produkt musí mít stanovené vstupní porty (přístup ke zdrojům) a výstupní porty (poskytnutí uživatelům)

Obecně platí, že libovolný způsob reprezentace dat, který poskytuje uživatelům nějakou hodnotu, je vhodným kandidátem na to stát se datovým produktem (Majchrzak et al. 2023). Jednoduše řečeno, datový produkt si lze představit **jako dataset splňující výše zmíněné charakteristiky**. Dataset může mít různou podobu, nemá pevně stanovenou formu. Mezi nejlépe představitelnou a zároveň dle autora zdaleka nejčastější podobou datového produktu je libovolný databázový dataset, ať už se jedná o databázovou tabulku, jejich propojení či databázové pohledy.

Datové produkty lze dle jejich orientace rozdělit na **3 základní druhy**:

- datové produkty **zaměřené na zdroj** (source-aligned) – tyto produkty mají za cíl jen přenášet data ze zdrojových systémů (či jiných zdrojů) s minimálními úpravami či transformacemi a poskytovat je ke konzumaci uživatelům (Mott 2022),
- datové produkty **orientované na spotřebitele** či konzumenta (consumer-aligned) - tyto datové produkty přinášejí byznysovou hodnotu díky expertíze a znalostem lidí z oboru (tedy konzumentů),
- **agregované datové produkty** – kombinace (agregace) různých datových produktů z různých domén.

Datový produkt nemůže být datovým produktem, pokud se k němu nesbírají a neuchovávají metadata. Ukázková struktura metadatového minima pro datové produkty je uvedena v tabulce 2.

Tabulka 2: Metadata datového produktu (zpracováno autorem podle zdrojů (Jochen et al. 2023; Majchrzak et al. 2023))

Metadata	Komentář
ID	Unikátní identifikátor datového produktu
Název	Unikátní (biznysový) název datového produktu
Popis	Biznysový popis datového produktu (k čemu slouží a co je jeho obsah)
Vlastník	Vlastník datového produktu (jméno, příjmení, kontakty apod.)
Vývojový tým	Vývojový (údržbový) tým datového produktu (jména, příjmení, kontakty apod.)
Zdroje	Zdrojové systémy a jiné zdroje datového produktu (výpis, případně i popis a odkazy)
Vstupní port(y)	Výpis vstupních portů datového produktu
Výstupní port(y)	Výpis výstupních portů datového produktu
Konzumenti	Výpis současných konzumentů datového produktu
Service-level objective(s)	Výpis SLOs
Podmínky konzumace	Za jakých podmínek lze datový produkt využívat (např. kdo, kdy)
Odkaz	Odkaz na datový produkt
Stav	Zobrazení stavu datového produktu (aktivní, neaktivní apod.)
Datum vytvoření	Datum vytvoření datového produktu (neměnný)
Datum platnosti	Určené datum platnosti datového produktu (mění se v průběhu času)

Verze	Aktuální verze datového produktu
Popis struktury	Popis datasetu datového produktu (popis jednotlivých sloupců, datových typů apod.)

7. Přínosy a rizika Data Mesh

7.1 Přínosy

- díky doménovému vlastnictví dat umožňuje využít naplno agilních způsobů řízení při práci s daty,
- vlastníci a vývojáři mohou rychle a pružně reagovat na změny v oblasti dat,
- vyšší škálovatelnost datového řešení (decentralizace),
- podporuje efektivnější správu dat a lepší spolupráci datových týmů a jednotlivců mezi sebou,
- pomocí publikování datových produktů dělá data (respektive v kontextu konceptu Data Mesh datové produkty) snadno dohledatelná a dostupná pro konzumenty dat k provádění analýz,
- tvorba datových produktů není vzdálena od byznysu a byznysových uživatelů, ba naopak je kladen důraz na propojení byznysové hodnoty,
- jasně definované vlastnictví dat, ale také pevně stanovené odpovědnosti.

7.2 Rizika

- chybí soubor osvědčených postupů, který by implementaci konceptu usnadňoval a usměrňoval,
- konzumenti dat musí dokázat popsat, co přesně potřebují (jaké mají byznysové potřeby),
- nedostatečný důraz na byznysové cíle či problémy,
- nedostatečná komunikace s konzumenty dat,
- velký zásah do firemní kultury (změna architektury, reorganizace a rekvalifikace či nábor zaměstnanců), pravděpodobně doprovázen také složitou implementací.

8. Role – vlastník datového produktu

Vlastník datového produktu je **osoba zodpovědná za nejen byznysové, ale také technické směřování a rozvoj datového produktu**. Tato pracovní role je stěžejní pro úspěšnou implementaci datových produktů. Role vlastníka datového produktu přirozeně vychází z principu, že data jsou vnímána v konceptu Data Mesh jako produkt. Důležitým pravidlem je, že by musí být **pouze jeden vlastník pro konkrétní datový produkt**.

Vlastník je podle (Majchrzak et al. 2023) **odpovědný za**:

- Stanovení důvodu tvorby datového produktu.
- Zjistit očekávání konzumentů a jejich byznysové potřeby.
- Neustálý sběr zpětné vazby.
- Plánování implementace datových produktů a jeho úprav.
- Strategické plánování vývoje datového produktu.

9. Závěry a doporučení

- Koncept Data Mesh je **nový přístup ke správě dat** v organizaci.

- Koncept se snaží řešit **problémy aktuálních datových architektur**, jako je například škálovatelnost či centralizace.
- Mezi **přínosy** patří například efektivnější správa dat, důraz na byznysový kontext, možnost pružně reagovat či vyšší škálovatelnost řešení.
- Mezi **rizika** patří například velký zásah do firemní kultury a architektury, nedostatečná komunikace s konzumenty dat a nedostatečný důraz na byznysové cíle a problémy.
- Koncept má **4 principy** – doménově orientované decentralizované vlastnictví, datové produkty, federated computational governance a self-serve datová platforma.
- Organizace by vždy měla pečlivě **zvážit, zda je pro ni koncept Data Mesh vhodné řešení**, jelikož je časově, finančně a architektonicky náročný na implementaci, a také k němu doposud chybí soubor osvědčených postupů, podle kterého by se mohla organizace řídit.

10. Zdroje

DEHGHANI, Zhamak, 2022. *Data Mesh: delivering data-driven value at Scale*. O'Reilly Media. ISBN 978-1-4920-9239-1.

JOCHEN, Christ, Larysa VISENGERIYEVA a Simon HARRER, 2023. Data Mesh Architecture. Online. Dostupné z: <https://datamesh-architecture.com>. [Citováno: 2023-08-05].

MAJCHRZAK, Jacek, Sven BALNOJAN, Marian SIWIAK a Mariusz SIERACZKIEWICZ, 2023. *Data Mesh in action*. Manning. ISBN 978-1-63343-997-9

TARTOW, Colleen a Andy MOTT, 2021a. *Data Mesh: Domain-oriented Ownership & Architecture*. Online. Starburst. Dostupné z: <https://www.starburst.io/blog/data-mesh-and-starburst-domain-oriented-ownership-architecture/>. [Citováno: 2023-11-20].

TARTOW, Colleen a Andy MOTT, 2021b. *Data Mesh: Self-Service Data Infrastructure*. Online. Starburst. Dostupné z: <https://www.starburst.io/blog/data-mesh-starburst-self-service-data-infrastructure/>. [Citováno: 2023-11-20].

[illegible]

C. Datový katalog

(pracovní dokument)



Nikola Pátková

VŠE Praha, 2023

Úvod

Datový management a zpracování firemních dat se poslední roky stává trendem a bohatstvím každé větší společnosti. Společnosti pomocí svých dat a dat dostupných o konkurenčních společnostech mohou zpracovávat analýzy a **data používat jako podklady pro svá strategická plánování**, mohou tedy tak získat konkurenční výhody, pokud dokáží data dostatečně využít.

Ne vždy je snadné získat data, která jsou k reportingu třeba, právě například ve velkých společnostech, které mají k dispozici mnoho dat na různých místech, z mnoha využívaných služeb a aplikací a bez celkového řízení datových toků.

Z tohoto důvodu je vhodné ve společnostech mít **interní tým expertů v oblasti zpracování dat**, kteří nejen že pomohou dostupná data zpracovat, získat z nich dostupné informace a vytvářet tak různé pohledy pro vedení společnosti, plní však také roli trenérů, kteří učí jiné zaměstnance konceptům Business Intelligence, aby si pohodlně z dostupných předpracovaných dat zvládli připravovat pohledy na data sami.

Data je účelné využívat jako **aktivum společnosti a ta musí být aktivně řízena**, včetně zajištění datové kvality. Řízení dat je složitá disciplína zahrnující všechny úrovně společnosti, a to od návrhu strategie až po provádění obchodních procesů. **Základem řízení dat však zůstává zodpovězení základních otázek**, jako jsou: Co jsou zdroje dat? Jakým způsobem jsou používána? Jaké jsou požadavky na data, aby plně podporovala obchodní procesy? Jak se řídí životní cyklus dat? Jak jsou data propojena mezi sebou? Kde data najít? Kdo je jejich vlastníkem a zodpovědnou osobou?

Dalším z **problémů je neefektivní využívání dat**. Mnoho různých týmů a manažerů využívá data ke svým činnostem, často však nevědí, kde a jak mohou data získat, a tak jsou nuceni hledat informace někdy velmi zdlouhavě. Stejně tak se může stát, že více týmů zpracovává a ukládá data stejným způsobem a data jsou tak zbytečně ukládána a zpracovávána vícekrát. Tuto problematiku by měl datový katalog pomoci vyřešit.

Velmi důležitým aspektem je **pravdivost dat**, tedy důvěra v datech pro zákazníky. Data musí být spolehlivým zdrojem, aby se dala využívat jako důležitý podklad ke strategickým rozhodováním celé společnosti. Nelze ji vybudovat ihned, avšak podpůrné procesy především, sdílené znalosti a programy by k tomu měly pomoci.

1. Analýza současných přístupů k řešení a užití datových katalogů, jejich možnosti a omezení

Tato část je zaměřena na **analýzu současného trhu s datovými katalogy**, která se nejprve soustředí na analýzu přístupů k datovým katalogům, následovanou současnými přístupy k užití datových katalogů a v poslední části je uveden přehled několika dostupných datových katalogů, které se liší právě přístupy.

1.1 Analýza současného trhu

K samotné aplikaci datového katalogu lze využít **4 přístupy řešení**, SaaS, Open Source, komerční a on-premises datové katalogy.

1.1.1 SaaS Cloud-based

Model SaaS (Software jako služba) umožňuje uživatelům připojit se ke cloudovým aplikacím a používat je přes internet. SaaS poskytuje **úplné softwarové řešení**, které lze zakoupit na základě

průběžných plateb od poskytovatele cloudové služby. **Pronajímá se jako možnost použití aplikace** a uživatelé se k ní připojují přes internet obvykle pomocí webového prohlížeče. Veškerá podpůrná infrastruktura, middleware, software a data aplikace jsou umístěné v datovém centru poskytovatele služeb. Poskytovatel služeb spravuje hardware a software a v rámci příslušné smlouvy o poskytování služeb zajišťuje dostupnost a zabezpečení aplikace a dat. SaaS dovoluje organizaci začít aplikaci rychle využívat s minimálními pořizovacími náklady. (Microsoft, 2023).

Výhody:

Toto řešení poskytuje jednodušší přístup odkudkoliv pomocí internetového připojení. Jelikož zákaznická organizace nemá pod správou hardware, tak automatické aktualizace a správa hardware je na straně poskytovatele služeb.

Nevýhody:

Nevýhodou tohoto řešení mohou být vysoké poplatky za využívání služby, i když zrovna aktivně nemusí být používána, nebo závislost na kvalitě a rychlosti internetového připojení.

1.1.2 Open source

Open source je software či produkt se zdrojovým kódem, který kdokoliv může kontrolovat, upravovat a vylepšovat. Původně termín vznikl v souvislosti se specifickým přístupem k vytváření počítačových programů, dnes však pojem označuje **širší soubor hodnot, kdy zahrnují principy otevřeného vývoje, výměny, spolupráce, rychlého vytváření prototypů, transparentnosti a komunitně orientovaného rozvoje**.

V tomto přístupu je tedy zákazníkům k dispozici zdrojový kód aplikace a zákazník kromě samotné implementace se tak může podílet na samotném rozvoji produktu. Často je na straně poskytovatele zřízeno **fórum, kde zákazníci, především vývojáři, mohou sdílet svou zpětnou vazbu** či dokonce své rozšířené verze produktu, návrhy opravy chyb a jiné. (Opensource.com, 2023)

Výhody:

Mezi výhody řešení tohoto typu patří rozhodně cena, jelikož nejsou zde žádné pořizovací náklady a jsou vysoce přizpůsobitelné díky přístupu ke zdrojovému kódu.

