

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky



Business a datové slovníky v rámci implementace Data Governance

DIPLOMOVÁ PRÁCE

Studijní program: Data a analytika pro business

Autor: Bc. Ondřej Musil

Vedoucí diplomové práce: doc. Ing. Jan Pour, CSc.

Praha, září 2024

Poděkování

Na tomto místě bych rád poděkoval vedoucímu diplomové práce, doc. Ing. Janu Pourovi, CSc., za cenné rady, konstruktivní připomínky a odborné vedení během celého procesu zpracování práce, které výrazně přispěly ke kvalitě této práce. Dále děkuji společnosti BNP Paribas Cardif za možnost řešit diplomovou práci na reálném projektu, za poskytnutí nezbytných podkladů a vstřícnost během celého procesu. V neposlední řadě bych chtěl poděkovat i svým kolegům za spolupráci, zpětnou vazbu a motivaci, která mi pomáhala práci dotáhnout do úspěšného konce.

Abstrakt

Diplomová práce se zabývá problematikou implementace byznys a datových slovníků v rámci konceptu data governance ve společnosti BNP Paribas Cardif. Cílem práce je zdokumentovat celý proces implementace – od úvodní analýzy a návrhu až po praktické zavedení – a nabídnout realizovatelná doporučení, jak čelit lidským, technickým a procesním výzvám, které jsou s tímto typem projektů spojeny.

Práce kombinuje teoretické poznatky z oblasti správy dat s reálným projektem v konkrétním firemním prostředí. Teoretická část poskytuje přehled klíčových oblastí data governance a předpokladů pro zavedení kvalitní datové dokumentace. Praktická část popisuje jednotlivé kroky realizace, včetně návrhu struktury dokumentace, tvorby definic obchodních pojmů a mapování mezi aplikačními a obchodními daty.

Součástí výstupu je funkční byznys a datový slovník doplněný o návrh dalšího rozvoje a identifikaci hlavních bariér včetně stručného zamyšlení nad využitím nástrojů umělé inteligence pro překlenutí lidských bariér. Práce se zaměřuje zejména na lidské a procesní komponenty správy dat, které bývají v praxi často opomíjeny.

Výsledkem je dokument, který může sloužit jako přehledný referenční materiál pro další organizace nebo jednotlivce, kteří čelí zavádění data governance, a to jak z pohledu teorie, tak i z hlediska praktické realizace.

Klíčová slova

Správa dat, řízení dat, datová dokumentace, byznys slovník, datový slovník, datový katalog, metadata, datový model, implementace.

Abstract

The thesis is focused on implementation of business glossaries and data dictionaries within the concept of data governance in BNP Paribas Cardif. The aim of the thesis is to document the entire implementation process - from initial analysis and design to practical implementation - and to offer actionable recommendations to address the human, technical and process challenges associated with this type of project.

It combines theoretical knowledge in the field of data governance with a real project in a specific corporate environment. The theoretical part provides an overview of the key areas of data governance and the prerequisites for implementing good data governance. The practical part describes the individual steps of implementation, including the design of the documentation structure, the creation of business term definitions and the mapping between application and business data.

The deliverable includes a functional business glossary and data dictionary complemented by a proposal for further development and identification of the main barriers, including a brief reflection on the use of artificial intelligence tools to bridge human barriers. The work focuses in particular on the human and process components of data governance, which are often neglected in practice.

The result is a document that can serve as a clear reference for other organizations or individuals facing the implementation of data governance, both in terms of theory and practical implementation.

Keywords

Data governance, data management, data documentation, business glossary, data dictionary, data catalog, metadata, data model, implementation.

Obsah

Úvod	12
1 Rešerše literatury	17
1.1 RO1: Jaký je současný stav poznání zvoleného tématu?	17
1.1.1 Rešerše vysokoškolských kvalifikačních prací	17
1.1.2 Rešerše knižních publikací a vědeckých článků	18
1.1.3 Rešerše internetových zdrojů	19
1.2 RO2: Jaké znalostní oblasti v rámci řízení dat definuje DAMA-DMBOK2?	19
1.2.1 DAMA-DMBOK2 Data Management Framework	20
1.2.2 Datová architektura	22
1.2.3 Datové modelování a design	26
1.2.4 Metadata	28
1.3 RO3: Jaké jsou hlavní komponenty data governance?	32
1.3.1 Lidé	35
1.3.2 Procesy	39
1.3.3 Nástroje a technologie	42
1.4 RO4: Jaký je význam byznys a datových slovníků a jaké jsou typické výzvy při jejich implementaci v kontextu DG?	43
1.4.1 Business Glossary	43
1.4.2 Data Dictionary	44
1.4.3 Lidské výzvy	46
1.4.4 Organizační a procesní výzvy	50
1.4.5 Technologické výzvy	52
1.5 Shrnutí kapitoly rešerše literatury	52
2 Současný stav	54
2.1 Představení společnosti	54
2.2 Aktuální stav správy dat	54
2.3 Shrnutí kapitoly současný stav	56
3 Implementace byznys a datových slovníků v BNP Paribas Cardif	58
3.1 Používané nástroje a technologie	59
3.1.1 Modelování a datová báze	59
3.1.2 Správa a řízení kmenových dat	61
3.1.3 Úložiště metadat	61
3.2 Analýza stávajícího stavu dokumentace	62

3.2.1 Aplikační data	65
3.2.2 Obchodní data	69
3.3 Návrh dokumentace	71
3.3.1 Metadata alignment	71
3.3.2 Metadata improvement	73
3.4 Implementace návrhu	82
3.4.1 Datový slovník	82
3.4.2 Byznys slovník	89
3.5 Zhodnocení přínosů a doporučení pro další rozvoj	90
3.5.1 Přínosy	90
3.5.2 Překážky a omezení	91
3.5.3 AI jako nástroj pro překlenutí lidských bariér	92
3.5.4 Doporučení pro další rozvoj	97
3.6 Shrnutí kapitoly implementace byznys a datových slovníků	98
Závěr	100
Použitá literatura	102

Seznam obrázků

Obrázek 1 DAMA-DMBOK2 Wheel (DAMA International 2017).....	21
Obrázek 2 Enterprise Data Model (DAMA International 2017).....	23
Obrázek 3 Horizontální závislost mezi datovými modely (DAMA International 2017)	24
Obrázek 4 Matice datových toků (DAMA International 2017).....	25
Obrázek 5 Diagram datových toků (DAMA International 2017).....	25
Obrázek 6 Přehled úkolů a výsledků návrhu databáze (Simsion a Witt 2004).....	27
Obrázek 7 Ukázka toku na úrovni datového prvku (DAMA International 2017).....	31
Obrázek 8 Ukázka toku na úrovni systémů (DAMA International 2017)	31
Obrázek 9 Komponenty správy dat a jejich vzájemné vztahy (Mahanti 2021)	33
Obrázek 10 Dopady chybějících Data Governance komponent (Mahanti 2021)	34
Obrázek 11 Obecně přijímané informační principy (Ladley 2012).....	40
Obrázek 12 Termín v byznys slovníku (Atlan 2023a)	44
Obrázek 13 Jednoduchý datový slovník (QIC WD 2020)	46
Obrázek 14 Obecný přehled datové architektury v BNPP Cardif (interní dokumentace)...	60
Obrázek 15 Příklad hlavních dimenzí OLAP kostky v BNPP Cardif (interní dokumentace)	60
Obrázek 16 Centralizovaná metadata architektura (DAMA International 2017)	62
Obrázek 17 Metadata alignment mezi datovými objekty a byznys termíny (vlastní zpracování).....	73
Obrázek 18 Pořadí atributů ve faktových tabulkách (vlastní zpracování).....	74
Obrázek 19 Naznačení datového toku pro ilustrační tabulky (vlastní zpracování).....	78
Obrázek 20 Proces implementace datového modelu (vlastní zpracování)	84
Obrázek 21 WhereScape 3D – Souhrn změn při slučování datových modelů (vlastní zpracování).....	84
Obrázek 22 WhereScape 3D – Nový datový model (vlastní zpracování)	85
Obrázek 23 WhereScape 3D – Zobrazení parametrů vybraného sloupce (vlastní zpracování)	85
Obrázek 24 WhereScape 3D – Mapování zdrojů dimenzionální tabulky (vlastní zpracování)	87
Obrázek 25 WhereScape 3D – Mapování zdrojů EPM OLAP (vlastní zpracování)	87
Obrázek 26 Zobrazení datového toku vybraného atributu v nástroji Collibra (vlastní zpracování).....	88
Obrázek 27 Zobrazení vazeb mezi aplikačními a byznys daty v nástroji Collibra (vlastní zpracování).....	89
Obrázek 28 Prompt pro popis transformační logiky vybraného sloupce (vlastní zpracování)	93
Obrázek 29 Popis transformační logiky vybraného sloupce pomocí ChatGPT (vlastní zpracování).....	93
Obrázek 30 Prompt přípravu definice pojmu Average Claim Amount (vlastní zpracování)	95
Obrázek 31 Obchodní definice pojmu Average Claim Amount pomocí ChatGPT (vlastní zpracování).....	95
Obrázek 32 Uvedené zdroje v odpovědi modelu ChatGPT	96

Seznam tabulek

Tabulka 1 Popis parametrů – datové objekty.....	66
Tabulka 2 Popis parametrů – datové atributy	67
Tabulka 3 Popis parametrů – obchodní data.....	69
Tabulka 4 Ukázka kompletně zdokumentovaného záznamu	80
Tabulka 5 Přehled naplnění dílčích cílů diplomové práce	100

Seznam výpisů programového kódu

Výpis 1 Ukázka sady pracovních tabulek pro kalkulaci KPIs (vlastní zpracování)	76
--	----

Seznam zkratek

AI	Artificial Intelligence
AVG	Average
BNPP	BNP Paribas
CE	Continental Europe
CSV	Comma-Separated Values
DAMA	Data Administration Management Association
DBMS	Database Management System
DG	Data Governance
DM	Data Management
DMBOK	Data Management Body of Knowledge
DWH	Data Warehouse
EA	Enterprise Achitect
EDM	Enterprise Data Model
EPM	Enterprise Performance Management
ETL	Extract, Transform, Load
FIS	Fakulta informatiky a statistiky
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
IFRS	International Financial Reporting Standards
IT	Informační technologie
KPI	Key Performance Indicator
MDA	Model Driven Architecture
MDS	Microsoft Master Data Services
ODS	Operational Data Store
OLAP	Online Analytical Processing

RMS	Referential Management Solution
RO	Rešeršní otázka
SQL	Structured Query Language
SSAS	SQL Server Analysis Services
WS	WhereScape

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 nazývána „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 význam dat 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. S naplněním 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. Klíčové je rovněž zapojení všech zúčastněných stran a jasná, nepřetržitá komunikace zajišťující spolupráci a porozumění napříč týmy. Postupná implementace pomáhá organizacím se efektivně přizpůsobit, zatímco multifunkční týmy s různými dovednostmi a školením budují kapacitu. Udržitelnosti je dosaženo začleněním řízení do standardních procesů s metrikami pro měření efektivit 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, integrace dat a interoperabilita, správa obsahu a dokumentů, správa referenčních a kmenových dat, data warehousing a business intelligence, metadata management, řízení datové kvality a řízení datové architektury (DAMA International 2017). Znalostní oblasti jsou velmi rozsáhlé a definice každé z nich není předmětem této práce. Pozornost bude tedy zaměřena především na oblast datové architektury, datového modelování a metadat ve vztahu k data governance.

V rámci 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 budují společný jazyk, snižují nejednoznačnost a podporují přesnost dat. Zavedením byznys datových slovníků mohou organizace zlepšit srozumitelnost a transparentnost dat, což je nezbytné pro informované rozhodování a dodržování regulatorních požadavků (Kimball a Ross 2013). Samotná oblast metadat však nestačí, protože ta se zabývá především tím, jak je publikovat, spravovat, uchovávat a podobně. Víze, 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 víze (DAMA International 2017).

V této diplomové práci, soustředěné na byznys a datové slovníky, bude věnována pozornost společnosti BNP Paribas Cardif. Jako významná finanční instituce spravuje BNP Paribas Cardif obrovské množství komplexních dat, která musí splňovat přísné regulační standardy (BNP Paribas Cardif 2023). Díky jedinečnému datovému prostředí a požadavkům je společnost vhodná pro zkoumání byznys a datových slovníků, jejich přínosů, problémů v udržitelnosti a překážek v jejich implementaci v rámci celého konceptu data governance.

Cíle práce

Hlavním cílem práce je analyzovat a zmapovat proces implementace byznys a datových slovníků a veškerá jejich úskalí a nabídnout realizovatelná doporučení k řešení lidských, technických i procesních problémů při implementaci data governance. Hlavní cíl lze rozpadnout na několik dílčích cílů:

1. popsat všechny relevantní oblasti data governance, které jsou klíčové pro realizaci hlavního cíle,
2. shrnout současný stav společnosti a identifikovat specifické požadavky na data governance,
3. zdokumentovat proces analýzy, návrhu a implementace byznys a datových slovníků,
4. na základě poznatků získaných z dostupné literatury a během procesu implementace identifikovat problémy spojené s byznys a datovými slovníky a nabídnout realizovatelná doporučení.

Použité metody

Pro dosažení stanovených cílů byla využita kombinace analýzy a syntézy odborné literatury, kvalitativního výzkumu a případové studie.

Teoretická část práce byla zaměřena na zodpovězení výzkumných otázek, které pokrývají současný stav poznání v oblasti správy dat, představení znalostních oblastí dle rámce DAMA-DMBOK2, klíčové komponenty data governance a roli byznys a datových slovníků

včetně známých bariér při jejich implementaci. K jejich zodpovězení byla použita metoda literární rešerše, komparace zdrojů a následná syntéza poznatků s cílem vytvořit pevný teoretický rámec pro navazující praktickou část.

Praktickou část lze, v širším pojetí, považovat za případovou studii realizovanou ve společnosti BNP Paribas Cardif, v rámci které autor aktivně participoval na zvyšování úrovně datové dokumentace. Byly použity metody strukturovaného pozorování, analýzy interních dokumentů nebo také semistrukturované rozhovory a konzultace s odborníky ve společnosti.

V návrhové a implementační fázi byl aplikován iterativní přístup, který umožnil flexibilní validaci jednotlivých kroků s byznys uživateli a technickými experty. Místy byly využity i principy design thinkingu, zejména pro hledání přístupů, jak lépe zpřístupnit složitá technická metadata obchodnímu publiku. Závěrečná doporučení vznikla syntézou praktických zkušeností, ověřených postupů a zpětné vazby z prostředí firmy.

Předpoklady a omezení

Jedním z hlavních předpokladů pro dosažení definovaných cílů práce je především zájem a ochota společnosti podílet se na řešení a sdílet informace. To zahrnuje přístup k interní dokumentaci pro přesnou analýzu, spolupráci zaměstnanců BNP Paribas Cardif napříč odděleními s cílem získat vhled do technických, procesních i lidských aspektů data governance a porozumění specifickým regulatorním požadavkům spojeným s odvětvím pojišťovnictví.

Mezi omezení lze zařadit především přístup k určitým typům dat, vzhledem k povaze a velikosti společnosti. Nemožnost poskytnout určitá data může omezit hloubku analýzy v určitých oblastech. Přizpůsobení se potřebám a procesům společnosti BNPP Cardif může přinést omezení, pokud jde o škálovatelnost nebo přenositelnost na jiné organizace s odlišnými strukturami správy dat.

Důvod výběru tématu

K výběru tématu a zkoumané společnosti pro diplomovou práci došlo během praxe absolvované v rámci oboru Data a analytika pro business. V rámci tohoto působení proběhla práce na skutečném implementačním projektu ve společnosti BNP Paribas Cardif s cílem zvýšení datové maturity v dílčích datamartech, respektive nalezení vazeb mezi aplikačními daty a byznys termíny a zvýšení kvality dokumentace. Činnosti popsané v této diplomové práci vychází ze zkušeností a aktivit, které byly provedeny během praxe ve společnosti. Při vypracování práce byly rovněž aplikovány znalosti získané v rámci pracovního působení v tomto oboru.

Struktura práce

První kapitola se zabývá hledáním odpovědí v dostupné literatuře na otázky týkající se data governance, byznys a datových slovníků a výzev souvisejících s jejich implementací. Cílem této kapitoly je vytvoření východiska pro další části práce.

Nejprve je prozkoumán aktuální stav poznání v této oblasti, poté jsou nalezeny a vymezeny relevantní znalostní oblasti. Následně jsou popsány klíčové komponenty v rámci správy dat. Poslední část rešerše se zabývá přímo byznys a datovými slovníky a výzvami v jejich implementaci.

Druhá kapitola popisuje pozici diplomové práce vzhledem k řešení data governance ve společnosti BNP Paribas Cardif a specifické požadavky na správu dat v dané společnosti.

Ve třetí kapitole je popsáno řešení, tzn. implementace byznys a datových slovníků, prezentovány výstupy této činnosti a doporučení pro další rozvoj.

V závěru práce jsou shrnuty hlavní poznatky a výstupy práce.

Očekávané přínosy

Na přínosy této diplomové práce lze nahlížet ve více rovinách.

V první řadě má práce přínos pro samotného autora, který během řešení získal hlubší porozumění oblasti data governance a praktické zkušenosti s implementací datové dokumentace ve firemním prostředí. Tyto zkušenosti jsou dále uplatnitelné v profesní dráze a přispívají k rozvoji odborných kompetencí.

V širším smyslu však práce přináší přidanou hodnotu i pro odbornou komunitu zabývající se správou dat. Na základě rešerše dostupné literatury práce poskytuje přehled klíčových znalostních oblastí, které jsou relevantní pro úspěšnou implementaci byznys a datových slovníků v rámci data governance. Tento teoretický základ je navíc propojen s podrobným popisem skutečného implementačního projektu ve společnosti BNP Paribas Cardif.

Důležitou složkou práce je, že vychází z reálného projektu realizovaného během působení ve společnosti BNP Paribas Cardif v rámci něhož došlo k návrhu a implementace zcela nové dokumentace technických a obchodních dat, včetně vytvoření datového modelu, definic obchodních pojmů, mapování mezi aplikačními a byznys daty apod. Výstupem je funkční datový a byznys slovník, který přispívá ke zvýšení datové kvality a posiluje celkovou úroveň správy dat v organizaci. Tento praktický přínos je doplněn o analytické zhodnocení a doporučení pro další rozvoj.

Díky této kombinaci teorie a praxe vzniká přehledný a prakticky využitelný dokument, který může sloužit jako podpora i pro jiné organizace či jednotlivce čelící podobné iniciativě. Práce ilustruje, jak celý proces implementace může probíhat, jaká specifika se mohou objevit v různých fázích, a jak s těmito výzvami efektivně pracovat.

Zvláštní důraz je kladen na lidské a organizační faktory, které jsou v praxi často opomíjeny na úkor technologických aspektů, přestože představují naprosto klíčové riziko i příležitost pro úspěšné zavedení efektivní správy dat.

Významným přínosem je také to, že veškeré aktivity popsané v této práci vycházejí z reálného prostředí a vedly ke konkrétním výstupům – od analýzy dokumentace, přes návrh struktury dokumentace a tvorbu definic a mapování, až po nasazení do prostředí

společnosti. Diplomová práce tak nepřináší pouze teoretické úvahy, ale také konkrétní doporučení podložené praktickou zkušeností.

Celkově tak práce nabízí ucelený pohled na proces zavádění správy metadat – od základních znalostí až po konkrétní implementaci – a tím přispívá k lepšímu pochopení i efektivnějšímu rozvoji data governance v praxi.

1 Rešerše literatury

Tato kapitola se zaměřuje na systematický přehled literatury, dle doporučení (Okoli a Schabram 2010). Struktura této kapitoly je dána rešeršními (výzkumnými) otázkami, které hledají odpověď v dostupné literatuře. Přehled z rešerše pomáhá identifikovat trendy v nalezených zdrojích a vytvořit teoretický podklad pro naplnění dílčích cílů diplomové práce.

Na základě rozboru cíle práce byly identifikovány tyto rešeršní otázky:

- RO1: Jaký je současný stav poznání zvoleného tématu? (1.1)
- RO2: Jaké znalostní oblasti v rámci řízení dat definuje DAMA-DMBOK2? (1.2)
- RO3: Jaké jsou hlavní komponenty data governance? (1.3)
- RO4: Jaký je význam byznys a datových slovníků a jaké jsou typické výzvy při jejich implementaci v kontextu DG? (1.4)

Výzkumné otázky mohou být dále strukturovány do podkapitol.

1.1 RO1: Jaký je současný stav poznání zvoleného tématu?

Před zodpovězením dalších výzkumných otázek je třeba položit si výchozí otázku, týkající se aktuálního stavu poznání dané problematiky v době řešení diplomové práce. V této kapitole tak budou prezentovány hlavní zdroje, které byly nalezeny při prvotní rešerši literatury. Rešerši lze dále rozdělit do tří kategorií: rešerše akademických publikací, knižních publikací a vědeckých článků a rešerše internetových zdrojů. Zdroje, které jsou uváděny byly použity v rámci přípravy na tvorbu práce anebo v práci samotné.

1.1.1 Rešerše vysokoškolských kvalifikačních prací

Pro vyhledání vysokoškolských kvalifikačních (diplomových) prací byl použit systém VŠKP na Vysoké škole ekonomické v Praze (VŠE [b.r.]), rešerše je tak zúžena na práce vytvořené přímo na půdě vysoké školy ekonomické. Tato činnost byla provedena jako první, aby bylo zjištěno, zda je vybrané téma relevantní. Zároveň jsou tyto práce zdrojem užitečné literatury, která byla dále využita.

Na užší téma „Business a datové slovníky v rámci implementace data governance“ nebyly nalezeny žádné podobné diplomové práce. Byla nalezena jedna bakalářská práce s tématem implementace business slovníku ve finanční instituci, kde se autorka zabývá návrhem a implementací business slovníku ve vybrané společnosti (Plotnikova 2021).

Naopak na obecné téma konceptu „Data Governance“ bylo nalezeno velké množství prací. Mezi starší práce lze zařadit například „Zavádění projektu data governance“ z roku 2014, kde autor popisuje nutné artefakty, které musí vzniknout v rámci projektu zavádění správy

dat (Zosinčuk 2014) nebo „Postup zavádění Data Governance“ z roku 2008, kde se autorka věnuje podobnému tématu (Slouková 2008). Z aktuálnějších prací byla nalezena například diplomová práce „Data governance v podnikové praxi“, kde autor (Frýba 2020) provedl analýzu trhu data governance nástrojů a nasazení vybraných nástrojů v konkrétním podniku. Za zmínku rovněž stojí práce autora (Václavík 2024), který provedl analýzu přínosu nasazení nové data governance v rámci projektu ve vybrané společnosti. Poslední relevantní diplomovou prací, která byla nalezena při rešerši je práce s názvem „Data governance assessment ve vybrané společnosti“ od autorky (Hozáková 2024), která na rozdíl od ostatních děl provádí analýzu a hodnocení již existujícího data governance a dává si za cíl vytvořit ucelený rámec pro hodnocení a zlepšení data governance programu v různých typech společností.

Na základě rešerše se ukázalo, že většina nalezených kvalifikačních prací se zabývá spíše počátečním zaváděním data governance ve společnosti, předpoklady a výběrem vhodných nástrojů. Tato diplomová práce se rovněž zabývá implementací vybraných konceptů a nástrojů správy dat, nikoliv však na samotném začátku, nýbrž několik let po zavedení DG ve společnosti. Zároveň klade větší důraz na lidskou složku, která je v tomto procesu velmi klíčová.

1.1.2 Rešerše knižních publikací a vědeckých článků

V této kapitole bude prezentováno několik klíčových zdrojů, které pomohly v rozšíření znalostí v této oblasti. K vyhledání těchto zdrojů byl z velké části použit webový portál knihovny vysoké školy ekonomické (VŠE [b.r.]), zejména přístup k citačním rejstříkům Scopus a Web of Science (WOS) a poté přístup k elektronickým knihám. Kromě toho byl také použit webový vyhledávací engine Google Scholar (Google [b.r.]).

Jak již bylo zmíněno v předchozí sekci, na téma data governance lze nalézt spoustu publikací. Jedním z hlavních podkladů pro tuto práci se stala kniha DAMA-DMBOK2 od společnosti (DAMA International 2017), která poskytuje komplexní pohled na osvědčené postupy, zásady a pokyny pro efektivní řízení dat. Kromě toho jako významný zdroj pro studium oblasti správy dat posloužily dvě knihy s podobným názvem a to „A Practitioner's Guide to Data Governance - A Case-Based Approach“ od (Gupta a Cannon 2020) a „Practitioner's Guide to Operationalizing Data Governance“ od (Hopper 2023). Obě knihy působí jako přehled doporučení a postupů od odborníků z praxe. Autoři se touto formou chtějí podělit o své zkušenosti a pomoci společností, na cestě k úspěšnému data governance programu. Za zmínku stojí rovněž kniha „Data Governance The Definitive Guide“ od skupiny autorů (Eryurek et al. 2021), kteří se snaží nabídnout celkový pohled na implementaci data governance ve společnostech, protože si uvědomují nutnou potřebu mít přehled nad svými daty související s jejich rostoucím objemem a množstvím regulatorních požadavků.