Nevýhody:

Ke vhodnému nastavení a správě je potřeba dostatečné technické znalosti, nemusí vždy poskytovat dostatečnou zákaznickou podporu, což se může zařadit mezi nevýhody. (Klipfolio, 2023)

1.1.3 Commercial Data Catalogs

Komerční datové katalogy jsou **placená řešení často nabízená s přidavnými funkcemi** a zákaznickou podporou. Jsou nejvhodnější pro větší společnosti, které vyžadují pokročilé funkce a dedikovanou podporu.

Výhody:

Výhodou tohoto typu řešení je, že poskytuje komplexní funkce, nástroje a vyhrazenou zákaznickou podporu.

Nevýhody:

Mezi nevýhody tohoto řešení lze řadit vyšší cenu a počet funkcí, které jsou s datovým katalogem implementovány, avšak nevyužité. (Klipfolio, 2023)

1.1.4 On-premises Data Catalogs

On-Premises datové katalogy jsou **instalovány a provozovány na vlastních serverech** dané společnosti, což je vhodným řešením pro společnosti, jejichž prioritou je kontrola a vlastní infrastruktura.

Výhody:

Výhodou tohoto typu je plná kontrola nad datovým katalogem a jeho zabezpečení, především co se týká hardwarové vrstvy, stejně tak jako že neexistuje přílišná závislost na kvalitě internetového připojení.

Nevýhody:

To, co pro jiné může být výhodou lze však zařadit i mezi nevýhody, jelikož hardwarové vybavení musí být tedy umístěno a spravováno zákaznickou společností a finanční náklady na pořízení a správu tak nemusí být vždy výhodné. Stejně tak jako prvotní nastavení datového katalogu, které může být nejen tedy drahé, ale také velmi náročné. (Klipfolio, 2023)

Datové katalogy tedy jako software mohou být nabízeny několika způsoby a je tedy na zákazníkovi si vybrat, který přístup mu vyhovuje nejvíce. Některé mohou být také **nabízeny jako součást balíčku produktů, či samostatně**. Lze se tedy rozhodovat na základě zabezpečení, ceny, dostupnosti zákaznické podpory, technických dovedností na straně zákazníka či dispozice hardware.

1.2 Současné přístupy k užití datových katalogů

Datový katalog lze přirovnat k existujícímu systému v knihovnách. Je to **v podstatě soubor metadat o dostupných datech**. Avšak aby data a informace získávané z nich byly dostatečně věrohodné, je potřeba nastavit proces, který důvěru uživatelům poskytne. V praxi se ustálil **3 krokový mechanismus**:

1. **Find:** Nejprve analyzovat data, kterými společnost disponuje, uvést jejich lokaci, datové typy a jiné užitečné metainformace.
2. **Explore:** Následuje objasnění vzniku dat a jejich reprezentace a kvalita.
3. **Trust:** Ohodnocení dat na základě vstupů z předcházejících kroků a zároveň vystavení procesu kontinuálního zlepšování, který zajistí zisk zpětné vazby od koncových uživatelů na poskytovaná data, případně odstraní identifikované kvalitativní nedostatky. (Lukšík, 2020)

Datový katalog má za úkol splnit **3 hlavní úkoly**:

1. *Dataset searching.*
2. *Dataset Evaluation.*
3. *Data Access.*

Dataset Searching pomáhá uživatelům **s hledáním požadovaných dat na základě klíčových slov, tagů, businessových výrazů** a vrací výsledky na základě relevance a četnosti. Vyhledávání pomocí přirozeného jazyka umožňuje méně technickým uživatelům najít svá data.

Dataset Evaluation, neboli vyhodnocování datasetu, slouží k **vyhodnocování dat na první pohled** a bez omezeného přístupu. Proto by lidé měli mít možnost zobrazit náhled datasetu, číst uživatelské recenze a anotace kurátorů, zobrazit podrobnosti o kvalitě dat a jejich metadata.

Data Access, tedy přístup k datům, zastává cestu od vyhledávání dat, jejich vyhodnocování až po přístup k nim, která by měla být bezproblémová. Součástí datového katalogu by měly být přístupové protokoly, aby byla zajištěna bezpečnost, soukromí a používání citlivých údajů v souladu s předpisy. (Wells, 2020)

1.2.1 Proč implementovat

Implementace datového katalogu v první řadě pomáhá zmapovat dostupná data ve společnosti za účelem jejich zpřehlednění. Dalším přínosem je **lepší dostupnost informací o datech** ve společnosti pro všechny zainteresované zákazníky, kterými jsou zaměstnanci společnosti, především businessoví a datoví analytici. Ti tato data analyzují za účelem jejich pochopení a nalezení nejlepšího využití. (Wells, 2020). Dalším z důvodů, proč implementovat datový katalog je **vize zlepšení kvality dat**.

Pojem datová kvalita může být definován jako „the degree to which a set of characteristics of data fulfills requirements“. (Transition Support, 2023)

Datová kvalita může být měřena dle metrik, definovaných v práci „Metriky, monitoring a řídicí proces Data Governance“. Jedná se o 6 identifikovaných metrik, kterými jsou Přesnost, Úplnost, Aktuálnost, Konzistence, Včasnost a Nestálost. Obecné příklady metrik jsou inspirovány zdrojem a vlastními zkušenostmi autorky.

1. Přesnost (Precision)

- a. Přesnost odráží **míru, jakou data správně odráží skutečnost**. Problémy s přesností si můžeme představit například syntaktické, jako jsou překlepy, nebo semantické, jakožto vyplnění hodnoty do nesprávného sloupce.
- b. Obecný příklad metriky:
 - i. Počet hodnot v neodpovídajícím sloupci.
- c. Konkrétní příklad:
 - i. V tabulce zaměstnanců se ve sloupci job_title nachází 17 % datumů, které neodpovídají žádné hodnotě označení pracovní pozice a svým formátem vyčnívají.

2. Úplnost (Completeness)

- a. Úplnost zobrazuje, **jak kompletně jsou vyplněny jednotlivé atributy**. V současné praxi není obvyklé rozlišovat mezi povinnými a nepovinnými atributy, a tak často může prázdná hodnota (NULL) dávat smysl, případně alternativní hodnota, jako „1900-01-01“ pro datumové hodnoty či „Neuvedeno“ pro textové. Existují i tzv. „nice to have“ atributy, kde je prázdná hodnota z technického hlediska povolena, avšak z hlediska podniku může být její vyplnění považováno za důležité. Při analýze úplnosti lze do negativního hodnocení úplnosti zahrnout i alternativní hodnoty, jejichž příklady byly uvedeny výše.
- b. Obecný příklad metriky:
 - i. Výskyt „1900-01-01“ u nenulových atributů (požadováno vyplnění).
- c. Konkrétní příklad metriky:
 - i. 60 % tréninkových dat nemá vyplněný začátek tréninku.

3. Aktuálnost (Currentness)

- a. Aktuálnost dat je definována **časem mezi změnou informací o sledovaném objektu na zdrojovém systému a její změnou v datech**.
- b. Obecný příklad metriky:
 - i. Časový interval mezi aktualizací dat ze zdrojového systému a replikou v datech.
- c. Konkrétní příklad:
 - i. Doba trvání přiřazení nově vytvořených účtů k definovaným rolím v interním systému je v 8 % případů až 48 hodin.

4. Konzistence (Consistency)

- a. Konzistence udává, **jak jednotný je obsah dat o jednom unikátním objektu napříč různými částmi organizace**. Optimálního stavu je dosaženo při existenci jediné platné verze dat označované jako „verze pravdy“ (*single point of truth*).
- b. Obecný příklad metriky:
 - i. Počet různých verzí informací o zaměstnanci společnosti napříč systémy udržující data o zaměstnancích.
- c. Konkrétní příklad:
 - i. Z celkových 5 systémů udržující data o zaměstnancích má paní Pátková ve 2 systémech lišící se informace o týmu v rámci organizace, jehož je v současné době součástí.

5. Včasnost (Timeliness)

- a. Včasnost je dimenze, která je orientovaná na čas, sleduje, **zda uživatelé mají svá data dostupná včas k tomu, aby mohli správně a včas provádět rozhodnutí**.

- b. Obecný příklad metriky:
 - i. Procentuální podíl situací, kdy manažeři oblastí podnikání nedostanou včas informace o aktuálním počtu zaměstnanců v jejich příslušné oblasti v porovnání s celkovým počtem sledovaných časových okamžiků.
- c. Konkrétní příklad:
 - i. V 90 % případů dostane manažer oblasti podnikání informaci o aktuálním počtu svých zaměstnanců včas, kdy v rámci měsíce bylo provedeno sledování aktuálního stavu na denní bázi a z 30, kdy ve 27 dnech byla obdržena data včas.

6. Nestálost (Volatility)

- a. Nestálost **sleduje frekvenci změn dat v čase**. Ačkoliv ji lze považovat spíše jako dimenzi, která slouží jako základ pro podpůrné metriky, může indikovat klíčové problémy a být důležitým aspektem například v případě fraud monitoringu, což je sledování podvodů, kde může sloužit k identifikaci potenciálních nepravostí.
- b. Obecný příklad metriky:
 - i. Počet změn hodnoty atributu za vybrané časové období.
- c. Konkrétní příklad:
 - i. Hodnota grade se u paní Pátkové za poslední týden změnila 8x.

(Vacek, 2014)

Podobně jako u prediktivní analytiky v podnikové informatice, datový katalog může napomoci v těchto oblastech, jakými jsou Analýza dostupných zdrojů dat, Organizace, integrace a čištění dat a Identifikace relevantních vztahů. (Pour, 2017)

V případě **analýzy dostupných zdrojů je kromě přehledu vyhodnocována i jejich kvalita, dostupnost, konzistence a konsolidace**. Pro organizaci, integraci a čištění dat datový katalog může podpořit právě svou přehledností a přístupností, kdy lze identifikovat právě nedostatky v kvalitě dat. A jelikož některé datové katalogy zahrnují funkce data mining, například pro data lineage či jiný pohled na identifikaci relevantních vztahů, lze je považovat také za nástroj, který svým způsobem podporuje i aktivity spojené s řešením prediktivní analýzy. (Pour, 2017).

2. Možnosti datových katalogů

Možnosti datových katalogů jsou různé dle daného softwaru. Nicméně obecně lze uvést několik příkladů, jaké možnosti a funkce mohou datové katalogy nabízet. Zdrojem pro tuto část práce byly především samotné datové katalogy a popis služeb, které nabízejí.

Jednou z možností datových katalogů je **katalogizace dat**, tzn. že data katalogizují metadata o datech a usnadňují tak identifikaci a porozumění datům. Fungují tedy dost podobně, jako systém knihoven. Tato možnost také úzce souvisí s další možností, **centralizace dat**, kdy datový katalog poskytuje centralizovaný přístup ke všem datovým zdrojům, což vede k usnadnění jejich správy a využití. S centralizovanými daty s metadaty je také možnost snadněji řídit **bezpečnost a řízení přístupu k datům**, což může vést k vyšší transparentnosti využívání dat.

Další z možností je **metadata management**, tedy uchování a správa metadat o datech, včetně popisů, schémat, klíčových slov a dalších, což podporuje další z možností, kterou je snadnější hledání a vyhledávání dat na základě různých kritérií, která právě prohledávají uvedená metadata. Tuto možnost nabízejí pomocí interaktivních uživatelských rozhraní. Další související možností ulehčující vyhledávání a samotnou organizaci dat je možnost přidávat tagy a anotace k datům.

Datové katalogy nabízejí často **možnost integrovat s existujícími nástroji a technologiemi**, jako například databázové systémy, nástroje Business Intelligence a další.

Obecně k řízení metadat většina datových katalogů nabízí podporu pro různé **metadatové standardy** umožňující konzistentní a interoperabilní správu metadat, ať už se jedná o firemní standardy nebo ISO standardy.

Některé datové katalogy nabízejí **oznamování a sledování změn v datech a metadatech**, což usnadňuje jejich správu a udržování dat aktuálních a relevantních. Tato oznamování a sledování lze u některých katalogů také získávat jako upozornění, lze je omezit pouze na vybrané změny či události v datech. Tato funkce může být podporována možností **auditování a sledování změn**, pro zajištění integrity dat. Mimo těchto oznámení nabízejí některé datové katalogy také **tvorbu reportů a statistik o používání metadat**, což lze využít nejen ke zlepšení samotné správy metadat, ale také k analýze a plánování.

Kromě výše zmíněných funkcí lze vybrané datové katalogy využít také jako **formu dokumentace** a nástroj pro vytváření a udržování dokumentace datových zdrojů.