Při rešerši vědeckých článků byly nalezeno více zdrojů zabývajících se výzvami spojenými se správou dat. Z důvodu omezeného přístupu k těmto zdrojům nebylo možné zahrnout všechny. Prvním článkem, který byl nalezen je konferenční příspěvek z mezinárodní konference „International Conference on Knowledge Management and Information Systems“, který se zabývá identifikací výzev pomocí rešerše literatury (Bassi a Alves-Souza

2023). Dalším relevantním zdrojem je článek od autora (Pansara 2023) s názvem „Unraveling the Complexities of Data Governance with Strategies, Challenges, and Future Directions“, kde se autor zabývá různorodým prostředím data governance, zkoumá strategie pro vytvoření účinných rámců správy dat a řeší výzvy spojené s touto oblastí. Poslední článek, který stojí za zmínku nese titul „Data Governance as a Collective Action Problem“ a zkoumá, proč je správa dat obtížná ve státních organizacích na základě rešerše literatury a empirické případové studie (Benfeldt et al. 2020).

1.1.3 Rešerše internetových zdrojů

Poslední kategorií rešerše je prohledávání webových stránek. V tomto ohledu bylo čerpáno například z blogu dedikovaného data governance od společnosti HubSpot (Amaresan 2019). Dalším zdrojem informací byla webová stránka společnosti Atlan, která se zabývá datovou demokratizací a na svém webu publikuje články o data governance, byznys slovnících, datech atp. (Atlan 2013). Kromě toho byla použita celá řada webů, které se přímo zabývají problematikou data governance a data managementu ať už v rámci propagace vlastního řešení nebo ve formě blogových a wiki článků. V případě že byly součástí článků originální zdroje, v některých případech byly prostudovány a použity namísto původně nalezeného příspěvku.

1.2 RO2: Jaké znalostní oblasti v rámci řízení dat definuje 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 této práce budou použity 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.

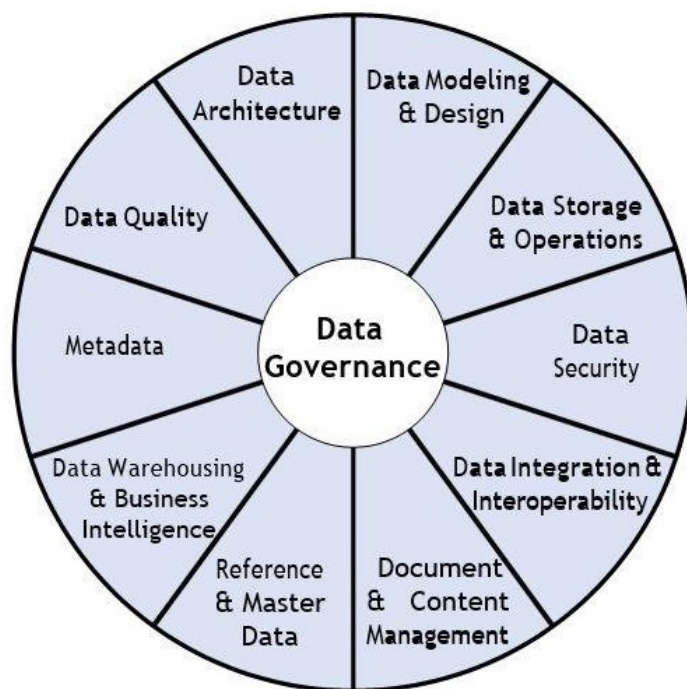
V úvodu práce byla zmíněna správa dat, jako aktiva, tzv. „data as an asset“. Podle (O’Sullivan a Sheffrin 2007), aktivem je jakýkoliv hmotný či nehmotný zdroj, který podnik nebo ekonomický subjekt vlastní nebo ovládá a který může být použit pro vytvoření kladné ekonomické hodnoty. Data lze považovat 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).

1.2.1 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, regulatorní 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.

Aplikační rámec, který podrobněji popisuje znalostní oblasti, které tvoří celkový rozsah DM je například DAMA-DMBOK framework. Jednou z možných vizualizací, které tento framework znázorňují je tzv. DAMA Wheel (viz Obrázek 1).



Obrázek 1 DAMA-DMBOK2 Wheel (DAMA International 2017)

DAMA Wheel umísťuje DG do centra data management činností, pretože je nutná pro zajištění konzistence všech dílčích funkcí a rovnováhy mezi nimi (DAMA International 2017). Ostatní znalostní oblasti (nebo také „funkce“) jsou umístěny okolo.

Kromě data governance 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 (DAMA International 2017). Jak je z výčtu patrné, existuje mnoho znalostních oblastí a vzhledem k jejich širokému rozsahu se nelze věnovat každé z nich.

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 metadata oblasti je třeba věnovat zvláštní pozornost obchodním metadatům, která poskytují datům kontext a význam, což usnadňuje jejich lepší pochopení a využití v rámci organizace.

V této práci 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 diplomové práce relevantní. Koncept data governance bude dále rozebrán v další rešeršní otázce (viz 1.3).

1.2.2 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 překleneje 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).

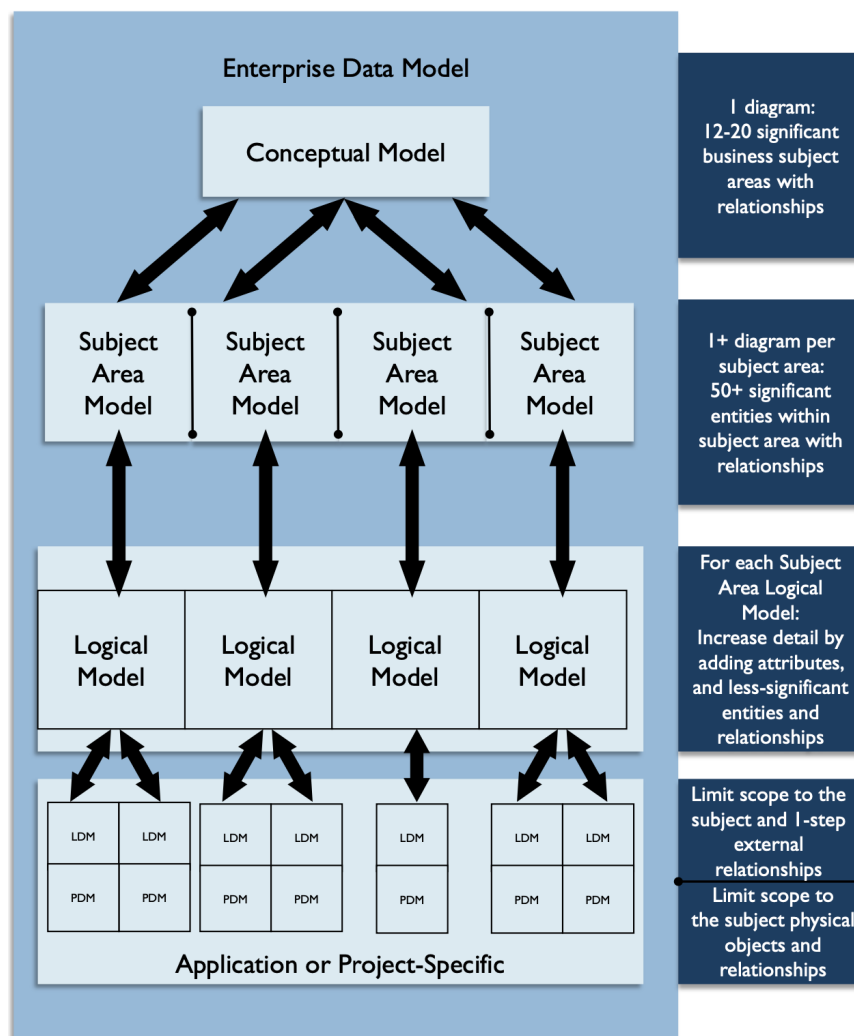
Jak data proudí v organizaci prostřednictvím kanálů nebo rozhraní, jsou zabezpečena, integrována, ukládána, zaznamenávána, katalogizována, sdílena, vykazována, analyzována a dodávána zúčastněným stranám. Po cestě mohou být data ověřována, vylepšována, propojována, certifikována, agregována, anonymizována 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).

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

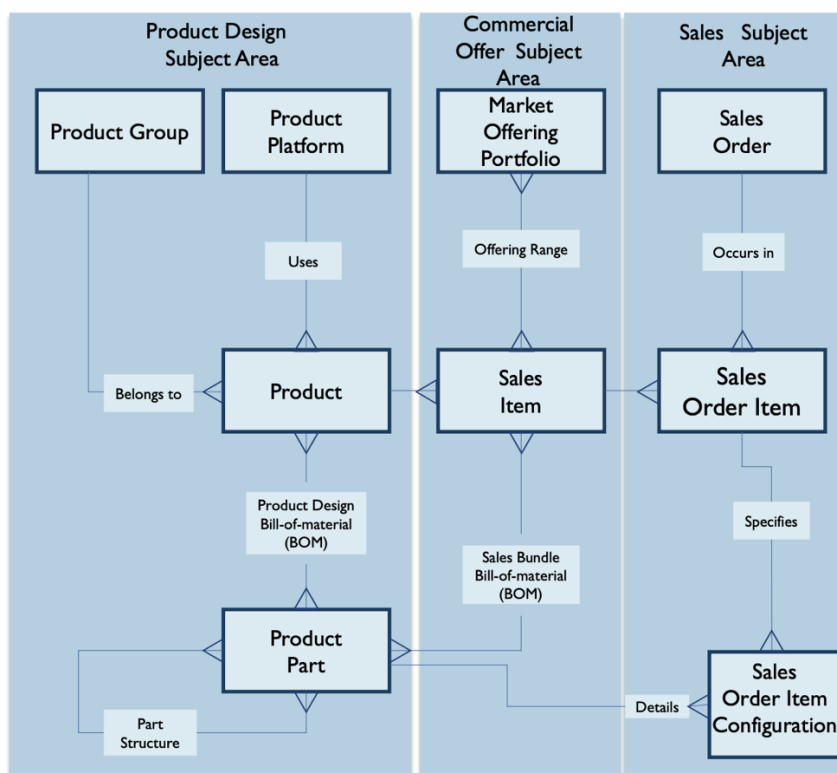
definice dat a metadat, klíčové datové entity podniku (tj. obchodní koncepty), jejich vztahy, kritická řídicí obchodní pravidla a některé kritické atributy. Stanovuje základ pro všechny projekty týkající se dat. Každý datový model na úrovni projektu musí vycházet z podnikového datového modelu. EDM pomáhá 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 aplikačními fyzickými datovými modely je k vidění na obrázku 2 (DAMA International 2017).



Obrázek 2 Enterprise Data Model (DAMA International 2017)

Horizontální vztahy mezi modely zobrazuje obrázek 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).

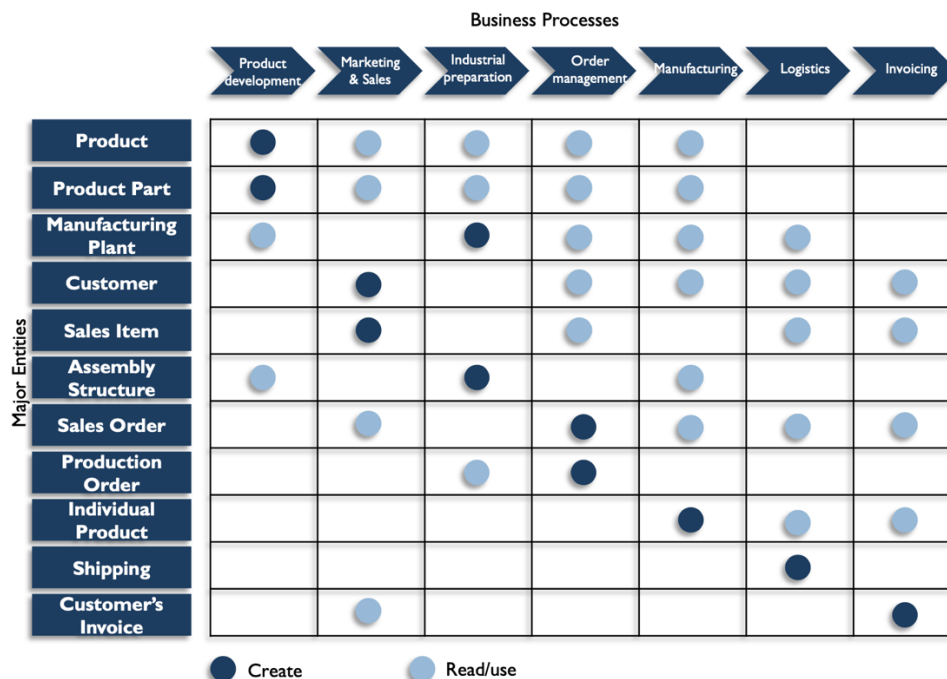


Obrázek 3 Horizontální závislost mezi datovými modely (DAMA International 2017)