Některé z datových katalogů nejsou omezeny pouze na malé datové sety, ale umožňují správu a zpracování větších objemů dat, tzv. Big Data. (Ham a Crockett, 2021)

3. Omezení datových katalogů

Jako mnoho jiných nástrojů, tak i datové katalogy mají svá omezení, které mohou narušovat efektivitu a zpomalovat či přímo bránit přínosům jejich implementace. Mezi taková omezení lze zmínit:

1. **Neaktuálnost metadat** – metadata o datech v katalogu je nutné udržovat aktuální, metadata mohou zastarat a nepřesně zobrazovat aktuální stav dat, především v dynamicky se měnících datových setech. Proto je nutné metadata pravidelně kontrolovat a aktualizovat.
2. **Složitost a náklady na implementaci** – vybrané typy datových katalogů může být náročné implementovat a následně spravovat, jiné zase mohou být příliš nákladné. Zejména pokud společnost nedisponuje dostatečnou kapacitou, financemi a technologickými zkušenostmi. Proto je důležité dbát na kvalitě a přesnosti prvotní analýzy při výběru řešení.
3. **Uživatelské rozhraní** – některé datové katalogy nemají vhodné uživatelské rozhraní a může být obtížné pro pochopení a užívání koncovými uživateli. V takovém případě je risk, že uživatelé budou hledat jiné cesty k přístupu k datům či více vytěžovat interní podporu datového katalogu.
4. **Kvalita metadat** – aby byla zajištěna efektivita datového katalogu, musí mít metadata o datech dostatečnou kvalitu. Pokud nejsou správná, úplná, aktuální, mohou vést ke ztrátě důvěryhodnosti dat.
5. **Bezpečnost a řízení přístupu** – pokud datové sety obsahují citlivá data, je klíčové data dostatečně zabezpečit, aby nebyla data ohrožena.
6. **Škálovatelnost** – pro společnosti s větším objemem dat a větším počtem uživatelů je důležité, aby byl datový katalog škálovatelný a umožnil pokrytí zdrojů hardwarových i softwarových.
7. **Nedostatek standardizace** – některé datové katalogy mohou aplikovat různé standardy a formáty metadat, která mohou mít problém s integrací s jinými systémy.
8. **Kontext** – datové katalogy mohou mít omezenou schopnost zachovat kontext kolem dat, což může vést k potížím uživatelů pochopit souvislosti s jinými daty.
9. **Náročnost udržování a správy** – nedostatek prostředků k udržování a správě datového katalogu může vést ke snížení jeho efektivitu.

Zdrojem této kapitoly byla inspirace v předchozích kapitolách vlastním zpracováním a internetové zdroje. (Klipfolio, 2023)

4. Přehled vybraných dostupných nástrojů datových katalogů

Výběr datových katalogů pro tento přehled byl kromě dostupných datových katalogů v rámci společnosti inspirován také trendem uvedeným společností Gartner. Byly vybrány 4 nástroje od různých poskytovatelů, a to **DataHub, Microsoft Purview, Informatica a Collibra**.

Pro nabízené funkce byly zpracovány tabulky shodné pro vybrané datové katalogy. Porovnání bylo vypracováno na základě informací dostupných na stránkách poskytovatele.

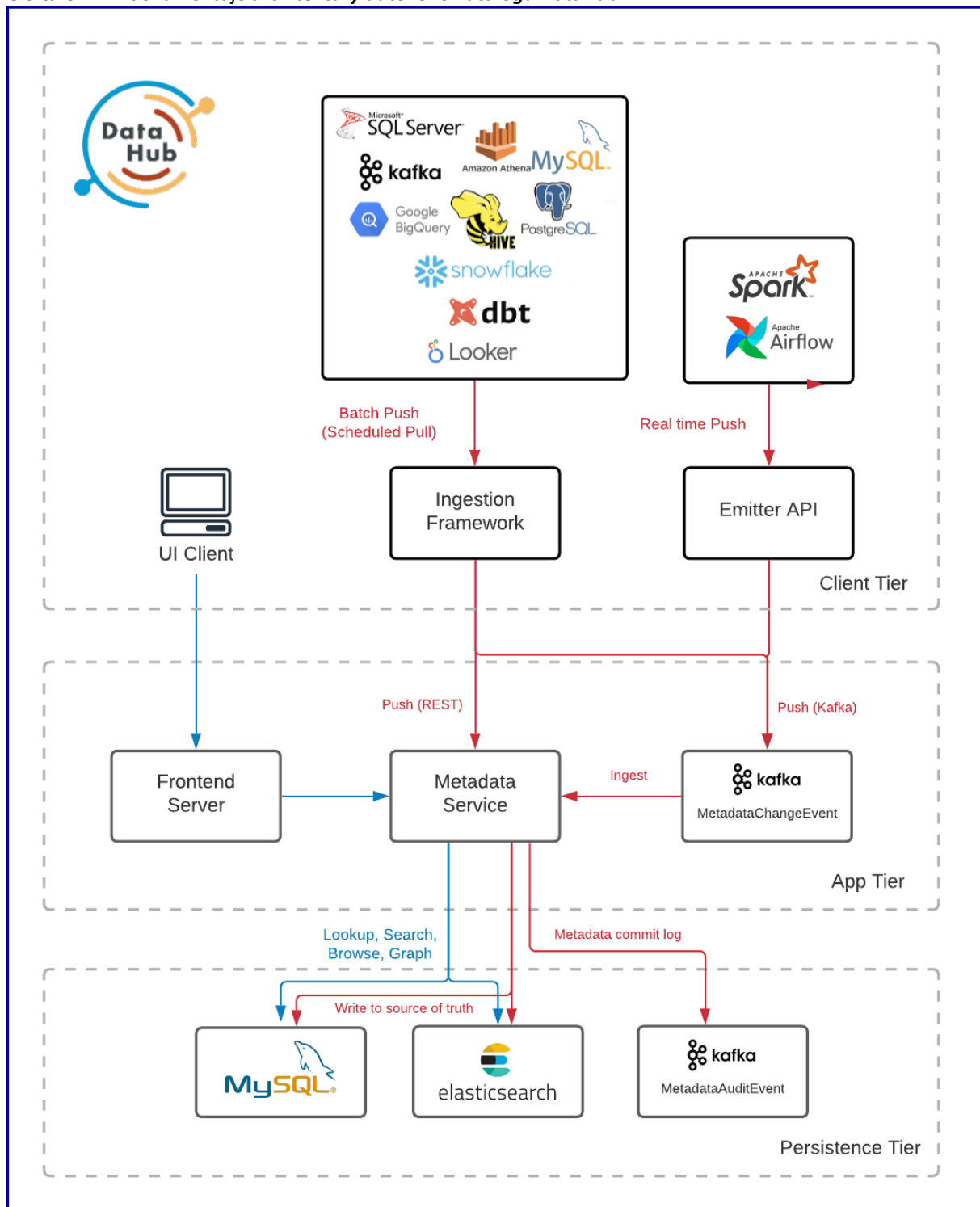
4.1 DataHub

DataHub je **open source datový katalog** vyvinutý ve spolupráci společností Acryl Data a LinkedIn. Poskytuje základní funkce datového katalogu a jako open source řešení navíc umožňuje vlastní rozšíření a správu. Nabízí například **předpřipravené integrace** se systémy Kafka, Airflow, MySQL, MS SQL Server, Postgres, LDAP, Snowflake, Hive, BigQuery a další. (DataHub, 2023)

Tabulka 1: DataHub provedení vybraných kritérií (vlastní zpracování)

Funkce	Provedení
Metadata Management	Správa a uchování metadat o datech, popisy, tagy, a jiné, dataset stats. Metadata lze spravovat také skrze UI.
UI/Vyhledávání	Vyhledávání umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožněno skrz administraci - vytáření uživatelů, skupin a pravidel přístupů.
Podpora pro metadatové standardy	Součástí repozitáře je metadata služba, podporuje však i federované služby komunikující s centrálním vyhledávacím indexem.
Reporty a statistiky	Nabízena možnost tvorby automatických reportů na základě vybraných kritérií.
Oznamování a zasílání upozornění	Ano, lze nakonfigurovat jako rozšíření, například do MS Teams, Slack.
Centralizovaný přístup	End-to-End Lineage
Auditování a sledování změn	Ano, v rámci Govern in Real Time
Integrace s dalšími nástroji	DataHub lze integrovat s nástroji třetích stran, například synchronizace Tagů do Snowflake systému, nebo vytváření JIRA ticketů a jiné.

Obrázek 4-1 dokumentuje **architektury datového katalogu** DataHub.



Obrázek 4-1 DataHub architektura (DataHub, 2023)

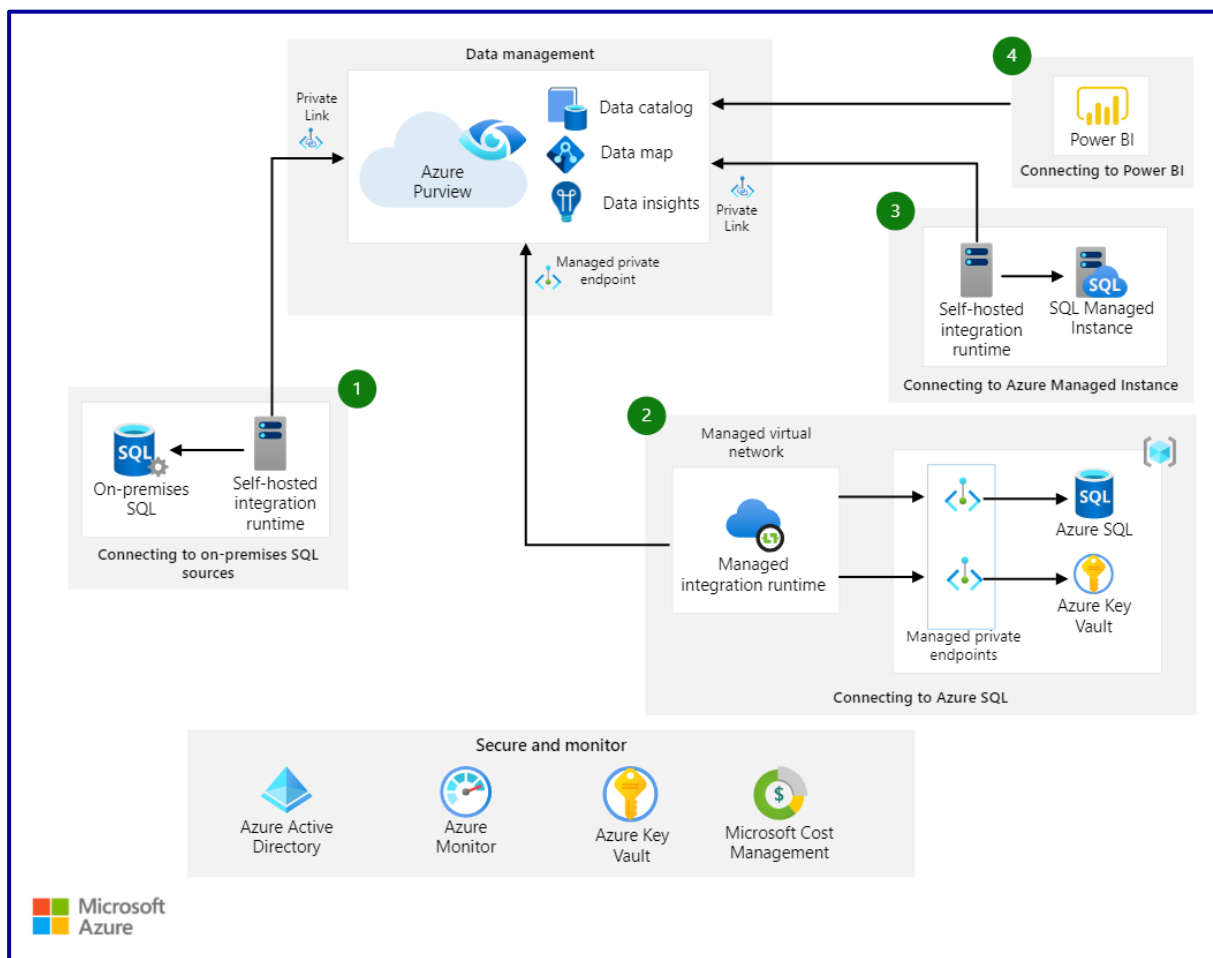
4.2 MS Purview

MS Purview je datový katalog společnosti Microsoft. Slouží **k centralizovanému řízení dat a metadat** v organizaci a lze ho plně integrovat s ekosystémem Microsoft, jako například MS Azure, Azure Data Factory, Power BI a jiné. (Microsoft, 2023).

Tabulka 2: MS Purview provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	MS Purview umožňuje správu a uchovávání metadat o datech, včetně popisů, schémat a dalších klíčových informací.
UI / Vyhledávání	Vyhledávání je umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožněno pomocí vytváření uživatelů, správu rolí. Další funkce pro zabezpečení dat a metadat, včetně řízení přístupu, jsou například služba Information Protection, Insider Risk Management, Data Loss Prevention, Customer Lockbox, Information Protection SDK, Message Encryption.
Podpora pro metadatové standardy	Purview podporuje různé metadatové standardy, což usnadňuje konzistentní správu metadat, například pomocí MS Data policies, Data Estate Insights.
Reporty a statistiky	Uživatelé mohou generovat reporty a statistiky o používání dat a metadat.
Oznamování a zasílání upozornění	Microsoft Purview umožňuje nastavit upozornění na změny v datech a metadatech nebo například na policies která nemají vyřešené problémy.
Centralizovaný přístup	Data Map
Auditování a sledování změn	MS Purview nabízí například moduly MS Purview Audit pro auditování, a MS Purview eDiscovery a Data Lifecycle Management pro sledování změn, nebo například životní cyklus dat.
Integrace s dalšími nástroji	Platforma Purview má silnou integraci s Azure Data Services, což zahrnuje Azure Synapse Analytics a Azure Data Lake Storage.

Obrázek 4-2 dokumentuje architekturu MS Purview:



Obrázek 4-2 MS Purview architektura (Microsoft, 2023)

4.3 Informatica

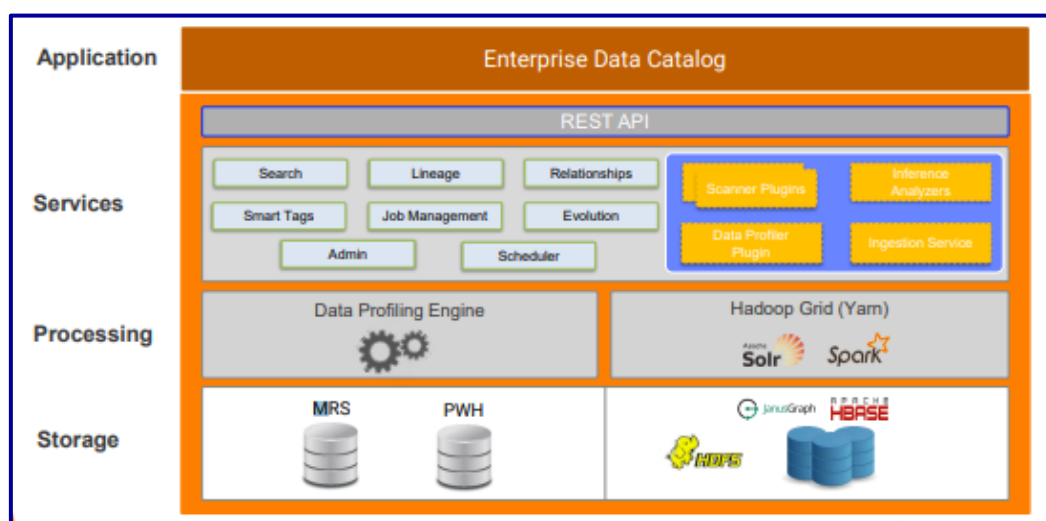
Americká akciová společnost Informatica, založena v roce 1993 zakladateli Gaurav Dhillon a Diaz Nesamoney v Kalifornii, USA. Dle informací dostupných na oficiálních stránkách byla společnost v roce 2023 lídrem na trhu v oblastech Data Quality Solutions, Data Integration Solutions a Master Data Management Solutions hodnocené společností Gartner, Magic Quadrant. Nabízí **celkové cloudové řešení** Intelligent Data Management Cloud na bázi engine CLAIRE umělé inteligence ke správě dat jakéhokoliv typu, schématu, složitosti nebo pracovní zátěže v jakékoliv lokaci na jediné platformě. (Informatica, 2023)

Tabulka 3 Informatica provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	Definice a správa metadat
UI/Vyhledávání	Umožněno skrze uživatelské rozhraní.

Bezpečnost a řízení přístupu	Obsahuje funkce pro zabezpečení dat a metadat a umožňuje řízení přístupu k nim.
Podpora pro metadatové standardy	Informatica podporuje různé metadatové standardy, a nabízí možnost integrace standardizace dat, ověřování a jiné, podporováno službou Informatica Cloud Data Governance and Catalog.
Reporty a statistiky	Uživatelé mohou generovat reporty a statistiky o používání dat a metadat.
Oznamování a zasílání upozornění	Informatica umožňuje nastavit upozornění na určité změny nebo události a zasílat oznámení srkze UI, email či jiné.
Centralizovaný přístup	End-to-End data lineage, holistické pohledy na vztahy mezi daty.
Auditování a sledování změn	Platforma umožňuje provádět audit změn v datech a metadatech, nebo například prediktivní analýzu.
Integrace s dalšími nástroji	Integrace s databázovými systémy, BI nástroji a další.

Obrázek 4-3 prezentuje architekturu produktů společnosti Informatica.



Obrázek 4-3 Informatica architektura (Informatica, 2023)

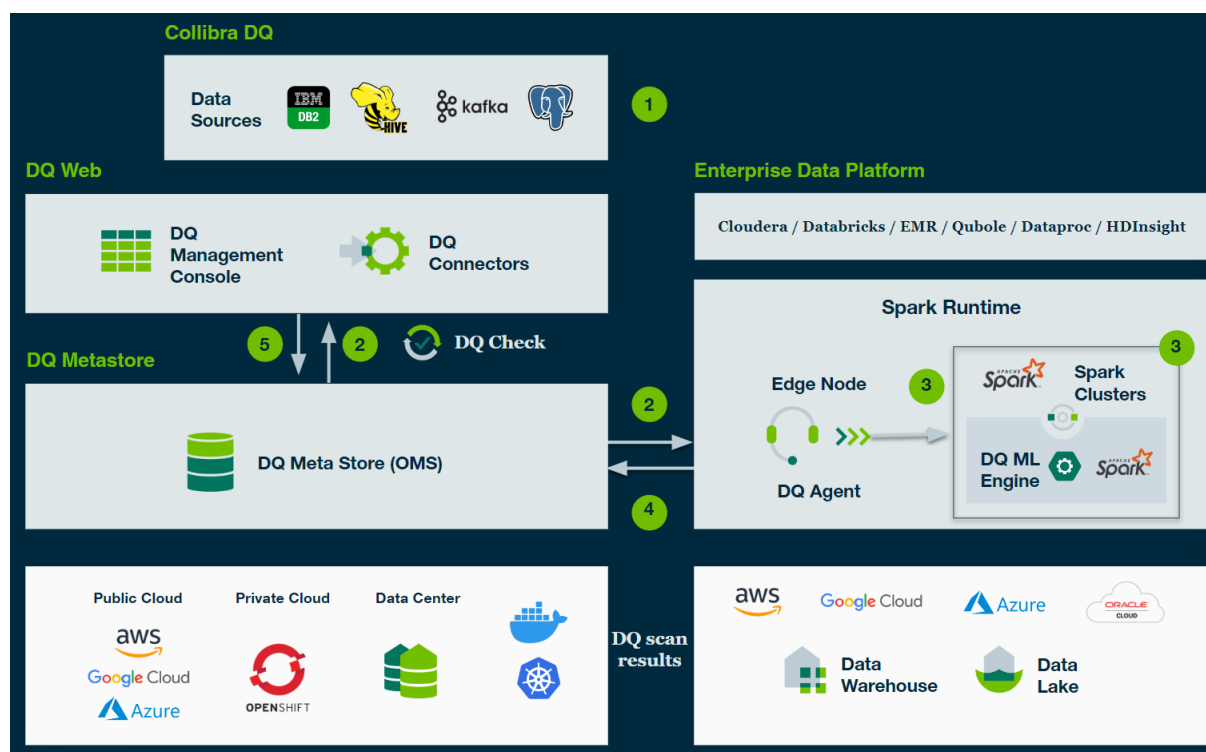
4.4 Collibra Data Intelligence Cloud

Společnost Collibra nabízí podnikové řešení Collibra Data Governance Center, jejíž součástí je právě Collibra datový katalog. Používá **sémantické vyhledávání, nabízí end-to-end vizualizaci datových toků, automatizaci klasifikace a kategorizace datasetů**, business procesů a intuitivní workflow engine. (Collibra, 2023)

Tabulka 4 Collibra provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	Business, technické i bezpečnostní metadata.
UI/Vyhledávání	Umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožňuje řízení přístupu, asset-level security, a současně dodržuje měnící se regulační požadavky v celosvětovém měřítku.
Podpora pro metadatové standardy	V rámci automatizace data governance.
Reporty a statistiky	Data quality reports, usage reports
Oznamování a zasílání upozornění	Skrze UI
Centralizovaný přístup	Nativní business a technické data lineage, one-stop-shop.
Auditování a sledování změn	Umožňuje sledovat změny v datasetech, zabezpečení i uživatelské změny.
Integrace s dalšími nástroji	Platforma Collibra umožňuje integraci s existujícími nástroji a systémy, včetně databázových systémů, business intelligence nástrojů a dalších.

Obrázek 4-4 dokumentuje architekturu Collibra Data Intelligence Cloud:



Obrázek 4-4 Collibra architektura (Collibra, 2023)

5. Závěr

Všechny datové katalogy poskytují **velmi podobné možnosti**, které byly zkoumány. Disponují vlastními rozhraními a často je lze integrovat s dalšími produkty poskytovatele.

Microsoft si zakládá na integraci s ostatními systémy a aplikacemi společností vyvinutými, je tedy vynikající volbou pro společnosti, která využívá i jiné produkty v rámci Azure. Je nabízen v rámci cloudových služeb Microsoft Azure.

Informatica datový katalog se specializuje na řešení integračních a datových správcovských potřeb. Nabízí silné integrační a ETL možnosti, je možnost nástroj nasadit jak v cloudu, tak on-premises. Je tak vhodný pro organizace hledající komplexní řešení pro integrační, datovou správu a data governance.

Collibra datový katalog funguje jako samostatný produkt, není přímo spojen s cloudovým ekosystémem a lze ho nasadit i v rámci on-premise. Cílí na širší škálu uživatelů s různými datovými infrastrukturami a specializuje se na data governance.

DataHub jako jediný z výše uvedených nástrojů poskytuje open source řešení. Klade tedy vyšší nároky na technické možnosti a znalosti svých zákazníků, nabízí však návody na svých stránkách a velkou komunitu uživatelů a vývojářů. Nabízí samostatné řešení s předpřipravenými rozhraními na nejvíce používané případy, například skennery na typy databází, které jsou nejčastěji používány.

6. Zdroje

APACHE SOFTWARE FOUNDATION, 2023. *Apache Hive* [online] [vid. 2023-12-02]. Dostupné z: <https://hive.apache.org/>

AREVALO, Isabel, 2023. *Správa dat napříč aktivy Azure SQL pomocí Microsoft Purview - Azure Architecture Center* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/cs-cz/azure/architecture/solution-ideas/articles/azure-purview-sql-estate-architecture>

BAARS, H., KEMPER, H., 2008. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework [online] [vid. 2018-09-23]. Dostupné z: <https://www.semanticscholar.org/paper/Management-Support-with-Structured-and-Unstructured-Baars-Kemper/2809521eecd9fb9a11332a96bab273bbae62e14>

COLLIBRA, 2023a. *Collibra Data Catalog | Collibra* [online] [vid. 2023-12-01]. Dostupné z: <https://www.collibra.com/us/en/products/data-catalog>

COLLIBRA, 2023b. *Data Citizens drive change with access to the right data* [online] [vid. 2023-12-01]. Dostupné z: <https://www.collibra.com/us/en/company/who-we-are>

DATABRICKS INC., 2021. What is Hadoop Distributed File System (HDFS). *Databricks* [online] [vid. 2023-12-02]. Dostupné z: <https://www.databricks.com/glossary/hadoop-distributed-file-system-hdfs>

DATAHUB, 2023a. *A Metadata Platform for the Modern Data Stack | DataHub* [online] [vid. 2023-12-01]. Dostupné z: <https://datahubproject.io/>

DATAHUB, 2023b. *Deployment Guides | DataHub* [online] [vid. 2023-12-01]. Dostupné z: <https://datahubproject.io/docs/category/deployment/>

DRESCH, Aline, Daniel PACHECO LACERDA a José Antônio Valle ANTUNES, 2015. *Design science research: a method for science and technology advancement*. Switzerland: Springer. ISBN 978-3-319-07374-3.