Návrh datových toků

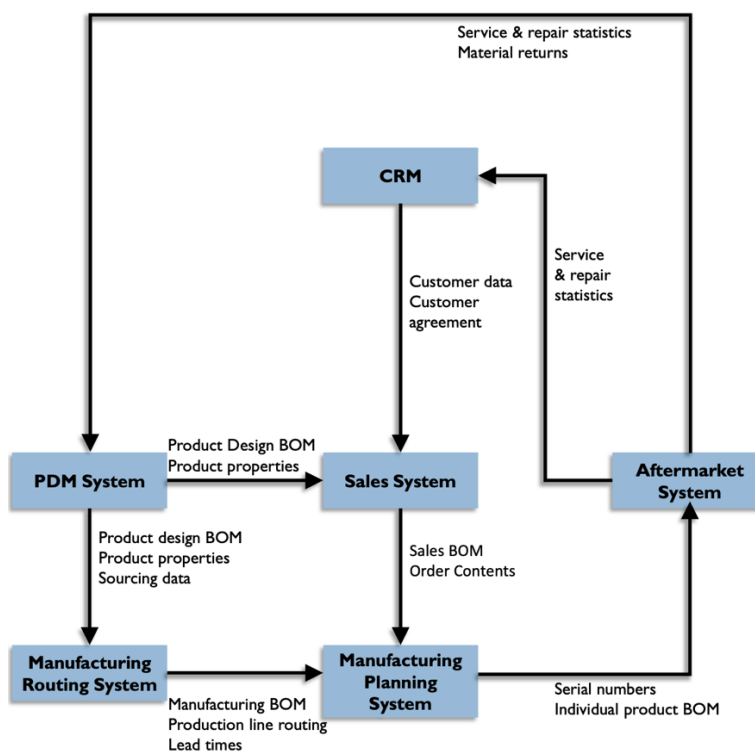
Návrh datových toků definuje požadavky a hlavní plán pro ukládání a zpracování dat v databázích, aplikacích, platformách a sítích. Tyto datové toky mapují pohyb dat k podnikovým procesům, místům, podnikovým rolím a k technickým komponentám – mezi různými systémy, aplikacemi atp. Tento proces zajišťuje, 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, že 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 (viz Obrázek 4) nebo diagramem datových toků (viz Obrázek 5).



Obrázek 4 Matice datových toků (DAMA International 2017)

Výhodou matice je, že poskytuje jasný přehled o tom, jaká data jsou procesy vytvářeny a používány. 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 závislostí dat mezi nimi, což následně zlepšuje dokumentaci procesů (DAMA International 2017).



Obrázek 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:

- 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 organizace. S růstem společnosti 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).

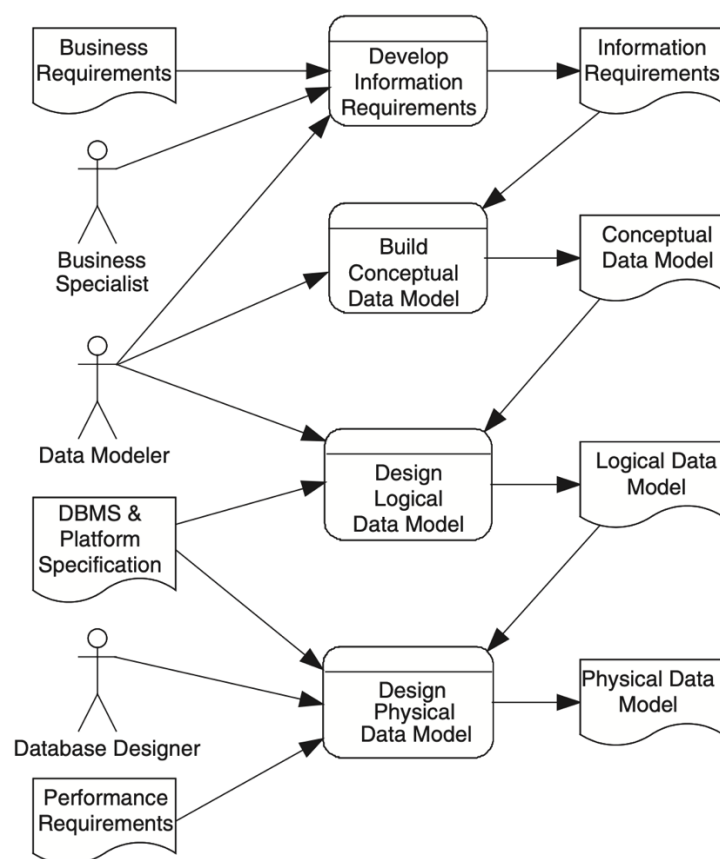
1.2.3 Datové modelování a design

V předchozí části byla představena znalostní oblast datové architektury, jejíž jednou 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 typicky existují na třech úrovních abstrakce – konceptuální, logické a fyzické. Každý model obsahuje sadu komponent – entity, vztahy, fakta, klíče, atributy atd. (DAMA International 2017).

Na obrázku 6 je zobrazen přehled úkolů a výsledků návrhu databáze od obchodních požadavků až po kompletní specifikaci databáze.



Obrázek 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 struktury, 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. 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 udržují porozumění datovým strukturám organizace. 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. Datové modely jsou rovněž zdrojem metadat pro datové slovníky (DAMA International 2017).

1.2.4 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 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 metadatech stejně jako techničtí uživatelé v obchodních metadatech (Hopper 2023).

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 této diplomové práce. 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 i součástí 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.

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, jejich vlastnosti, fyzické datové modely, přístupová oprávnění, pravidla CRUD a zdokumentované vztahy mezi modely a aktivy. Zahrnuje 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).

Provozní metadata

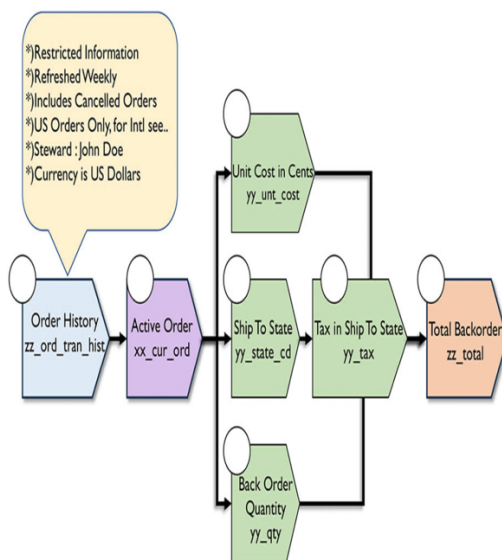
Provozní metadata popisují charakteristiky rutinních operací s daty a souvisejícími statistikami. Patří sem scheduling, popisy datových úloh, statusy přesunu a transformace dat, statistiky read/update výkonu, statistiky objemu dat, statistiky dostupnosti a informace o zálohování a archivaci. Denní informace 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).

Data lineage

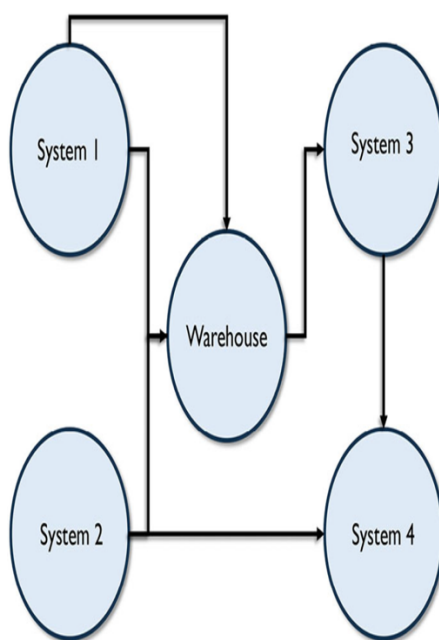
Významným přínosem zjišťování a dokumentování metadat o fyzických aktivech je poskytování informací o tom, jak se data při pohybu mezi systémy transformují – tzv. data lineage (někdy také mapování datových toků). Výsledkem je ucelená vizualizace dat při jejich pohybu z původních míst (oficiální zdroj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

Lineage lze dokumentovat na různých úrovních. Na obrázku 7 je ukázka datového toku na úrovni datového prvku, která popisuje, co se děje s vybraným sloupcem v databázi (DAMA International 2017).



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

Vyšší úrovně lineage (např. systémová lineage) shrnují pohyb na úrovni systému nebo aplikace. Obrázek 8 ukazuje systémový tok, kde je vizualizován obecný pohyb dat na úrovni systému nebo aplikace (DAMA International 2017).



Obrázek 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 toho, jaká data organizace má, jde často spolu s řízením datové kvality o výchozí body při definování data governance ve společnosti (Hopper 2023).

1.3 RO3: Jaké jsou hlavní komponenty 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, jakou 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 a 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.]).

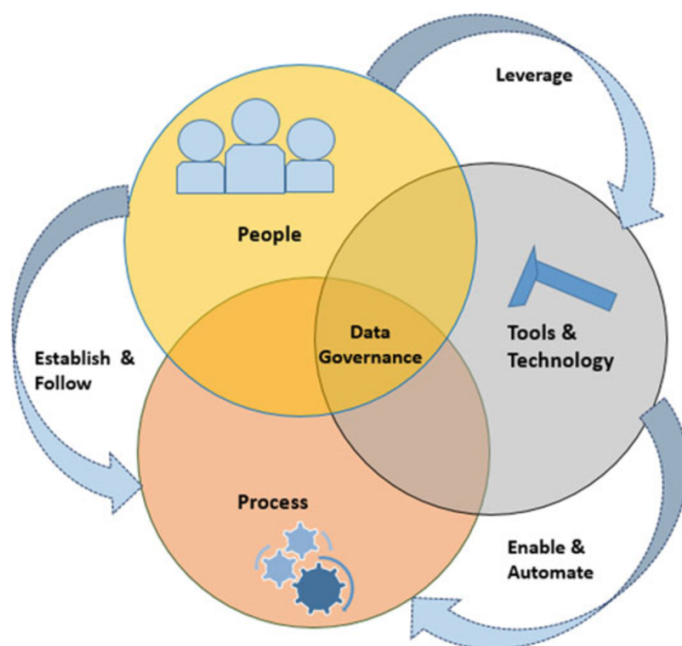
Komponenty správy dat

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 (viz Obrázek 9).

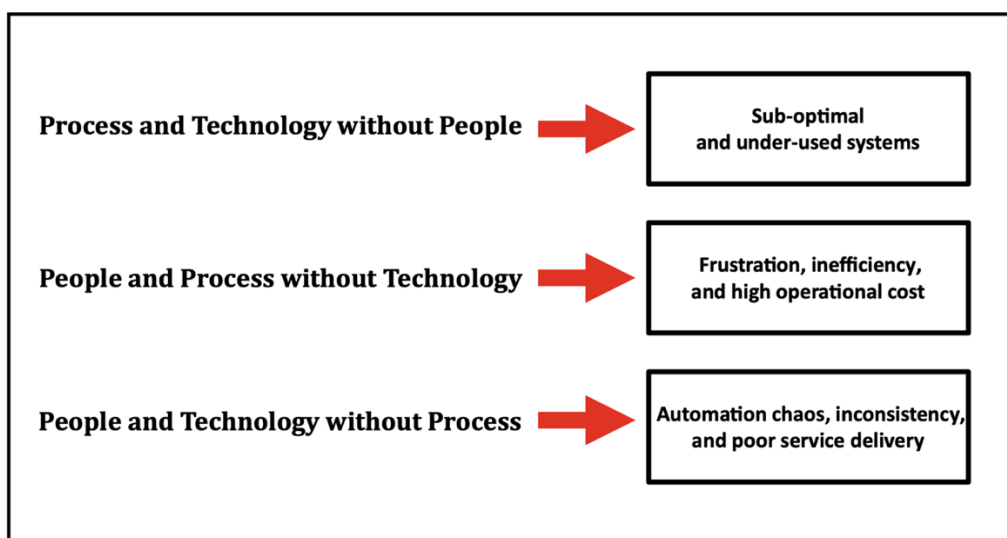


Obrázek 9 Komponenty správy dat a jejich vzájemné vztahy (Mahanti 2021)

Aby bylo možné čerpat benefitů správy dat, musí všechny komponenty vzájemně spolupracovat. Pokud pouze jedna z komponent chybí, může to znamenat neúspěch při implementaci DG programu (Mahanti 2021).

- *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 (Mahanti 2021).
- *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 (Mahanti 2021).
- *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 zmateků, špatné poskytování služeb (například dlouhé dodací lhůty a zpoždění) a zvýšené náklady (Mahanti 2021).

Výše uvedené skutečnosti jsou shrnuty v obrázku níže (viz Obrázek 10).