HAM, Kenneth a Dustin CROCKETT, 2021a. *Evaluation of Data Catalog Software for Hanford Site Environmental Datasets* [online]. PNNL--31960, DVZ-RPT--066, 1832173. [vid. 2023-12-01]. Dostupné z: doi:[10.2172/1832173](https://doi.org/10.2172/1832173)

HAM, Kenneth D. a Dustin L. CROCKETT, 2021b. *Evaluation of Data Catalog Software for Hanford Site Environmental Datasets* [online]. PNNL-31960; DVZ-RPT-066. B.m.: Pacific Northwest National Lab. (PNNL), Richland, WA (United States) [vid. 2023-12-01]. Dostupné z: doi:[10.2172/1832173](https://doi.org/10.2172/1832173)

INFORMATICA, 2023. *Informatica*. *Informatica* [online] [vid. 2023-12-01]. Dostupné z: <https://www.informatica.com>

JSON, 2023. *JSON* [online] [vid. 2023-12-01]. Dostupné z: <https://www.json.org/json-en.html>

KLIPFOLIO, 2023. *What is a Data Catalog? Definition, Challenges, and More | Klipfolio* [online] [vid. 2023-12-01]. Dostupné z: <https://www.klipfolio.com/resources/data-stack/data-catalog>

KUBERNETES, 2023. *Production-Grade Container Orchestration. Kubernetes* [online] [vid. 2023-12-01]. Dostupné z: <https://kubernetes.io/>

LUKŠÍK, Marek, 2020. *Datová kvalita a datové katalogy* [online]. Dostupné z: <https://kata-log.vse.cz/Record/000658742>

MBI, 2023. *MBI - Management Byznys Informatiky - home page* [online] [vid. 2023-12-01]. Dostupné z: <https://mbi.vse.cz/>

MBI-AF, 2023. *MBI-AF | IT A ANATOMIE FIRMY* [online]. [vid. 2023-12-01]. Dostupné z: <https://www.mbi-af.cz/>

MICROSOFT, 2023a. *Co je SaaS? Software jako služba | Microsoft Azure* [online] [vid. 2023-12-01]. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-saas>

MICROSOFT, 2023b. *Learn about Microsoft Purview* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/en-us/purview/purview>

MICROSOFT, 2023c. *Prostředí Integration Runtime - Azure Data Factory & Azure Synapse* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/cs-cz/azure/data-factory/concepts-integration-runtime>

NEDZELSKÝ, Toman, 2019. *Metoda dynamické alokace lidských zdrojů v projektovém řízení - Vysokoškolské kvalifikační práce - Vysoká škola ekonomická v Praze* [online] [vid. 2023-12-01]. Dostupné z: <https://vskp.vse.cz/77486>

OPENSOURCE.COM, 2023. *What is open source?* | Opensource.com [online] [vid. 2023-12-01]. Dostupné z: <https://opensource.com/resources/what-open-source>

POUR, Jan, 2017. *Prediktivní analytika v podnikové informatice* | Česká společnost pro systémovou integraci [online] [vid. 2023-12-01]. Dostupné z: <http://old.cssi.cz/prediktivni-analytika-v%C2%A0podnikove-informatice>

REDHAT, 2020. *Red Hat OpenShift enterprise Kubernetes container platform* [online] [vid. 2023-12-01]. Dostupné z: <https://www.redhat.com/en/technologies/cloud-computing/openshift>

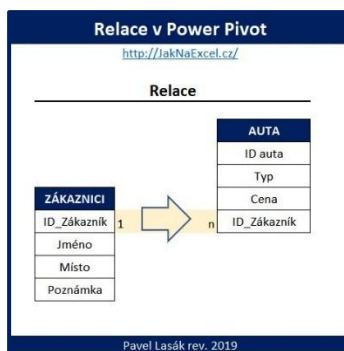
SHEVEK, 2021. *What Data Catalogs Don't Reveal*. DevOps.com [online] [vid. 2023-12-01]. Dostupné z: <https://devops.com/what-data-catalogs-dont-reveal/>

TRANSITION SUPPORT, 2023. *Quality - what it means* [online] [vid. 2023-12-01]. Dostupné z: https://transition-support.com/what_is_quality.html

VACEK, Martin, 2014. *Metriky, monitoring a řídicí proces Data Governance - PDF Stažení zdarma* [online] [vid. 2023-12-01]. Dostupné z: <https://docplayer.cz/10899103-Metriky-monitoring-a-ridici-proces-data-governance.html>

VONDRA, Marek, 2023. *Lekce 1 - Docker - Teorie a instalace* [online] [vid. 2023-12-01]. Dostupné z: <https://www.itnetwork.cz/docker-teorie-a-instalace>

WELLS, Dave, 2020. *What Is a Data Catalog? - Importance, Benefits & Features* | Alation [online] [vid. 2023-12-01]. Dostupné z: <https://www.alation.com/blog/what-is-a-data-catalog/>



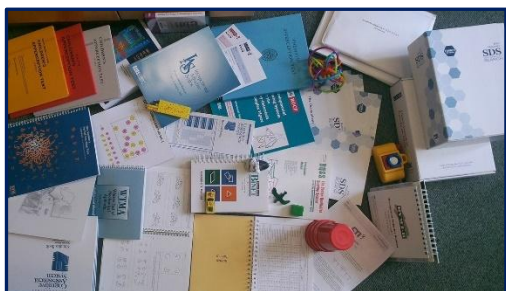
D. Metodika datové migrace

(pracovní dokument)



Marek Slezák

VŠE Praha, 2023



[1] Metodika a postup datové migrace	
[2] Nástroje pro datovou migraci	[3] Práce s daty
[4] Testování	[5] Rozpočet
[6] Přehled zdrojů	



Účelem tohoto dokumentu je poskytnout **přehled všech podstatných doporučení a návodů** ve vztahu k řízení a řešení **datové migrace** v rámci zejména aplikačních projektů. Dokument představuje speciální dílčí **dokument k základnímu dokumentu Řízení IT** na portále MBI-AF. Dokument vychází **z celé řady zdrojů obsažených v závěrečném přehledu** a na bázi příslušných citací vybírá z nich ta nejpodstatnější doporučení.

1. Metodika a postup datové migrace

1.1 Používejte formální metodiku, která byla vyzkoušená a otestovaná [1]

Tento osvědčený postup se podílí na úspěšném dokončení projektů. Rozlišovat také musíme skutečnost, zdali se jedná o metodiku interní, která byla vytvořena vlastními silami, nebo metodiku, kterou nám poskytl systémový integrátor, či dodavatel. Metodiky poskytované systémovými integrátory a dodavateli mají totiž větší úspěšnost než metodiky interní, což může být dáno skutečností, že jsou interní metodiky méně otestované [1].

Využívejte proto metodiky, které již byly ověřeny. Metodiku můžete také stanovit na základě předchozích zkušeností z migračních projektů, pokud jimi disponujete. **Výhodou migračních projektů** je skutečnost, že i přes to, že se liší v různých oblastech, **jádro zůstává pořád stejné**, a tudíž i postup se diametrálně neliší, a proto můžete stejný, či obdobný postup využít také příště. Počítejte ale také s tím, že v určitých oblastech budete kvůli různým požadavkům zainteresovaných stran a dalším skutečnostem nuceni metodiku modifikovat. **Na závěr každého projektu si definujte tzv. lessons learned** – co fungovalo, a co naopak nefungovalo a je potřeba do budoucna zlepšit. Také to vám může pomoci vylepšovat metodiku pro hladší průběh budoucích projektů.

1.2 Metodika migrace dat by měla postupovat od posouzení, přes plánování až po přesun a validaci [12]

Jak již bylo zmíněno, snažte se využívat metodiku, která byla otestována. Váš **postup by měl zahrnovat plánování**. Během této fáze společně s dalšími zainteresovanými stranami **naplánujte, jak dojde k přesunu dat ze zdrojového do cílového prostředí a vytvořte projektový plán**, který budete v průběhu projektu dodržovat. Ve fázi plánování také stanovte, jaký hardware nebo software budete k migraci potřebovat, jaká bude architektura migrace, specifikujte zdroje, postupy migrace a plány nasazení a testování. Do plánování zahrňte nejen pracovníky vašeho týmu, ale také další strany, které se na projektu podílí. Také ostatní strany budou mít v rámci projektu zodpovědnost za vykonávání některých činností, proto je důležité, aby měly o projektovém plánu přehled.

Ve druhé fázi se zaměřte na **přesun dat ze zdrojového do cílového prostředí**. Získejte, nainstalujte a nakonfigurujte potřebný software, který budete k migraci potřebovat, získejte data a analyzujte zdrojové a cílové struktury, a začněte tvořit transformační kód, který data transformuje do cílových struktur. Počítejte také s tím, že vám fáze transformací dat pravděpodobně zabere nejvíce času společně s následným testováním dat. Nepodceňte testování, které může výrazně ovlivnit průběh projektu. V první řadě proveďte systémové testování, dále poté uživatelské testování.

Validace by pak měla být poslední fází, během které se ověří, **zda byla data úspěšně přesunuta ze zdrojového do cílového prostředí** podle plánu. Pokud se na projektu podílíte jako externí systémový integrátor, nezapomeňte připravit pro vedení společnosti, pro kterou jste data migrovali, prezentaci se shrnutím projektu. V závěru projektu také vypracujte dokumentaci a předejte ji klientovi. Získané zkušenosti zpracujte během tzv. lessons learned, kdy si stanovíte, co během projektu fungovalo, co nefungovalo a můžete zlepšit do dalšího projektu. Na základě těchto zjištění můžete modifikovat metodiku pro další projekty.

1.3 Mějte plán [12] [15]

Plánování migrace může pomoci **určit, kde by se mohly vyskytnout problémy a způsoby**, jak se jim vyhnout. Mějte plán a plňte ho. Plánujte migraci dat (nebo integraci) jako projekt a sledujte svůj postup podle plánu. **Plánujte neúspěch**; plánujte špatné požadavky; plánujte klienty, kteří nereagují nebo jsou nerozhodní. Plánujte zkrátka vše. [15]

Zároveň se snažte mít **projektový plán co nejvíce konkretizovaný** tak, aby obsahoval kompletní seznam všech činností, které je potřeba vykonat, spolu s časem potřebným pro jejich provedení a identifikací odpovědných osob. Právě projektový plán vám umožní všechny tyto skutečnosti vidět.

Počítat je také nutné se situacemi, ve kterých může dojít ke změně požadavků a projektový plán tak bude muset být přizpůsoben novým podmínkám. Počítejte proto s tím, že změny mohou nastat také v průběhu projektu, a to nejen v důsledku předem špatně stanovených požadavků, ale také například v souvislosti s opomenutím některých činností nebo v důsledku změn požadavků a rozhodnutí zainteresovaných stran.

1.4 **Mějte dostatek času na plánování projektu [15][16]**

Je důležité, abyste si **na plánování projektu vyhradil dostatek času**. Plánování totiž může ovlivnit průběh celého projektu, a proto by neměly být opomenuty žádné kritické kroky. Fáze plánování by měla zahrnovat jednotlivé kroky, které je během průběhu migračního procesu provést, a zohlednění zdrojů odpovědných za jednotlivé úkoly [16].

Pokud je v projektu datové migrace **zainteresováno více stran, nezapomeňte je do procesu tvorby projektového plánu zapojit**. Každá strana bývá zodpovědná za provedení různých kroků v rámci projektu. Proto je klíčové, aby se všechny strany aktivně zapojily do tvorby projektového plánu a byly informovány o svých odpovědnostech. Nezapomeňte také všem stranám zajistit přístup k projektovému plánu, díky čemuž budou moci pravidelně monitorovat jeho postup.

1.5 **Zkontrolujte, zda jsou k dispozici potřebné zdroje [16] [19]**

Zajištění dostupnosti zdrojů, které jsou zodpovědné za jednotlivé úlohy, **hraje klíčovou roli** během projektu. Klíčové zdroje by měly být zapojeny od začátku projektu a neměly by se jeho součástí stávat až v jeho průběhu. Pokud je do migrace zapojeno více pracovišť, je potřeba, aby se zapojily zdroje z každého pracoviště v případě, že mezi pracovišti existují rozdíly [16].

V případě, že vás v průběhu migračního projektu opustí některý člen týmu, je nezbytné nalézt kvalifikovanou náhradu. Současně zajistěte, aby byl **nový člen týmu plně seznámen se všemi aspekty** projektu a jeho rolí v něm. Pro tyto účely mu poskytněte odpovídající školení nebo konzultace. Je důležité počítat s možností náhrady člena týmu a také s tím, že následné zaškolení nového člena bude vyžadovat čas a úsilí. Nepodceňte tuto činnost, jelikož správné zaškolení nového člena týmu je klíčové pro zachování kontinuity a efektivitu projektu.

Zároveň zajistěte, aby byly **od počátku projektu k dispozici zdroje, od jejichž činností se odvíjejí další kroky**. Například, pokud je některý člen týmu odpovědný za implementaci prostředí, musí tuto činnost dokončit v předem stanoveném termínu, aby nezdržoval postup dalších kroků projektu, které jsou na implementaci prostředí závislé.

1.6 **Zdokumentujte proces migrace [12][15]**

Zdokumentujte celý proces migrace. To vám pomůže vyškolit zaměstnance, předat know-how a zjednodušit nebo zefektivnit další projekt datové migrace, což pomůže snížit náklady i rizika. Tvorbou dokumentace rovněž umožníte přístup k informacím o provedené migraci a poskytnete auditorům komplexní přehled o průběhu projektu.

Pokud se na projektu podílíte jako externí systémový integrátor, **předejte vytvořenou dokumentaci se v závěru projektu klientovi**. Tato dokumentace klientovi poskytne odborné znalosti, které může využít a aplikovat v případě budoucích projektů datové migrace. Snažte se také dokumentaci předat ve stručné a srozumitelné formě, aby byla čitelná a použitelná. Díky tomu může dokumentace sloužit také pro další zlepšování procesů datové migrace.

1.7 **Vyviňte interní kompetence ohledně migrace dat a nespolehejte zcela na externí zdroje [1]**

Pokud jako společnost nedisponujete dostatečným know-how nebo zdroji pro provedení migrace, můžete **oslovit externího systémového integrátora**, který má s podobnými nebo stejnými projekty zkušenosti, a který migraci dat provede za vás. I přesto je však důležité, abyste byly součástí migračního

projektu od začátku do konce, jelikož i vy budete mít v rámci projektu zodpovědnost za vykonávání některých činností. Díky tomu budete mít zároveň přehled o průběhu projektu. Vaše zapojení do projektu je klíčové a je nutné, abyste aktivně komunikovali a spolupracovali s ostatními zainteresovanými stranami v projektu.

1.8 Podnik musí být zapojen ve všech fázích projektu, od zahájení až po testování [1] [6]

Jedná se podle [1] o **nejkritičtější faktor úspěchu**. Zákazníkům, produktům nebo službám totiž rozumí mnohem více byznys než IT. Pokud se pracovníci těmito vztahy a jejich zachováním nezabývají, může projekt selhat. Podobně hovoří také [6], podle které je třeba **určit, kdo rozumí obchodním datům**, se kterými se pracuje. Je proto důležité zapojit také jiné podnikové útvary a lidi, kteří data používají [6].

Je třeba si uvědomit, že pokud v projektu vystupuje **více zainteresovaných stran, není podnik samotný jedinou stranou, která by do projektu migrace měla být zapojena**. Z hlediska jeho úspěšného dokončení by do něj měly být aktivně zapojeny všechny strany, které se na projektu podílí, jelikož každá z těchto stran je v projektu zodpovědná za vykonávání některých činností. Zajistěte také, aby byly všechny strany zapojeny od začátku projektu, díky čemuž budou mít dostatečný přehled a rozumět kontextu.

1.9 Získejte písemný závazek vendora [16]

Pro zajištění úspěchu v projektu je důležité, aby byl **zapojen také dodavatel systému**, pokud jeho služeb využíváte. Dodavatel systému, do kterého data migrujeme, by měl být přítomen po celou dobu projektu [16]. Zároveň zajistěte, aby byl dodavatel zapojen od samého počátku projektu a mohl se tak věnovat svým činnostem. Zajistěte také **pravidelný kontakt s dodavatelem**, nechte se informovat o postupu na jeho straně a informujte ho o postupech na vaší straně.

Pokud se na projektu podílíte **jako externí systémový integrátor**, podpis o spolupráci s dodavatelem systému pravděpodobně bude mít váš klient, který jeho služeb využívá. Proto v případě, že vendor neplní závazky, které ze smlouvy vyplývají, reportujte tuto skutečnost klientovi, který s vendorem smlouvu uzavřel, a který následně stanoví rozsah škod a jejich náhradu. V případě, že klienta o problémech nebudete informovat, sami se do nich můžete dostat, což zkomplikuje průběh migrace.

1.10 Odstraňte komunikační bariéry [16]

Mnoho problémů, které během projektu vznikají, jsou způsobeny komunikačními bariérami – **špatné pochopení problému, nejasné vysvětlení, jazykové rozdíly, nejasnosti v odpovědnostech, nedostatečné informace** atd. [16]. Odstraňte proto komunikační bariéry mezi všemi členy týmu a zároveň mezi všemi stranami projektu. To vám pomůže usnadnit průběh migračního projektu.

V případě, že nerozumíte významu dat nebo způsobu, jak data mapovat, **komunikujte s relevantními stranami a diskutujte řešení problému**, aniž byste se pokoušeli o vlastní řešení. Následné chyby mohou mít negativní vliv na průběh celého projektu.

Pokud pracujete **na mezinárodním projektu**, popřípadě na projektu, kde se používá jiný jazyk než váš mateřský jazyk, zajistěte, že **členové týmu ovládají potřebný jazyk** a jsou schopni aktivně komunikovat s ostatními stranami a řešit případné problémy.

Na závěr každého meetingu shrňte, co bylo na dané schůzce probráno, jaké jsou následující kroky, a kdo je za ně zodpovědný. Dělejte si také během každé schůzky poznámky, které pak mohou sloužit jako vodítko nebo důkaz, pokud některá ze zainteresovaných stran bude tvrdit něco jiného, než tvrdila minule.

1.11 Automatizujte, pokud je to možné [15]

Vše, co lze automatizovat, by mělo být automatizováno. **Automatizace zahrnuje všechny přesuny a transformace dat, včetně přesunu původních zdrojových dat do stagingového prostředí.** Automatizujte úlohy, které provádíte opakovaně se stejným postupem, včetně přípravných prací po dodání nových dat, jako je kontrola datové kvality nebo aplikace transformací, abyste minimalizovali riziko chyb a zvýšili efektivitu procesu.

Pro dosažení automatizace můžete využít **vytváření skriptů nebo procedur, které umožní automatické nahrávání, či formátovat** potřebných dat podle stanovených pravidel. Tím ušetříte čas a minimalizujete riziko chyb, které by mohly vzniknout při ručním zpracování dat.

2. Nástroje pro datovou migraci

2.1 Používejte nástroje pro profilování dat před zahájením projektu i v jeho průběhu [1]

Když společnost **Bloor Research provedla v roce 2011 průzkum** mezi společnostmi, které se projekty datové migrace zabývají, zjistila, že zhruba **polovina společností, které profilování využily, tak učinily před stanovením rozpočtu projektu.** Z této poloviny poté 72% společností dokončilo projekt včas a v rámci rozpočtu. Naopak společnosti, které profilování využili později, měly úspěšnost dokončení pouze 52% z nich. Jednou z hlavních příčin byl uveden nedostatečný přehled o problémech s kvalitou dat, který profilování dat poskytuje [1].

Profilovací nástroje vám pomohou porozumět datům a ušetřit velké množství času. Doporučuji používat profilovací nástroje také v případě, že migrujete menší množství dat. Díky nim lépe porozumíte kvalitě dat, se kterými pracujete, čímž se budete schopni vyvarovat chybám v datech, jako jsou duplicity nebo chybějící hodnoty, a nutnosti jejich oprav během testování. Využíváním nástroje pro profilování dat získáte také dostatečný přehled o datech, se kterými pracujete, díky čemuž můžete odstranit chyby a zvýšit tak jejich kvalitu ještě před samotným transformováním.

2.2 Pro čištění dat používejte nástroje pro čištění dat. Nepokoušejte se o čištění ručně [1]

Studie společnosti Bloor Research [1] uvádí, že společnosti, které v dané studii **během řešení migračního projektu nevyužívali nástroje pro čištění dat, měli pouze 55% šanci na úspěšné dokončení** projektu. Jak již bylo zmíněno v předchozí kapitole, špatná kvalita dat je jednou z příčin neúspěšně dokončeného dokončení projektu. Tyto nástroje vám mohou pomoci vyčistit data od duplicit nebo chybných formátů dat, a zvyšovat tak jejich kvalitu.

Abyste mohli s fází čištění začít, je nejprve nutné **získaná zdrojová data analyzovat.** Důkladnou analýzou můžete odstranit mnoho chyb v datech již v jejich počátcích. S čištěním dat tak začnete co nejdříve. Chyby, které v datech v počátcích neodstraníte se následně mohou objevovat během fáze testování, což může prodloužit tuto fázi a být tak v rozporu s projektovým plánem.

Důležité také je, abyste data **nečistili dle vlastního uvážení, pokud jim dostatečně nerozumíte.** Pokud v projektu vystupujete například jako externí systémový integrátor, většinou nemáte dostatečný přehled o byznysovém významu dat, se kterými pracujete. Je proto vhodné dohodnout se se zaměstnanci klienta, kteří datům, se kterými pracují, rozumí, jakým způsobem data čistit. Nikdy nečistěte data dle svého vlastního uvážení.

V neposlední řadě **využívejte nástroje pro čištění dat** a nepokoušejte se o manuální čištění. Zejména při zpracování rozsáhlých datových souborů vyžaduje manuální čištění nadměrnou časovou náročnost

a je náchylnější ke vzniku chyb. V případě, že by došlo k migraci více sad dat, bylo by nutné každou tuto sadu ručně čistit. Tento proces pomohou nástroje pro čištění dat usnadnit.

2.3 Používejte nástroje pro integraci dat. Nepokoušejte se o ruční kódování [1] [13]

Podobně jako v případě nástrojů pro čištění dat, by měly být **používány také nástroje pro integraci dat**. Tyto nástroje umožní automatizovat některé úlohy a vyhnout se tak ručnímu kódování, které může být zdoluhavější, náročnější na provedení, a náchylné na chyby. Společnosti, které integrační nástroje ve studii [1] nevyužívaly, měly pouze 50% šanci na úspěšně dokončení projektu.

Nástroje pro datovou integraci mají v migračním projektu svoji nezpochybnitelnou roli. Ruční kódování s sebou nese riziko v podobě nesouladu a obtížné údržbě kódu. Každý vývojář v tomto případě vytváří kód podle svého uvážení, což může vést k nekonzistenci výsledného produktu.

Využívejte proto **centrální nástroje pro datovou integraci**. Ty vám navíc mohou pomoci generovat metadata a šablony stejném stylu pro každý mapping, což vám zajistí konzistenci a snadnou orientaci, i když bude každý mapping vytvářet někdo jiný. Díky tomu budete moci kontrolovat a upravovat práci kolegů, když to bude potřeba. Využíváním integračních nástrojů dojde nejen k úspoře času, ale také k minimalizaci rizika vzniku chyb a větší konzistenci mezi jednotlivými mappingy, což oceníte například při jejich kontrole.

2.4 Ujistěte se, že pro migraci dat používáte správný software [12] [15] [17] [19]

Produktů, které je možné k migraci využít, je celá řada. Každý produkt obsahuje různé funkcionality a nabízí se za různé ceny. Důkladně proto zvažte, který softwarový produkt nejlépe odpovídá vašim požadavkům.

Existují případy, kdy mají **společnosti vyvinuté vlastní nástroje pro tyto účely**. Zaměstnanci jsou s těmito nástroji seznámeni, což následně usnadňuje jejich použití během migračního projektu. Ne vždy je ale vhodné interní nástroj pro migraci dat využít, jelikož ne vždy tento nástroj splňuje všechny vaše požadavky. Pokud tomu tak je, vyberte jiný nástroj, který je na trhu k dispozici.

2.5 Je důležité zvážit, zda je možné migraci snadno ukončit a znova spustit [12]

Požadavkem, který je třeba zvážit, je **možnost vrátit migraci zpět**. V podstatě jde o to, že pokud se během migrace něco pokazí, je možné migraci ukončit a spustit znovu. To může být u některých technologií problematické, zatímco u jiných je to snadné [12].

Během migračního projektu je **zásadní, abyste měli možnost provést rollback migrace**. Existuje několik způsobů, jak toho dosáhnout. Pravidelně si vytvářejte **zálohu migrační databáze**, abyste v případě jejího smazání nebo poškození mohli data obnovit ze zálohy. Zároveň ukládejte různé verze záloh a nepřepisujte stále jen jednu zálohu. Tím získáme více možností pro provedení rollbacku v případě potřeby.

Dalším způsobem, který vám umožní udělat částečný rollback, je **ukládání transformačních skriptů pomocí verzování**. Při každé změně transformačního kódu ukládejte tyto změny do nové verze mapování a ponechte si verze původní. V případě, že vaše poslední verze mapování způsobí v databázi nežádoucí změny v datech, můžete využít verze původní a obnovit data do původního stavu. Pokud váš migrační nástroj umožňuje tuto funkcionalitu, je to nesporná výhoda.

3. Práce s daty

3.1 *Pochopte data ve zdrojovém systému ještě předtím, než začnete migrovat [6] [15]*

Společnosti často podceňují důležitost posouzení stavu dat před samotným procesem migrace. **Datová kvalita má významný vliv na úspěch migrace, a proto plně pochopte zdrojová data a jejich struktury, než začnete s jejich transformací.** Ptejte se, komu budou data sloužit, co s nimi hodlají uživatelé dělat, či zda v nich něco nechybí. To může výrazně usnadnit následný proces migrace. Správnost mapování dat do cílového systému závisí kromě mnoha dalších faktorů právě na pochopení, a proto tento krok nepodceňte.