Obrázek 10 Dopady chybějících Data Governance komponent (Mahanti 2021)

Vzájemná spolupráce všech komponent je tedy skutečně velmi důležitá. Zároveň však lze říci, že některé komponenty jsou významnější než jiné. Tato skutečnost vyplývá i z popisu dopadů chybějících komponent, z něhož je patrné, že chybějící lidská či procesní komponenta znamená 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 zjednodušují jejich úkoly a umožňují lepší spolupráci (Mahanti 2021).

Důraz na lidskou a procesní komponentu je obzvlášť důležitý především protože spousta společností, i v současné době, stráví 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).

1.3.1 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 zodpovědností, a ne pouze záležitostí vedoucího managementu je jedním z nejtěžších úkolů, kterému 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 obchodní 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 zavádí proto, že organizace něco nedělá správně a proto je třeba to změnit. Je nezbytné se orientovat, školit, vzdělávat, komunikovat, vést, povzbuzovat.

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 přiřazení odpovědnosti za výsledky související s daty, což často zahrnuje zavedení rolí, jako jsou správci a vedoucí správy dat. 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).

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 jejichž 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 směrnici 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 udržitelné.

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 obzvláště 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 (viz Kultura). 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 využít k posouzení požadavků uživatele a jejich uplatnění v dokumentaci (Indeed Editorial Team 2024a).

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 projevat 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).

1.3.2 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.

Datové principy

Datové principy jsou tzv. high-level výroky o tom, čeho chce organizace dosáhnout zavedením správy dat. Neměly by být zaměřovány 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, doktrína, předpoklad, pravidlo či kodex chování.

Organizační principy jsou jádrem efektivní správy dat, jedná se prohlášení o filozofii, které si lze představit jako listinu práv a svobod – 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:

- 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 jsou k vidění na obrázku 11.

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 11 Obecně přijímané informační principy (Ladley 2012)

Datové politiky

(DAMA International 2017) definuje politiky či zásady 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 jakéhokoliv 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. Konkrétní zásady se mohou soustředit například na přístup, užití, uchovávání dat nebo například datovou kvalitu a integraci (Ladley 2012).

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).

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. Tyto procesy by měly být pravidelně revidovány a aktualizovány, aby odpovídaly organizačním změnám, aktualizacím předpisů a vyvíjejícím se potřebám správy dat. 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).

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).

1.3.3 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 (viz kapitola 1.4).

1.4 RO4: Jaký je význam byznys a datových slovníků a jaké jsou typické výzvy při jejich implementaci v kontextu DG?

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.

1.4.1 Business Glossary

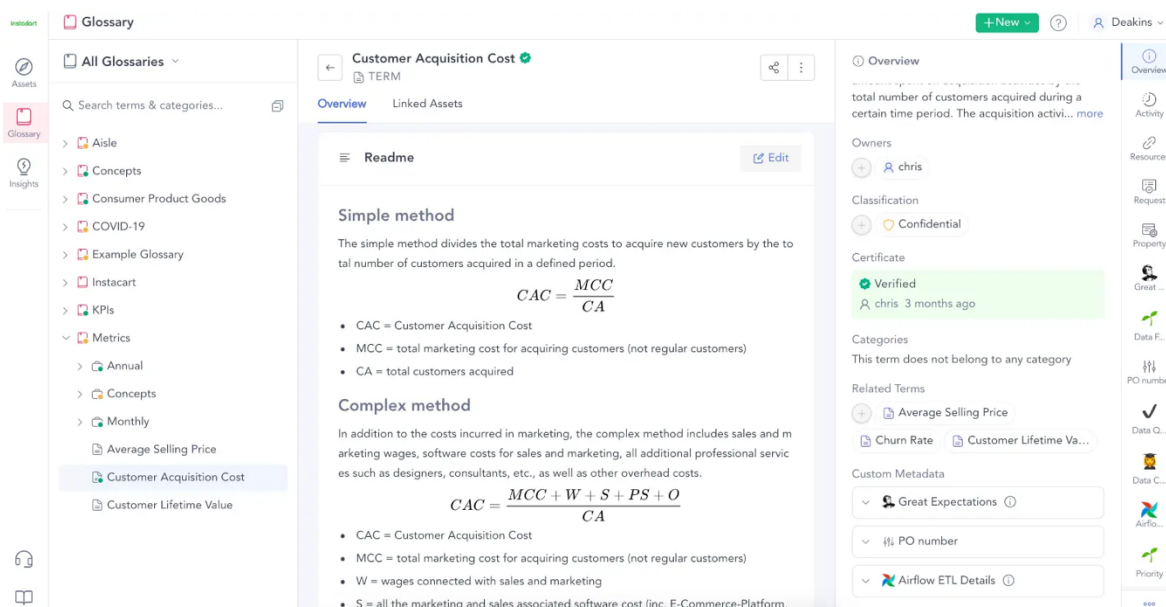
Byznys slovník (také „business glossary“, „obchodní slovník“ nebo „slovník obchodních pojmů“) je zásadním data governance nástrojem, jehož účelem je dokumentovat a uchovávat obchodní pojmy a terminologii organizace, definice a vztahy mezi těmito pojmy – zabývá se tedy obchodními metadaty (viz kapitola 1.2.4). 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. Nejdůležitější vlastností slovníku je však to, aby jeho obsah byl robustní. Dle (DAMA International 2017) by slovníky obchodních pojmů měli obsahovat například 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) odpovědní za jejich správu, techničtí uživatelé. Obchodní 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 a vývoji a k provádění dopadové analýzy (DAMA International 2017).



Obrázek 12 Termín v byznys slovníku (Atlan 2023a)

Byznys slovníky by neměli 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 je k vidění na obrázku 12.

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, který jim umožňuje 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 (viz kapitola 1.2). 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, 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 obsah datového slovníku bude zpravidla odlišný pro každý datový, 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 (DAMA International 2017). Ukázka jednoduchého datového slovníku je k vidění na obrázku 13.

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 13 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. Vyvíjecí 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 (viz kapitola 1.3) 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 (Mahanti 2021). Výzvy lze tedy rovněž volně rozdělit do stejných kategorií.

1.4.3 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 dat, 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.

Znalosti

V rámci lidské složky hrají významnou roli znalosti (viz kapitola 1.3.1). 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).

V kapitole 1.3.1 bylo zdůrazněno, že v případě byznys a datových slovníků jsou klíčové silné dokumentační dovednosti. 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 správný, kromě toho však existuje celá řada dalších prvků, které je nutné zvážit. Podle (Allemann 2023; Walery 2020) by se definice měly řídit předem definovanými pravidly:

- **Publikum** – znalost publika je klíčová, definice musí být přizpůsobené 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 či doménově specifických termínů.
- **Rozsah** – definování pojmů 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.

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 dat je proces, který je potřeba neustále vědomě rozvíjet (DAMA International 2017; Walery 2020).

Lidé 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, náboru a retenci kvalifikovaných odborníků, aby mohly účinně podporovat svá úsilí v oblasti správy dat (Almeida 2024).

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 (Raden 2023b).

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 náležité školení, podporu a zdroje, aby usnadnily přechod a adaptaci (Raden 2023a).

Inkluzivní rozhodovací procesy, v nichž mají jednotlivci 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 vcítěním se do nich mohou organizace a společnosti zavést strategie, které usnadní hladší přechod a umožní lidem přijmout nové technologie a postupy (Raden 2023a).

Kultura

Význam podnikové kultury už byl v této práci 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 jej 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 integrovali činnosti správy, nikoli vnucovat 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).

1.4.4 Organizační a procesní výzvy

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 (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)

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. Vyvážení 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).

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 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 (Gupta a Cannon 2020).

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 vyžaduje souhlas různých klíčových zúčastněných stran. Poté, co jsou procesy 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).

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).

1.4.5 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; Mahanti 2021).

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 (DAMA International 2017; Mahanti 2021; Gupta a Cannon 2020).

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.

1.5 Shrnutí kapitoly rešerše literatury

Rešerše literatury potvrdila, že implementace byznys a datových slovníků v rámci data governance je komplexní výzvou, která vyžaduje znalost celé řady oblastí – od datové architektury a modelování, přes metadata, až po organizační a lidské faktory. Zvláštní pozornost je nutné věnovat zejména propojení technického a obchodního pohledu na data, protože právě zde často vznikají komunikační bariéry a neefektivita.

Z poznatků řešerše vyplývá, že úspěšné zavedení slovníků není možné bez jasně definovaných rolí, procesů pro validaci a údržbu, a aktivní participace uživatelů. Významnou roli hrají také kulturní a vzdělávací aspekty – pouze technické řešení nestačí.

Pro další části práce je tedy důležité:

- Zohlednit provázanost mezi datovou architekturou, modelováním a metadaty.
- Soustředit se nejen na samotné nástroje, ale i na to, jakým způsobem se do správy dat zapojují lidé.
- Zaměřit se na překážky, které literární zdroje často popisují – například odpor ke sdílení informací, složité schvalovací procesy nebo nedostatek motivace – a hledat, zda se objevují i v praxi konkrétní organizace.
- Při praktické analýze si všímat nejen technického stavu, ale i zralosti datové kultury a nastavených pravidel.

Řešerše tak vytváří pevný teoretický základ pro praktickou část práce, a zároveň definuje kritéria, podle kterých bude možné hodnotit úspěšnost realizovaného řešení.

2 Současný stav

Tato kapitola se zaměřuje na definování současného stavu v době řešení diplomové práce. Nejprve dojde ke stručnému představení společnosti, poté bude zhodnocen aktuální stav data governance a uvedeny výzvy či bariéry, které byly na počátku identifikovány a staly se podnětem pro řešení diplomové práce.

2.1 Představení společnosti

BNP Paribas Cardif je mezinárodní pojišťovací divize renomované finanční skupiny BNP Paribas (dále také „BNPP“). Společnost nabízí široké spektrum pojistných produktů, které se zaměřují na ochranu klientů v různých oblastech, přičemž její hlavní zaměření leží na pojištění schopnosti splácet úvěry a další finanční závazky (BNP Paribas Cardif [b.r.]; Brezina 2024).

Obchodní model společnosti je postaven na principu B2B2C – společnost disponuje širokou distribuční sítí tvořenou obchodními partnery včetně světových bank, nebankovních poskytovatelů finančních služeb, maloobchodních služeb a nezávislých finančních zprostředkovatelů. Cardif tak prodává své produkty ostatním společnostem (B2B), a ty je následně prodávají spotřebitelům (B2C). BNPP Cardif v současné době působí ve 33 zemích a za svou dobu působení, od roku 1996, získala společnost více než sto milionů klientů po celém světě (BNP Paribas Cardif 2023).

V české republice existují dvě entity: BNP Paribas Cardif Pojišťovna, která poskytuje zmiňované pojistné produkty, a BNP Paribas Cardif Services, která se specializuje na IT a další sdílené služby pro evropské pojišťovny ze skupiny BNPP (BNP Paribas Cardif 2023; [b.r.]). V rámci IT služeb se Cardif Services zabývá zajištěním provozu, údržby a rozvoje aplikací a IT systémů pro sjednávání pojistných smluv a řešení pojistných událostí. Kromě služeb v oblasti IT poskytuje Cardif Services například služby v oblastech řízení lidských zdrojů, marketingu, účetnictví a pojistné matematiky. Hlavním cílem je poskytovat pojišťovnám kvalitní služby, aby skupina BNP Paribas Cardif mohla maximálně těžit ze sdílených informací a automatizovaných procesů a pracovat tak na zkvalitnění svých služeb, které poskytuje svým klientům (BNP Paribas Cardif [b.r.]).

2.2 Aktuální stav správy dat

Vzhledem k tomu, že BNP Paribas Cardif operuje ve vysoce regulovaném finančním sektoru a spravuje rozsáhlé množství citlivých a komplexních dat, je efektivní správa dat klíčovým faktorem pro zajištění souladu s legislativními požadavky i pro optimalizaci vnitřních procesů. Data jsou nejen základním stavebním prvkem analytických a reportingových procesů, ale také klíčovým aktivem pro rozhodování na všech úrovních organizace.

S účastí v tomto sektoru je spojená nutnost dodržovat řadu regulatorních rámců a standardů zaměřených na správu dat a informační bezpečnost. Mezi nejvýznamnější patří:

- Solvency II – regulace pro pojišťovací sektor vyžadující transparentní řízení rizik, kvalitu dat a robustní reporting (EIOPA 2015).
- GDPR (General Data Protection Regulation) – přísná pravidla pro správu osobních údajů, která ovlivňují procesy uchovávání, přenosu a mazání dat (MVČR [b.r.]).
- IFRS 17 – účetní standardy vyžadující přesnou evidenci a kategorizaci pojistných smluv a finančních údajů (IFRS 2024).

Pro zajištění souladu s těmito regulacemi je nezbytné mít definovanou strategii data governance, která zajišťuje integritu, dostupnost a bezpečnost dat napříč všemi procesy. V době řešení diplomové práce bylo v BNP Paribas Cardif data governance již čtyři roky formálně zavedeno. Společnost definovala metodiky, procesy a používala nástroje pro správu dat. Přesto však výsledky neodpovídaly očekáváním – data governance nepřineslo zamýšlenou hodnotu, nebylo plně implementováno a v praxi nefungovalo efektivně. Společnost přistoupila ke správě dat systematicky a postupovala podle osvědčených rámců. Governance iniciativa zahrnovala:

- Definované procesy a metodiky – existovaly dokumentované postupy, jak spravovat metadata, business a datové slovníky, jak evidovat datové zdroje a jak řídit kvalitu dat.
- Zavedené nástroje – společnost využívala specializované softwarové řešení pro metadata management a datové katalogy (viz kapitola 3.1).
- Školení a komunikace – zaměstnanci byli seznámeni s principy správy dat a s tím, jaké postupy mají dodržovat.

Navzdory těmto krokům nebyl stav data governance programu ve společnosti uspokojivý. Problémům, kterým společnost čelila, byly především nízká kvalita dokumentace a nekonzistence mezi obchodními a technickými daty. Toto východisko představovalo počáteční bod diplomové práce v rámci kterého bylo nutné analyzovat existující problémy, jakožto kontext pro samotné řešení.

Přestože byly byznys a datové slovníky již implementovány, pro práci s nimi byly definovány procesy a postupy a byl vybrán nástroj, ve kterém jsou spravovány, nebyly pravidelně aktualizovány, jejich obsah řádně provázán a integrován a v mnoha případech dokumentace nebyla dostupná nebo nebyla dostatečně kvalitní na to, aby s ní bylo možné pracovat. Tento fakt lze považovat jako důsledek výzev spojených s lidskou a procesní komponentou správy dat (viz kapitoly 1.4.3 a 1.4.4). Na základě analýzy klíčových problémů byly definovány bariéry, které jsou popsány v dalších odstavcích této kapitoly.

I přes zavedené metodiky a nástroje se data governance v praxi neseťkalo s dostatečnou adopcí mezi zaměstnanci. Bylo vnímáno spíše jako administrativní zátěž než jako přidaná hodnota, což vedlo k jeho omezenému využití. Dokumentace dat nebyla prioritizována, protože byla považována za vedlejší úkol – zaměstnanci často neměli čas ani motivaci

detailně popisovat data nebo jim chyběly potřebné znalosti pro vytvoření kvalitní dokumentace (viz kapitoly 1.3.1 a 1.4.3).

Kromě toho hráli roli také organizační a sociologické bariéry. Sdílení informací o datech bylo vnímáno jako ohrožení osobního know-how a ztráta vlivu, což vedlo k neochotě spolupracovat. Neochota spojená se sdílením informací a spoluprací může být zčásti důsledkem přísných bezpečnostních opatření týkajících se práce s daty a přístupu k interním i externím zdrojům. Přísná bezpečnostní opatření mají bezpochyby dopad nejen na ochranu dat, ale také na firemní kulturu, produktivitu a ochotu ke spolupráci. Když se lidé potýkají s omezeními, která brání snadné práci s daty, může to posilovat i některé negativních jevy (Kumar a Ansari 2016; Redman et al. 2025).

Přestože mohou být tyto iniciativy vytvořeny s dobrým záměrem a jasnou potřebou na ochranu citlivých dat a zajištění souladu s regulacemi, mohou vést k omezené flexibilitě zaměstnanců při práci s daty a ke snížení ochoty sdílet znalosti. Nejasnost ohledně toho, co je bezpečné sdílet, vede k přehnané opatrnosti, kvůli které si zaměstnanci raději informace ponechávají pro sebe. Mimo jiné, omezený přístup ke zdrojům zpomaluje dokumentaci a práci s daty. To přispívá k uzavřenější firemní kultuře, kde si jednotlivé týmy chrání své know-how místo spolupráce na sdílené datové platformě. Výsledkem je pomalá adopce data governance, protože zaměstnanci nemají motivaci se aktivně zapojit do správy metadat a obchodních i datových slovníků a tvoří se tak prostředí s nízkou mírou spolupráce a sdílení dat, což jsou přitom základní pilíře úspěšné správy dat (Redman et al. 2025; Raden 2023b).

Z hlediska data governance je to paradoxní situace. Organizace chce mít kvalitní, dobře dokumentovaná a sdílená data, ale zároveň nastavuje tak přísné bariéry, že to samo o sobě brzdí iniciativy pro lepší řízení dat. Navíc lidé nemusí rozumět rozdílu mezi bezpečným a efektivním sdílením dat versus neoprávněným sdílením citlivých údajů, což může vést ke zbytečnému utajování běžných informací, které by měly být dostupné v rámci firmy.

Některé výše uvedené problémy byly definovány na základě analýzy a diskuze se zaměstnanci společnosti na počátku projektu, jiné byly odkryty během fáze posouzení aktuálního stavu dokumentace (viz kapitola 3.2). Aby byly odstraněny problémy spojené s nízkou kvalitou dokumentace a nesouladem mezi byznys a technickými daty, byl jsem pověřen analýzou a nápravou těchto nedostatků. Mým cílem je nejen opravit existující chyby a nesrovnalosti, ale také ukázat, že vhodně provedená správa metadat je reálně uskutečnitelná a přináší praktické benefity. Cílem těchto činností je zlepšit kvalitu dokumentace a umožnit zaměstnancům lépe rozumět datům s kterými pracují a využívat je efektivněji. Dosažením tohoto cíle by mělo dojít ke zvýšení důvěry v data a ukázat, že systematická správa metadat může být užitečná a vytvořit tak předpoklady pro širší adopci data governance napříč organizací. Detailnímu popisu těchto aktivit a jejich výsledků se věnuje další kapitola – viz kapitola 3.

2.3 Shrnutí kapitoly současný stav

Analýza současného stavu ve společnosti BNP Paribas Cardif ukázala, že přestože je data governance formálně zakotvena a nástroje pro správu metadat jsou zavedené, jejich

praktické využití je omezené. Klíčové slabiny byly identifikovány především v oblasti srozumitelnosti dokumentace a nedostatečné integrace mezi technickými a obchodními daty.

Z pohledu uživatelů je problematická také omezená dostupnost informací a nejasné odpovědnosti za obsah dokumentace. Bez dostatečného porozumění a motivace jednotlivých týmů dochází k degradaci kvality metadat a jejich využitelnosti.

Tyto zjištění potvrzují, že implementace byznys a datových slovníků nepředstavuje pouze technický úkol, ale vyžaduje důslednou koordinaci napříč celou organizací, včetně jasně definovaných rolí, komunikačních kanálů a schvalovacích procesů.

Pro návrh řešení je proto nutné:

- Zajistit, aby nově vytvořená dokumentace propojovala technický a obchodní pohled na data.
- Vytvořit strukturu dokumentace, která bude snadno udržitelná a srozumitelná pro všechny uživatelské skupiny.
- Identifikovat příležitosti, jak navržený přístup škálovat i mimo pilotní oblast (např. další datamarty nebo regiony).
- Zohlednit existující organizační omezení a připravit řešení, které bude akceptovatelné v daném firemním kontextu.

Zjištěné nedostatky tak definují prostor, ve kterém se návrh a realizace praktického řešení bude pohybovat.

3 Implementace byznys a datových slovníků v BNP Paribas Cardif

Předchozí kapitola ukázala, že i přes formální zavedení data governance ve společnosti, přetrvávaly jisté lidské a procesní překážky, jejichž důsledkem byla především nízká kvalita dokumentace a nekonzistence mezi obchodními a technickými daty a s tím spojená nízká úroveň obchodních a datových slovníků. Přestože existovaly metodiky a technologie pro správu metadat, v praxi nebyly systematicky využívány.

Cílem této kapitoly je podrobný popis konkrétních kroků, které byly podniknuty k nápravě těchto problémů, se zaměřením na zlepšení práce s byznys a datovými slovníky a jejich kvality. V rámci řešení byla provedena řada kroků zaměřených na revizi aktuální dokumentace, odstranění nesrovnalostí v metadatech, propojení byznys termínů s technickými daty, revizi a úpravu současného datového modelu pro zajištění souladu s high-level designem a příprava komplexního dokumentu na základě kterého byly provedeny změny v aplikačních a obchodních datech.

Jednotlivé kroky řešení jsou popsány v následujících částech této kapitoly. Provedené aktivity lze logicky rozdělit do těchto celků:

1. Analýza stávajícího stavu dokumentace.
2. Návrh dokumentace.
3. Implementace návrhu.
4. Zhodnocení přínosů a doporučení pro další rozvoj.

V teoretické části bylo několikrát avizováno, že data governance je dlouhodobá iniciativa, která vyžaduje iterativní přístup. Snahy o její implementaci by neměly být zaměřeny na okamžité vyřešení všech problémů najednou, ale na postupné kroky vedoucí k měřitelnému zlepšení kvality dat a jejich využitelnosti v organizaci.

Cardif je globální korporací působící ve třiceti třech zemích, což s sebou přináší velmi složité datové prostředí, různé regulační požadavky a technologickou heterogenitu napříč regiony. V tak rozsáhlém ekosystému je klíčové hledat příklady, kde data governance může přinést okamžitý přínos, a na nich ukázat reálnou hodnotu těchto aktivit. Ve společnosti se prosazuje princip „jedné verze pravdy“ napříč všemi zeměmi, což znamená, že se udržuje jednotný model datového řízení, který je následně propagován do dalších zemí, přičemž specifické požadavky jednotlivých zemí se řeší operativně.

I když je model sdílený pro všechny země, práce byla zaměřena primárně na region kontinentální Evropy (CE; Česko, Slovensko, Německo, Bulharsko, Maďarsko, Rakousko), který byl využit jako hlavní referenční model. Tento přístup umožnil vyhnout se komplikacím spojeným se specifickými požadavky ostatních zemí a soustředit se na vytvoření konzistentního základu, který bude možné následně škálovat a rozšířit do dalších regionů. I v rámci tohoto regionu byla práce soustředěna pouze na vybrané datové prvky,

kteře budou dále specifikovány v následujících kapitolách, včetně popisu používaných technologií.

Tato strategická volba umožnila zaměřit se na konkrétní datové procesy, odstranit klíčové bariéry v práci s metadaty a ukázat, jak efektivní správa byznys a datových slovníků může přispět ke zlepšení celkové kvality dat v organizaci s předpokladem, že úspěšné provedení těchto aktivit může umožnit rozšířit iniciativu i do dalších datových prvků a později dalších regionů.

3.1 Používané nástroje a technologie

Tato kapitola se věnuje popisu specifických nástrojů a technologií, které jsou ve společnosti používány pro efektivní správu a monitorování datových aktiv, a které byly použity v rámci řešení. V dalších částech této práce budou technologie opakovaně zmiňovány, a proto je vhodné je stručně představit.

3.1.1 Modelování a datová báze

Vývoj probíhá formou MDA (model driven architecture), takže každý objekt musí být zdokumentován. Veškerá dokumentace a modelování podnikové datové architektury probíhá pomocí standardizovaného modelovacího jazyka ArchiMate v nástroji Sparx Enterprise Architect (EA).

Fyzické datové modelování probíhá v prostředí nástroje WhereScape 3D (WS 3D), do kterého jsou přenesena metadata z EA. Ve WhereScape 3D je připraven datový model pro každou úroveň datového skladu. V rámci designu lze rozlišovat mezi master modelem a working modely. Working model je souborem workových tabulek, které se váží ke specifické user story, a odpovídá určité části řešení, vyvíjené různými členy týmu. Jednotlivé kusy řešení se pak sloučí a vznikne master. Kromě toho existuje také generační model, který slouží pro přípravu na export z WhereScape 3D do WhereScape Red.