Jedním z prvních kroků po získání zdrojových dat je jejich analýza. Zjistěte množství a účel entit, jejich atributy, vazby mezi entitami, formáty dat, primární a cizí klíče, datové typy nebo datové objemy. Pokud si něčím nejste jisti nebo nechápete význam poskytnutých dat, vždy tyto skutečnosti konzultujte a nepokoušejte se o vlastní řešení. To platí také v případě, pokud jste externím systémovým integrátorem, který nemá hlubší byznysové porozumění migrovaných dat a plně nechápe jejich využití. **Vždy proto identifikujte klíčové osoby, které migrovaným datům rozumí** a snažte se správně pochopit význam dat, aby došlo k jejich správné migraci. Nesprávné pochopení dat povede k chybné transformaci a následným problémům během testování.

Pouhé pochopení zdrojových dat tedy nestačí. **Pochopit musíte také struktury cílového systému,** do kterých data migrujete. Pokud budete chápat pouze zdrojové struktury, nikoliv však cílové, je málo pravděpodobné, že se vám transformace dat povede tak, jak očekáváte. V případě nejasností vše konzultujte s klientem nebo dodavatelem systému, kteří mají hlubší povědomí o datech.

3.2 *Zaměřte se na správné mapování dat, aby byl zajištěn kompletní přenos dat [16]*

Mapování je proces, během kterého **se přiřazují atributy původních datových struktur k atributům cílových datových struktur.** Kromě správného mapování zjistěte, zda se shodují datové typy, délky znaků jednotlivých atributů nebo povinnosti atributů. Kvalita mapování má přímý vliv na kvalitu výsledných dat a platí, že čím kvalitněji mappingy provedete, tím menší je pravděpodobnost výskytu chyb v datech.

Správné mapování dat je úzce souvisí s jejich správným pochopením. Při mapování je důležité mít detailní znalosti o struktuře a významu původních dat, stejně jako o očekávané struktuře a požadavcích cílového systému. Proveďte proto před samotným mapováním důkladnou analýzu datových struktur.

Kvalitu mapování nelze jednoznačně posoudit pouhým pohledem na mapované atributy, a proto je nezbytné **provést důkladné testování.** Přesto je možné snížit riziko nesprávného mapování. Pokud se objeví situace, kdy není jasné, jak daný atribut mapovat do cílového systému, konzultujte tuto skutečnost s někým, kdo má znalosti týkající se migrovaných dat a cílových struktur. Tyto kroky mohou přispět k minimalizaci rizik spojených s nesprávným mapováním atributů.

3.3 *Rámec pro řízení dat má zásadní vliv na celý migrační projekt [1] [6]*

Je vhodné, aby společnosti měly **zavedený rámec pro řízení dat bez ohledu na to, zda souvisí s migračním projektem** [1]. Pokud nemáte takový rámec, může pro vás projekt migrace představovat dobrý odrazový můstek. Kvalitní rámec pro řízení dat pro vás bude nespornou výhodou, pokud se k datové migraci odhodláte. **Rámec vám umožňuje vyšší kvalitu dat,** což má velký vliv na průběh migračního projektu. Bez rámce pro řízení dat je pravděpodobné, že se bude kvalita dat časem zhoršovat. To bude v případě migračních nebo jiných projektu vyžadovat větší úsilí během jejich čištění. Mějte proto zavedený rámec pro řízení dat a pokud ho ještě nemáte, zvažte jeho implementaci. Implementace rámce pro řízení dat už je ale na samostatný projekt, který vyžaduje specifické požadavky.

3.4 Pravidelně zálohujte data pro případ, že by během procesu migrace došlo k problému [17] [19]

Záloha dat zajišťuje, že pokud během datové migrace **dojde k problémům, či chybám, které způsobí ztrátu dat, zůstanou i přesto tato data bezpečně zálohována** a dostupná pro další práci. Zálohu dat provádějte pravidelně během migračního projektu. Nemělo by se vám stát, že poslední záloha je několik dní nebo týdnů stará. Vytvářejte si také různé verze záloh a nepřepisujte jednu zálohu stále dokola. V případě potřeby tak můžete vybrat, jakou zálohu obnovíte.

Pokud to váš migrační nástroj umožňuje, **zálohujte také transformační kód pomocí verzování**. Tím získáte různé verze transformačního kódu, které je možné v případě potřeby obnovit, stejně jako v případě zálohování databáze. Zálohování transformačního kódu je klíčové pro minimalizaci dopadů chyb či ztráty dat.

Provádějte také kontroly zálohy dat a testujte její obnovení, abyste ověřili, že jsou zálohy správně vytvořené a fungují, jak mají. To pomáhá minimalizovat riziko překvapení během obnovy.

3.5 Identifikujte formáty dat, které budete přenášet [17] [19]

Během migračních projektů se může stát, že datové typy ve zdrojovém a cílovém systému nejsou shodné. Může také nastat situace, kdy zdrojový atribut povoluje více znaků než atribut v cílovém systému, což může způsobit problémy při přenosu některých dat. Proveďte proto **analýzu datových struktur v počáteční fázi migračního procesu, abyste identifikovaly formáty, umístění, citlivost dat, entity**, atributy, vazby mezi entitami, primární a cizí klíče, datové typy, objemy, se kterými budeme pracovat.

Pokud se objeví situace, kdy například zdrojový atribut neodpovídá formátu nebo pravidlům atributu v cílovém systému, **řešte co nejrychleji tyto problémy ve spolupráci s ostatními stranami**, které se na projektu podílejí, a diskutujte vhodné řešení. Nepokoušejte se o řešení sami bez předchozí konzultace.

3.6 Nahrajte data, aby bylo možné zahájit proces migrace [17]

Aby bylo možné spustit proces migrace, je nezbytné **zajistit dostupnost dat, která budou přenášena do nového systému**. Data musí být získána s ohledem na jejich citlivost a zabezpečení. Domluvte si vhodný způsob předávání dat, který minimalizuje riziko změny, poškození nebo ztráty. Nezapomeňte také připravit prostředí, do kterého budou data nahrána.

Pochopte entity ve zdrojovém a cílovém systému, jejich atributy, vazby mezi entitami, formáty dat, primární a cizí klíče, datové typy a objemy dat. To je nezbytné pro správnou analýzu dat. V případě nejasností komunikujte problémy se všemi relevantními stranami projektu a vyjasněte si je.

3.7 Investujte do následné údržby, abyste se ujistili, že byla zachována integrita dat [17]

Tento postup neplatí, pokud jste například externí systémový integrátor, který daný systém nespravuje. Pokud však jako klient nebo dodavatel **vlastníte informační systém, do kterého byla data migrována, je vhodné po dokončení migrace zajistit, aby migrovaná data zůstala v bezproblémovém stavu** a nedošlo k jejich ztrátě. Pro minimalizaci tohoto rizika investujte do následné údržby systému. To umožní zlepšení a optimalizaci výkonu, zajištění bezpečnosti, odstranění chyb a prodloužení životnosti systému. Je také vhodné vytvořit rámec pro řízení dat v případě, že ho ještě nemáte. To pomůže stanovit pravidla pro ukládání dat a zvýšit jejich kvalitu.

3.8 Omezte počet zprostředkovatelů (vrstev abstrakce) [15]

Při přesunu dat z jednoho systému do druhého nebo z jednoho formátu do jiného hrozí, že se něco ztratí nebo změní. Pokud jsou například data v souborech .xlsx, vy je otevřete v aplikaci Excel a

rozhodnete se je uložit ve formátu .csv, aplikace Excel může vypustit počáteční nuly nebo převést text na čísla a přidat desetinná místa. Výsledkem tak mohou být soubory, které nejsou správně naformátované. **Vyhnete se proto otevírání dat v kancelářských balíčcích. Zároveň požadujte, aby byla data dodávána v původním formátu a snažte se omezit počet zprostředkovatelů.** To je jedním z klíčových faktorů pro udržení dat v původní podobě.

Požádejte také společnost, která zdrojová data posílá, o **vytvoření a zdokumentování postupu exportu dat a posílání**. To umožní standardizaci a minimalizaci rizika změn v situacích, kdy data exportuje někdo jiný. Pokud by se však stalo, že data v souboru obsahují jiné formáty než posledně, reportujte tuto skutečnost společnosti, která vám data poskytl.

Také vy si **vytvořte dokumentaci pro nahrávání datových souborů do databáze**. Tímto krokem zajistíte standardizaci postupu při nahrávání dat. S pečlivě zdokumentovaným postupem budete vždy postupovat stejně a zajistíte, že data budou nahrána tak, jak potřebujete.

3.9 Vytvořte si knihovnu opakovaně použitelného kódu [15]

Některé transformace se musí provádět opakovaně, a to nejen v rámci jednoho projektu. Pište proto **kód tak, aby se dal používat opakovaně**. Příkladem takové transformace je formátování telefonních čísel [15]. Vytvoření knihovny opakovaného kódu vám ušetří mnoho času a pomůže minimalizovat chyby. **Knihovna může obsahovat kódy pro různé části projektu, včetně skriptů pro nahrávání a transformaci dat**, funkcí pro změny formátování záznamů v attributech, nebo kódů pro spouštění procedur. Díky opakovaně použitelnému kódu snadněji udržíte konzistenci v transformacích dat a budete schopni aplikovat změny na více místech najednou.

3.10 Opravujte kód, ne data [15]

Během procesu migrace je možné narazit na data, která je nutná opravit. **Opravujte je pomocí transformačního kódu, neupravujte data samotná**. To zajistí, že v případě nové migrace budou nová data správně transformována a nemusí tak docházet k jejím dalším opravám [15]. Tento přístup nejen šetří čas a snižuje riziko chyb, ale také zajišťuje snadnější sledování provedených úprav. Pokud byste upravovali přímo zdrojová data, mohlo by se časem stát obtížným identifikovat, která konkrétní data byla upravena.

V souvislosti s tím je také důležité uvést, **abyste neupravovali kód tak, aby jeho podmínky platily pro konkrétní výjimky ve zdrojových datech**. Naprogramujte obecnější podmínky, které lze aplikovat na každý záznam atributu. Pokud se v migrovaných datech objeví například některé emailové adresy obsahující dva zavináče, není vhodné vytvářet v transformačním kódu výjimku pro tyto konkrétní emailové adresy, ale pro všechny adresy, které mohou dva zavináče obsahovat. Tento přístup vám umožní snadnou aplikaci transformačního kódu na různé sady dat.

4. Testování

4.1 Provedte v cílovém systému testování, abyste se ujistili, že byla data úspěšně přenesena [17] [19]

Testování tvoří podstatnou část migračního procesu a nemělo by být opomíjeno. Právě **testování ověřit, že byla data namigrována do cílového systému v požadovaném stavu a neobsahují chyby, duplicitu**, neodpovídající datové typy, které by mohly mít negativní vliv na následné rozhodování podniku. Zajistí se také to, že budou migrovaná data bezpečná a na správném místě [19].

Testování je fází, kterou se nevyplatí podceňovat a vždy si pro ni **vyhradte dostatek času**. Množství chyb, které se v datech mohou objevit, ovlivňuje množství nutných oprav, s čímž je nutné počítat.

Testování samotných dat ne vždy ověří, zda byla data migrována správně do cílového systému a zda jsou v požadovaném stavu. Samotná data se mohou zdát na první pohled v pořádku, nemusí však odpovídat kontextu ostatních dat. **Testujte proto data v kontextu zbylých dat.** Dále testujte byznys procesy nad těmito daty. Bez těchto činností nepoznáte, zda jsou migrovaná data v požadovaném stavu.

Důležité je, abyste vy, nebo jakákoliv strana zodpovědná za testování, měly **připravené vhodné testovací prostředí, které co nejvíce odpovídá reálným podmínkám cílového systému.** Tato skutečnost vám pomůže odhalit chyby v datech a zjistit, zda jsou data migrována správně.

4.2 Pravidelně během migračního projektu testujte [16]

Testování je dalším krokem, který hraje v migračním projektu významnou roli, jelikož **pomáhá identifikovat a odstranit chyby v datech, které ovlivňují datovou kvalitu.** Může se stát, že nad daty v původním systému nebyl aplikován rámec řízení dat, který se ovšem v cílovém systému nachází. Vznikají tak nové požadavky na migrovaná data a je potřeba otestovat, zda data tyto požadavky splňují.

Začněte také **s testováním co možná nejdříve a vyhradte si na něj dostatek času.** Během testovací fáze je pravděpodobné, že se v datech objeví velké množství chyb, které bude následně nutné opravit. To může mít za následek prodloužení fáze testování a následný časový skluz.

Vytvořte si také **jednotné místo, kam se budou ukládat veškeré nalezené chyby (tzv. bugy).** Neposílejte si informace o chybách v datech a jejich řešení prostřednictvím emailů. Brzy se v těchto vláknech začnete ztrácet a přestanete tak mít přehled o řešených chybách. Použití specializovaných nástrojů pro správu chyb umožní sledování, přiřazování odpovědností, prioritizaci a zaznamenávání stavu řešení chyb. Tím zvýšíte přehlednost a efektivitu během průběhu testování.