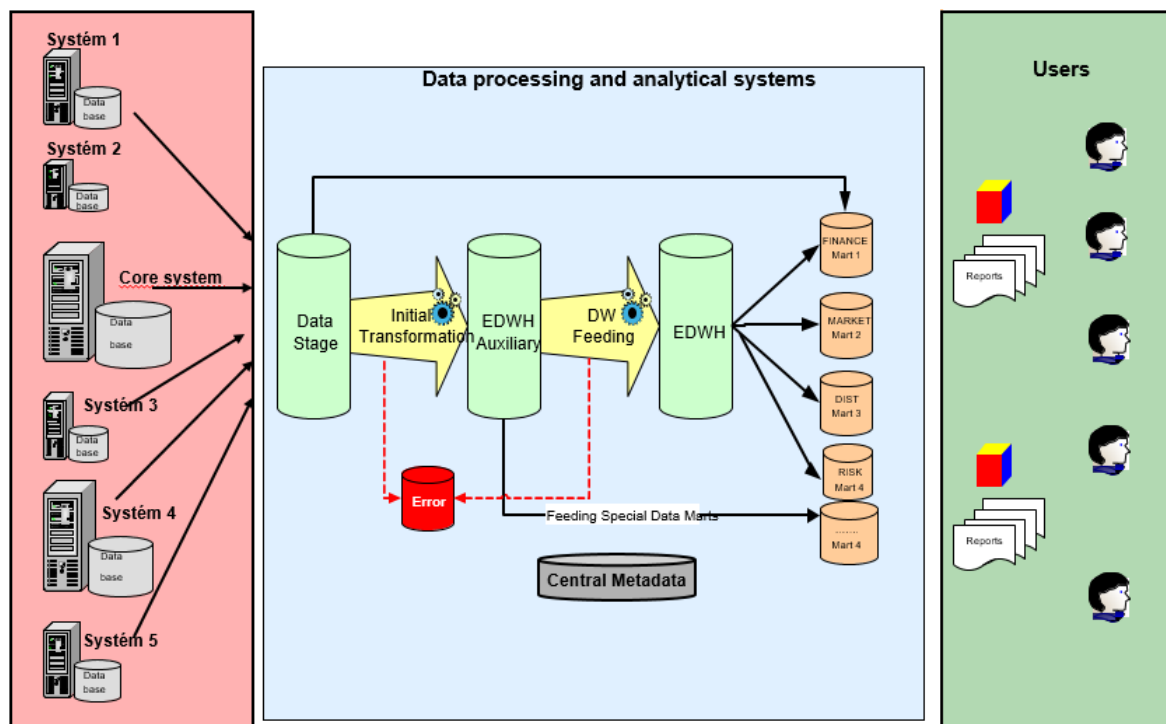
Pro správu datové báze se používá nástroj SQL Server Management Studio, nad kterým je jako nadstavba používán WhereScape Red (WS Red), což umožňuje zjednodušení práce s datovou infrastrukturou, neboť díky exportu modelu z WS 3D není třeba rozsáhlého psaní kódu. Ve WhereScape Red dochází k plnění tabulek pomocí scheduleru na bázi denních loadů a k transformaci a agregaci některých atributů.

Datový sklad (DWH)

Společnost využívá hybridní model datového skladu (tj. kombinace mezi Kimballovou a Inmonovou metodologií) normalizovaný na třetí normální formu (3NF), který podporuje analytické a reportingové potřeby (Naeem 2024).

Cardif má celou řadu zdrojových systémů kterými jsou pojišťovací a účetní systémy a systém pro správu kmenových dat. Ke sběru dat ze zdrojových systémů dochází v landing zóně (LDZ). Do LDZ se dostávají surová data bez jakýchkoliv transformací. Z LDZ se data dostávají do operativního úložiště dat (ODS), kde dochází k drobným změnám (například

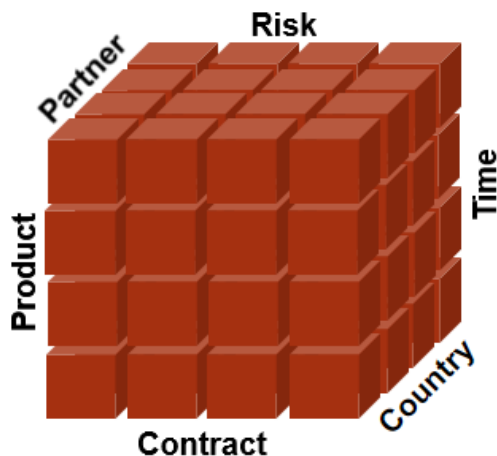
narovnání datových typů) a error handlingu. Z ODS data putují do jádra datového skladu (CORE), kde dochází k transformacím a normalizaci na třetí normální formu. Poslední úrovní jsou aplikace a datamarty. Každá vrstva datového skladu má přidělený specifický trojčíslíkový identifikátor, který je součástí jmenné konvence datových objektů.



Obrázek 14 Obecný přehled datové architektury v BNPP Cardif (interní dokumentace)

Na rozdíl od jádra DWH, je struktura datamartů de-normalizovaná a odpovídá hvězdicové struktuře (star schema). Přístup k datům v datových tržištích je umožněn pomocí OLAP kostek, ke kterým koncoví uživatelé přistupují ve formě kontingenčních tabulek díky SSAS spojení (SQL Server Analysis Services) prostřednictvím Microsoft Excel.

Obsah OLAP kostky je řízen požadavky byznys uživatelů, kteří určují rozsah atributů a pořadí v kterém mají být k dispozici. Sloupce v kostce jsou kategorizovány do logických celků, které představují dimenze a metriky (viz Obrázek 15).



Obrázek 15 Příklad hlavních dimenzí OLAP kostky v BNPP Cardif (interní dokumentace)

V době psaní této práce existují celkem čtyři datová tržiště. Jak bylo vymezeno v úvodu kapitoly 3, kvůli komplexnímu datovému prostředí bylo řešení zjednodušeno na vybraný geografický region a vybrané datové prvky. Řešení bylo aplikováno na úroveň datových tržišť, přesněji na datamart s názvem EPM (Enterprise Performance Management).

EPM je klíčovým prvkem podnikového strategického výkonnostního řízení. Nabízí konzistentní pohled na business trendy v různých časových obdobích a napříč různými regiony a usnadňuje strategické plánování a prognózování díky integraci klíčových ukazatelů (KPIs) na jednom místě. Vzhledem k jeho důležitosti byl vybrán jako vhodný kandidát, kde data governance může přinést okamžitý přínos a hodnotu, protože obsah datamartu je konzumován přímo byznys uživateli, kteří díky těmto aktivitám lépe porozumí datům, která konzumují.

Kromě EPM datamartu existují v BNP Paribas Cardif ještě další specializované datamarty, které se zaměřují na jiné klíčové oblasti podnikových aktivit, například k podpoře procesů finanční uzávěrky, plnění požadavků regulačních rámců a transparentní výkaznictví vůči regulatorním orgánům nebo podporu marketingových aktivit, analýzu kampaní a segmentaci zákazníků pro efektivnější cílení pojišťovacích produktů. Vzhledem ke zmiňovaným vysokým nárokům na bezpečnost a ochranu dat a informací nebudou tyto datamarty, a některé další prvky architektury, podrobněji popisovány.

3.1.2 Správa a řízení kmenových dat

Systémy pro správu kmenových dat zajišťující jednotnost a konzistenci klíčových datových entit. V BNPP Cardif je pro tyto účely užíván systém zvaný RMS (Referential Management Solution), který je provozován pomocí Microsoft Master Data Services (MDS). Důležitým subsetem, který se v tomto systému nachází, je sada entit, která jsou ručně spravována byznys uživateli a přenášena do datového skladu pro další využití. Jedním z těchto entit jsou také entity byznys dat a byznys termínů, které jsou zdrojem pro byznys slovník. Kromě toho se zde nachází připravené šablony pro import aplikačních dat do EA či WS 3D.

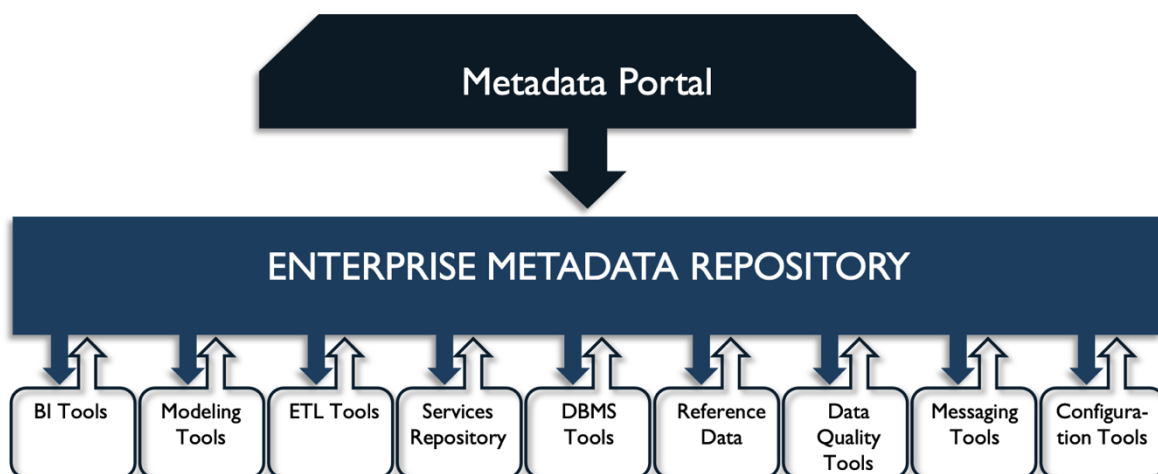
3.1.3 Úložiště metadat

Poslední složkou, kterou je vhodné zmínit je úložiště pro správu metadat. V BNPP Cardif je správa a řízení metadat centralizována v data governance platformě Collibra, která funguje jako data katalog. Ten organizaci umožňuje řízení metadat napříč všemi zeměmi, čímž poskytuje jednotný přehled o podnikových datech. Toto řešení je v Cardifu označováno jako Datapedia.

Platforma je strukturována do dvou úrovní:

- Globální úroveň – zastřešuje centrálně definované metadatové standardy a procesy, které jsou platné napříč celou organizací.
- Lokální úroveň – umožňuje jednotlivým regionům přizpůsobit správu metadat podle specifických požadavků.

Společnost využívá centralizovanou metadata architekturu (viz Obrázek 14), která se skládá z jediného úložiště metadat, které obsahuje kopie metadat z různých zdrojů. To nabízí jisté výhody, jako například vysoká dostupnost, rychlé vyhledávání a možnost obohacení či transformování sesbíraných metadat. Přináší s sebou však i jisté výzvy, z nichž největší je složitost synchronizace metadat se zdrojovými systémy pro zajištění konzistence (DAMA International 2017).



Obrázek 16 Centralizovaná metadata architektura (DAMA International 2017)

Úložiště je naplňováno prostřednictvím pravidelného plánovaného importu metadat z různých nástrojů. Centralizované úložiště je možné prohlížet prostřednictvím metadata portálu (v tomto případě Collibra). Globální vyhledávání napříč metadaty shromážděnými z různých nástrojů je možné díky jejich centralizovanému sběru v jednom úložišti (viz Obrázek 16).

Datapedia se skládá ze tří komponent:

- Datový slovník – obsahuje popisy databázových objektů, aplikací a jejich vzájemných vztahů.
- Byznys slovník – uchovává byznys pojmy, definice datových prvků a strukturu reportů, což usnadňuje propojení mezi byznys a technickými metadaty.
- Databáze pro integraci dat – umožňuje konsolidaci, verifikaci a publikování dat do metadata portálu.

Kromě katalogizace dat zahrnuje datapedia také data lineage a měřítko kvality dat (přesnost, konzistence, aktuálnost atp.), což napomáhá transparentnosti a důvěryhodnosti dat napříč organizací. Tím se zajišťuje, že všichni uživatelé pracují s jednotným, aktuálním a ověřeným datovým modelem, což je zásadní pro úspěšnou implementaci data governance.

3.2 Analýza stávajícího stavu dokumentace

Prvním krokem při řešení problematiky byznys a datových slovníků v rámci data governance bylo pochopení aktuálního stavu dokumentace ve společnosti BNP Paribas Cardif. Tato úvodní analýza sloužila jako nezbytný základ pro další práci, protože bez

pochopení kontextu by bylo obtížné identifikovat konkrétní problémy a navrhnout efektivní řešení. V této kapitole proto budou popsány hlavní aspekty analýzy současného stavu dokumentace – jaké byly zjištěné nedostatky, jakým způsobem byla metadata spravována a jaké překážky bránily efektivnímu využívání byznys a datových slovníků.

Nutnou podmínkou pro tuto aktivitu (a všechny následující) je dostatečná znalost kontextu a problematiky v oblasti pojišťovnictví. Stejně jako každé odvětví, i pojišťovnictví má řadu doménově specifických termínů, metrik a regulatorních požadavků, které je nutné znát, aby bylo možné správně analyzovat současný stav správy dat.

Kromě toho má každá organizace vlastní terminologii, kterou zaměstnanci používají, vlastní pracovní postupy a specifické požadavky neboli firemní kulturu, se kterou je třeba se seznámit, než lze začít efektivně pracovat na zlepšení dokumentace a řízení metadat.

Již v samotném úvodu řešení bylo tedy třeba eliminovat 2 výzvy:

- Získání dostatečné míry odborných znalostí.
- Adopce firemní kultury.

Překonání těchto výzev je naprosto klíčové pro další práci, protože bez toho nelze pochopit současný stav, správně interpretovat dostupná data a navrhnout funkční řešení. Z tohoto důvodu je vhodné vysvětlit, jakým způsobem byly tyto výzvy redukovány či eliminovány.