Vytvořte **po každé opravené chybě kód, který ověří, zda byla chyba v datech opravena** a následně vytvořte tabulku, která bude sloužit jako souhrn těchto kódů, ověřujících, zda jsou data opravena. Tímto postupem získáte přehled o tom, zda se opravené chyby v datech nevyskytují znova.

4.3 Musí být provedeno systémové a uživatelské testování a před spuštěním ostrého provozu je nutné získat podpis [15]

Fáze testování by měla být **rozdělena na systémovou a uživatelskou část. Jako první proveďte testování systémové,** kdy ověříte, zda jsou data správně integrována do cílového systému, a respektují jeho strukturu a pravidla. Jakmile je systémové testování dokončené, přejděte na testování uživatelské. To mohou provádět pouze skuteční uživatelé systému. **Uživatelské testování je důležité, protože uživatelé mají přímý kontakt s daty a mohou odhalit chyby a nedostatky,** které by jinak mohly být přehlédnuty.

Před spuštěním samotné migrace je nezbytné, aby proběhlo **její schválení a došlo k podpisu zúčastněných stran projektu,** kterými mohou být například zástupci klienta a dodavatele. Podpisem dojde k potvrzení, že zúčastněné strany souhlasí s migrací dat a uvedením systému do ostrého provozu. Podpis zároveň zajišťuje, že jsou všechny strany o datové migraci informovány a souhlasí s ní.

Důležité je uvědomit si, že pokud data nebudou odpovídat požadavkům klienta a dodavatele systému (v případě, že je v projektu přítomen), může se stát, že souhlas k migraci nedají, například z důvodu chyb v datech, a migrace neproběhne. Zajistěte proto, **aby byla data v požadované kvalitě, a to ideálně v termínu, který je pro migraci stanoven.**

4.4 Plánujte prodloužené období systémových a uživatelských testů [15]

V rámci testování může být nalezeno mnoho chyb. Zároveň je těžké odhadnout, kolik času budou mít uživatelé v rámci uživatelského testování. **Naplánujte proto dostatečný čas na celou fázi testování, včetně systémového a uživatelského testování.** Pokud migrujete data do systému dodavatele, je důležité, aby také on byl na systémové testování připraven a vyhradil si dostatek času. **Informujte také uživatele o začátku uživatelského testování a zahajte testování co nejdříve,** aby bylo možné včas

odhalit všechny chyby. V neposlední řadě si naplánujte si prodloužené období pro systémové a uživatelské testování.

4.5 Vytvořte si vztah s uživateli, kteří testují [15]

Pokud se data migrují do vašeho systému a nacházíte se ve fázi uživatelského testování, je důležité **vysvětlovat uživatelům, co děláte a proč se ptáte na určité otázky**. Projevujte zájem o problémy uživatelů a ochotu jim pomoci [15]. Pokud se na projektu podílíte jako externí systémový integrátor, máte o starost méně. V kontaktu s uživateli je především strana, do jejichž systémů jsou data migrována. Vytvořte si **komunikační linku s klientem, který s vámi bude sdílet informace od uživatelů**. V případě potřeby je důležité, abyste klienta v komunikaci podporovali, například vysvětlováním nebo řešením vzniklých problémů.

4.6 Ujistěte se, že váš kód loguje vše, co dělá [15]

Definujte, **co se bude během projektu datové migrace logovat**. Zároveň se ujistěte, že kód zaznamenává vše, co dělá: kdy se proces spustí, kdy se dokončí, počty zpracovaných záznamů, ale také informace o tom, jaké kroky byly v průběhu migrace provedeny, kdo je provedl a jaké byly výsledky každého kroku. Když se následně zjistí **problém s migrací, většina klientů vás požádá, abyste prošetřili, co se pokazilo a proč**. Mít k dispozici logy vám může ušetřit dny práce [15].

Logování však není důležité jen pro dohledávání chyb, ale slouží také auditorům během auditu, kdy může pomoci rekonstruovat průběh migrace dat. **Logy proto zahrňte také do dokumentace projektu**. Také ostatní zainteresované strany budou mít přehled o tom, jaké kroky byly během migračního projektu prováděny.

4.7 Získejte novou sadu dat pro každé kolo uživatelských testů [15]

Během projektu datové migrace **získejte novou sadu dat pro každé kolo uživatelských testů**. Nezapomeňte, že chcete proces otestovat kompletně od začátku do konce. To znamená, že musíte otestovat nejen své postupy nasazení a kód, ale také schopnost exportovat data konzistentním a přesným způsobem [15].

Podobně jako bylo popsáno v postupu, který se týkal minimalizace počtu prostředníků, i zde je vhodné si **zdokumentovat, jakým způsobem budete data exportovat**. Vytvořte si dokument, který krok po kroku popíše tento postup. To umožní standardizaci a minimalizaci rizika změn v situacích, kdy data exportuje někdo jiný a exportovaná sada dat je stále ve stejném formátu.

5. Rozpočet

5.1 Je-li migrace součástí většího projektu, řešte ji zcela samostatně, a to od sestavování rozpočtu až po testování [1]

Existence samostatného rozpočtu je důležitá, pokud je migrace součástí většího projektu. Podle [1] snižuje neoddělení rozpočtu migrace od celkového projektu šanci na úspěšné dokončení projektu až 4x. Kromě odděleného rozpočtu od celkového projektu by měla být migrace také odděleně testována. Je-li testování datové migrace začleněno do celkového testování, snižuje se pravděpodobnost úspěšného dokončení projektů včas nebo v rámci rozpočtu.

5.2 Zhodnoťte velikost projektu a zjistěte, jaké zdroje bude vyžadovat a kolik na něj potřebujete v rozpočtu [15] [17] [19]

Aby bylo možné projekt realizovat, stanovte **rozsah aktivit, které je nutné v rámci projektu provést**. Rozsah projektu může být ovlivněn mnoha faktory. **Zvažte počet systémů, které je potřeba migrovat a také na počet systémů, do kterých máte migraci provést**. Dále zvažte rozdílnost datových struktur. V případě, že jsou datové struktury 1:1, zabere jejich transformace méně času, než když jsou struktury rozdílné. To má samozřejmě také vliv na pracnost a časovou náročnost projektu, a tedy i na rozpočet. Dále **berte ohled také na komplexnost projektu**.

V případě, že se rozhodnete pro koupi zahraničního portfolia (popřípadě pokud migrujete zahraniční portfolio jako externí systémový integrátor), může dojít k situaci, kdy se jednotlivé právní normy obou zemí liší. Rozdílné jsou tak procesy, číselníkové hodnoty a další skutečnosti, které mohou mít následně vliv na pracnost. Všechny tyto faktory proto zvažte, plánujte a stanovte na základě nich rozpočet. Zároveň **určete vhodné zdroje pro realizaci projektu, včetně lidských zdrojů** (vývojáři, analytici, testéři atd.) a technologií (hardware a software). Jsou-li známy rozsah a zdroje, stanovte náklady projektu a na základě toho poté stanovte rozpočet projektu.

6. Použitá literatura

- [1] HOWARD, Phillip. Data Migration – 2011. [online]. 2011. [cit. 2023-01-11]. Dostupné z: <https://www.bloorresearch.com/restapi/download/YrXfE8OGLjFOZi1SheSH9fterlzCB7ZqN91HIYk7Op7tgplC>
- [2] Balakrishnan Tara, Gnanasambandam Chandra, Santos Leandro, Srivathsan Bhargs. Cloud-migration opportunity: Business value grows, but missteps abound. [online]. 2021. [cit. 2023-01-11]. Dostupné z: <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/cloud-migration-opportunity-business-value-grows-but-missteps-abound>
- [3] MAHANTI, Rupa. Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools. 2021. [cit. 2023-01-20]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6723145?accountid=17203>.
- [4] MAHANTI, Rupa. Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. 2019. [cit. 2023-01-23]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6262212?accountid=17203>.
- [5] AZEROUAL, Otmane a Meena JHA. Without Data Quality, There Is No Data Migration. Big Data and Cognitive Computing [online]. 2021, vol. 5, no. 2, s. 24. [cit. 2023-01-24]. Dostupné z: <https://www.proquest.com/scholarly-journals/without-data-quality-there-is-no-migration/docview/2544472698/se-2>
- [6] BERTOLUCCI, Jeff. 10 Big Data Migration Mistakes. Informationweek – Online [online]. 2012. ISSN 19383371. [cit. 2023-01-24]. Dostupné z: <https://www.proquest.com/trade-journals/10-big-data-migration-mistakes/docview/1032866116/se-2>.
- [7] WORSCH, Filip. Fúze společností z hlediska projektu migrace dat [online]. Praha, 2013 [cit. 2023-05-25]. Dostupné z: <https://theses.cz/id/isa7tp/>. Diplomová práce. Vysoká škola ekonomická v Praze.
- [8] NĚMCOVÁ, Alžběta. Vytvoření a aplikace metodiky pro hodnocení kvality provedení datové migrace [online]. Praha, 2015 [cit. 2023-01-25]. Dostupné z: <https://theses.cz/id/xrsh1r/>. Diplomová práce. Vysoká škola ekonomická v Praze.
- [9] CLÉMENT, Delphine & Hassine-Guetari, Soumaya & Laboisie, B. Data Quality as a Key Success Factor for Migration Projects. 2012. [cit. 2023-01-25]. Dostupné z: https://www.researchgate.net/publication/280751278_Data_Quality_as_a_Key_Success_Factor_for_Migration_Projects
- [10] MORRIS, J. Practical Data Migration, 2020. [cit. 2023-01-25]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6370586?accountid=17203>.

[11] CHEMUTURI, M. Mastering IT Project Management: Best Practices, Tools and Techniques. [online]. 2013. [cit. 2023-01-25]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/3319551?accountid=17203>

49

[12] IBM. Best practices for data migration. [online]. 2009. [cit. 2023-01-25]. Dostupné z: https://www.academia.edu/9778539/IBM_Global_Technology_Services_Best_practices_for_data_migration_Methodologies_for_assessing_planning_moving_and_validating_data_migration_Best_practices_for_data_migration_Contents

[13] RUSSOM, P. Best Practices in Data Migration. 2006. [cit. 2023-01-25]. Dostupné z: http://download.101com.com/pub/TDWI/Files/TDWI_Monograph_BPInDataMigration_April2006.Pdf

[14] TSAKUNOVA, Margarita. Migrace datového skladu z MS SQL Server on-premise prostředí do cloudu [online]. Praha, 2022 [cit. 2023-01-25]. Dostupné z: <https://theses.cz/id/ayyil/>. Diplomová práce. Vysoká škola ekonomická v Praze.

[15] MASRI, D. Developing Data Migrations and Integrations with Salesforce: Patterns and Best Practices, 2018. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/5622564?accountid=17203>.

[16] NOREEN, Neysa, R.H.I.A. Data Migration Lessons Learned. Journal of AHIMA [online]. 2013, vol. 84, no. 9, s. 40-42. ISSN 10605487. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/trade-journals/data-migration-lessons-learned/docview/1437254005/se-2>.

[17] IntelligentHQ. Why is data migration important? [online]. London: Newstex, 2020. Copyright – Copyright Newstex Nov 4, 2020; Poslední aktualizace - 2022-10-19. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/blogs-podcasts-websites/why-is-data-migration-important/docview/2457158229/se-2>.

[18] SARMAH Simanta Shekhar. Data migration. [online]. [cit. 2023-01-26]. Dostupné z: <http://article.sapub.org/10.5923.j.scit.20180801.01.html>

[19] Abou_el_ela Abdou Hussein. Data Migration Need, Strategy, Challenges, Methodology, Categories, Risks, Uses with Cloud Computing, and Improvements in Its Using with Cloud Using Suggested Proposed Model (DMig 1). [online]. [cit. 2023-01-26]. Dostupné z: <https://www.scirp.org/journal/paperinformation.aspx?paperid=106592>

[20] ZKOUMALOVÁ, Barbora. Migrace systémové databáze elektronického obchodu [online]. Brno, 2016 [cit. 2023-01-27]. Dostupné z: <https://theses.cz/id/fkgh2e/>. Diplomová práce. Vysoké učení technické v Brně.

[21] Sebastian-Coleman, L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. [online]. 2013. [cit. 2023-01-28]. Available from: <https://www.proquest.com/legacydocview/EBC/1106491?accountid=17203>.

[22] Microsoft Azure. What is Data Migration?. [online]. [cit. 2023-01-28]. Dostupné z: <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-data-migration/#defining-data-migration>

[23] IBM. What is data governance?. [online]. [cit. 2023-01-28]. Dostupné z: <https://www.ibm.com/topics/data-governance>

50

[24] MAHANTI, R. Data Governance Success: Growing and Sustaining Data Governance. 2021. [online]. [cit. 2023-01-30]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6826318?accountid=17203>.

[25] WRIGHT, G. Best Practise. [2022]. [cit. 2023-03-20]. Dostupné z: <https://www.techtarget.com/searchsoftwarequality/definition/best-practice>