Získání odborných znalostí

Získání dostatečných odborných znalostí je klíčové nejen pro pochopení současného stavu ale rovněž i pro přípravu nové dokumentace. V případě tvorby a údržby efektivních slovníků obchodních pojmů, dostatečné znalosti zajišťují, že definice ve slovníku jsou relevantní, přesné a užitečné pro organizaci, což v konečném důsledku zlepšuje komunikaci, spolupráci a rozhodovací procesy. Při analýze současného stavu lze na základě těchto znalostí posoudit stav existující dokumentace. Kromě toho může nesprávné pochopení termínů způsobit chybnou interpretaci dat a způsobit nesrovnalosti ve výpočtech různých atributů.

Nabytí takových znalostí nemusí být z počátku jednoduché a vyžaduje značnou míru úsilí, vlastní iniciativy a proaktivního přístupu. Počáteční znalosti lze získat primárně třemi způsoby – studiem interních a externích materiálů a opakovanou komunikací se zkušenými odborníky ve společnosti.

Z interních materiálů šlo především o studium dostupné technické a obchodní dokumentace, prezentací a vizuálů, propagačních materiálů, onboardingových materiálů a kurzů, interních portálů a interní wiki. Kromě toho byl využit i samotný byznys slovník v Collibře, který však sloužil spíše doplňkový zdroj pro získání základního přehledu o některých termínech. Teoreticky by měl byznys slovník poskytovat oficiálně validované termíny a jejich definice, což by z něj činilo ideální nástroj pro rychlou orientaci v doménové terminologii. Nicméně, vzhledem k tomu, že jedním z cílů bylo analyzovat kvalitu této dokumentace, bylo nutné k němu přistupovat kriticky a s jistou mírou skepticismu.

V případě externích zdrojů byla využita především celá řada slovníků pojišťovacích termínů, například (National Association of Insurance Commissioners [b.r.]), (Association of British

Insurers [b.r.]) nebo (Česká asociace pojišťoven [b.r.]). Kromě toho byly využity i další weby, které se souhrnně zabývají problematikou pojišťovnictví. Zde je vhodné zdůraznit, že studium externích zdrojů trpělo následky zmiňovaných silných nároků na zaručení bezpečnosti a ochrany, protože podstatná část webů byly zablokovány, což znemožnilo přístup k celé řadě informací, které by jinak bylo možné využít. Kromě toho by v současné době dávalo smysl především pro získání nějaké počáteční znalosti jakékoliv problematiky využít jeden z dostupných jazykových modelů umělé inteligence, které mají široké znalosti napříč různými oblastmi a umožňují rychlou orientaci v odborné problematice. I tyto externí zdroje byly však bez výjimky blokovány, takže nebylo možné je jednoduše využít.

Získání počátečních znalostí o pojišťovnictví a správě dat ve společnosti BNP Paribas Cardif nebylo možné pouze studiem interních materiálů a veřejně dostupných zdrojů. Klíčovou roli hrály konzultace s odborníky z různých oblastí společnosti, díky kterým bylo možné nejen lépe pochopit specifickou terminologii, ale také získat hlubší vhled do interních procesů a kontextu nad rámec dokumentace. Díky odborníkům z finančního oddělení byl získán přehled, jak se v praxi počítají klíčové metriky. Zaměstnanci z datového týmu, poskytnuli informace o tom, jak jsou data v organizaci strukturována, kde jsou uložena a jak se transformují mezi jednotlivými systémy. Odborníci z oddělení rizik a regulací nabídli informace o tom, jaké regulatorní požadavky ovlivňují správu dat a jaký dopad mají na data governance. Byznys analytici umožnili pochopit, jakým způsobem jsou v Cardifu spravovány pojistné produkty a jaké pojmy jsou důležité pro jejich řízení. Pochopit veškeré detaily z technické i obchodní stránky je obtížné a nemožné během jednorázové komunikace, je třeba znovu zdůraznit, že jde o opakovanou komunikaci s postupnými přírůstky a rozšiřování povědomí v průběhu celého řešení.

Kromě aktivní komunikace se ukázalo být výhodné účastnit se pasivně dalších podnikových meetingů pro získání většího povědomí o širším kontextu a účastnit se schůzek různých oddělení, zaměřených na představení různých konceptů.

Adopce firemní kultury

Firemní kultura nemá žádnou formalizovanou definici, ale běžně zahrnuje poslání, vizi a hodnoty organizace a ovlivňuje způsob, jakým zaměstnanci vzájemně komunikují, přijímají rozhodnutí a obecně jakým způsobem přistupují k práci (viz kapitoly 1.3.1 a 1.4.3). V tomto případě je klíčové především porozumění internímu jazyku, běžně používaným zkratkám, výrazům a interním názvům, referencím a postupům, aby bylo zajištěno, že jsou definice v souladu s existující terminologií, komunikačními vzorci a kontextuálními rámci organizace. Pochopení jedinečného jazyka, citlivosti a mocenské dynamiky v rámci BNP Paribas Cardif zajišťuje, že definice budou rezonovat se zaměstnanci, odkazovat na příslušné obchodní procesy a respektovat organizační hranice. Toto kulturní povědomí přímo ovlivňuje míru přijetí, protože definice, které odpovídají tomu, jak společnost skutečně funguje, budou s větší pravděpodobností přijaty a využívány.

Porozumět firemní kultuře není něco, co lze zvládnout za jeden den nebo i první týden v organizaci. Existují však aktivity, pomocí kterých ji lze rychle pochopit a začít se v ní pohybovat více přirozeně. V podstatě jde opět o kombinaci studia dokumentace a aktivní participace a komunikace. Některé činnosti provedené pro získání odborných znalostí zároveň přispěly i k seznámení s firemní kulturou – studium některých materiálů,

konzultace s odborníky a pozorování schůzek, rozhodovacích procesů a každodenního provozu.

Kromě formálních rozhovorů s odborníky přispěli k adopci firemní kultury velmi výrazně také neformální rozhovory. Neformální rozhovory se zaměstnanci jsou jedním z nejúčinnějších způsobů, jak poznat firemní kulturu, neboť odhalují nepsané normy, sdílené hodnoty a praktické poznatky, které formální dokumentace často opomíjí. Během přestávek na kávu, obědů nebo diskusí na chodbách lze zjistit, jak se rozhodnutí skutečně přijímají, které pojmy mají v týmech specifické konotace a kde se skrývají potenciální citlivé body. Kolegové často sdílejí upřímné pohledy na to, jaké předchozí iniciativy byly úspěšné nebo neúspěšné a proč, což poskytuje cenný kontext. Tyto neformální výměny názorů také pomáhají budovat vztahy s klíčovými nositeli znalostí, kteří mohou později potvrdit vaše obchodní definice nebo se zasadit o jejich přijetí. Navíc věnování pozornosti jazyku, příkladům a odkazům, které jsou dalšími zaměstnanci přirozeně používány, poskytne autentickou terminologii, kterou lze začlenit do dokumentace, a zajistit tak, že bude odrážet, jak práce skutečně probíhá.

Zároveň pro pochopení firemní kultury pomohli různé další firemní zdroje nad rámec těch, zmíněných při popisu získání odborných znalostí, například standardní operační postupy nebo příručky, projektová dokumentace, nebo firemní zásady, směrnice, vize a cíle a podobně.

Tyto dvě prerekvizity byly základním kamenem pro všechny další aktivity. Bez nich by bylo téměř nemožné správně analyzovat existující dokumentaci a navrhnout konzistentní byznys a datové slovníky. Na druhou stranu je důležité zdůraznit, že tyto kroky nelze provést jednorázově, firemní kultura se stále mění a znalosti je nutné pravidelně rozšiřovat, jde spíše a výchozí bod, který umožnil získat dostatečné množství znalostí k tomu, aby bylo možné začít pracovat na skutečné implementaci.

Aby byl stav dokumentace důkladně posouzen, bylo provedeno několik kroků – revize aplikačních dat (tj. entity a atributy) v databázi datapedie, revize aplikačních dat v EPM datamartu pro posouzení konzistence mezi skutečnými objekty v datamartu a objekty dostupnými na platformě Collibra (důležité především kvůli centralizované metadata architektuře). Kromě toho byla posouzena také byznys data a úroveň integrace mezi obchodními a aplikačními daty.

3.2.1 Aplikační data

Pro účely analýzy aplikačních dat byly navrženy vlastní reporty s vybranými atributy v nástroji WhereScape 3D. WS 3D umožňuje přípravu vlastních reportů, díky kterým lze vytvářet přehledy z datových modelů, což umožňuje uživatelům získat specifické poznatky, například testovat návrhy schémat oproti skutečným datovým výstupům, kontrolovat konzistenci metadat nebo prohlížet vzorová data při vytváření prototypů. Přehledy lze následně exportovat pro externí analýzu či dokumentaci.

Celkem byly připraveny dva přehledy, jeden pro entity a jeden pro jejich atributy. Pro každý report bylo nutné vybrat vhodnou sadu parametrů, které budou sledovány pro datové

objekty a jejich komponenty. Výběr sledovaných parametrů vycházel jednak z dostupných atributů, které již byly podporovány ve WhereScape 3D a mohly být jednoduše dokumentovány, a jednak z dalších pomocných parametrů, které dosud dokumentovány nebyly, ale jejich zavedení mohlo významně podpořit kvalitu dokumentace. Parametry, které byly ve WS 3D již podporovány, vycházely z požadavků BNPP Cardif na data governance, které byly definovány nezávisle na tomto řešení (viz kapitola 2).

Tabulky

Pro analýzu kvality metadat u tabulek bylo zvoleno pět parametrů, které jsou shrnuty v tabulce 1, jež se skládá z názvu parametrů, stručného popisu a příkladu. Příklady slouží spíše k ilustračním účelům, přestože byly zvoleny tak, aby byly podobné datovým elementům v BNPP Cardif, nejde o skutečný příklad vybraný z databáze. Parametry byly zvoleny v anglickém jazyce, aby byly srozumitelné pro anglicky mluvící členy týmu, čímž se usnadnila širší spolupráce a zapojení více stakeholderů.

Tabulka 1 Popis parametrů – datové objekty

Parametr	Popis	Příklad
Object name	název tabulky	fact_401_Premium
Table type	typy tabulek slouží jako logická kategorizace, která určuje, jak jsou tabulky strukturovány a zpracovávány při převodu a exportu modelu	20_fact_m100
Comment	formální popis účelu objektu	The table contains information on premiums written and paid under individual insurance contracts. It is used to analyze the development of premiums and to monitor the fulfilment of clients' obligations.
Transformation rule	přehled o tom, jak jsou data v daném objektu upravována a agregována	Calculation of the premium based on the product's tariff: Premium Amount = Base Premium × Risk Factor × Discounts. Data is updated monthly based on payments received. If a payment is missing, the premium is marked as "Overdue".

Term reference	vazba objektu na obchodní termín v byznys slovníku	3#Premium
----------------	---	-----------

Sloupce

V případě atributů bylo do reportu zvoleno celkem 12 parametrů. Parametry jsou shrnuty v tabulce 2, která má stejnou strukturu, jako tabulka číslo 1. Uvedené parametry a příklady se řídí stejnou logikou, která je popsána výše při výběru parametrů pro datové objekty.

Tabulka 2 Popis parametrů – datové atributy

Parametr	Popis	Příklad
Attribute name	název sloupce	GWP_Policy_Active
Table name	název tabulky, k níž sloupec náleží	fact_401_Premium
Comment	formální popis účelu atributu	The total amount of Gross Written Premium (GWP) for active policies in a given period. Includes all policies in force without cancellation or cancellation.
Data type	datový typ	decimal
Precision	celkový počet číslic, které lze do čísla uložit	38
Scale	počet povolených číslic napravo od desetinné čárky	12
Column Order	pořadí sloupce v tabulce	6
Is Nullable	označení, zda sloupec může obsahovat nulové hodnoty	True
Is Primary	označení primárního klíče	False
Transformation rule	popis, jakým způsobem je sloupec kalkulován	The total amount of Gross Written Premium (GWP) for active policies in a given period. Includes all policies in force without cancellation or cancellation.

Source table	původ sloupce nebo výčet objektů, které jsou použity při jeho kalkulaci	Policy, Premium_Transactions
Term reference	Vazba objektu na obchodní pojem v byznys slovníku	36#GWP Policy Active

Oba připravené reporty byly aplikovány v EPM datamartu a databázi datapedie odkud byly vybrány pouze záznamy spojené s EPM. Veškeré výstupy byly exportovány do formátu CSV a následně nahrány do Microsoft Excel pro externí analýzu.

Pro kontrolu shody mezi datapedií a EPM byla nejprve použita kombinace podmíněného formátování a funkce MATCH (POZVYHLEDAT) v MS Excel, která umožnila automatické vyhledání shod a nesrovnalostí mezi objekty a atributy. Výsledek byl následně manuálně ověřen, aby se zajistila přesnost detekovaných nesrovnalostí a správná interpretace nalezených rozdílů. U záznamů, které se shodovali bylo následně zkontrolováno, zda se shodují i další parametry. Při komparaci byla jako referenční bod považována data z datamartu, protože odpovídají skutečně nasazeným a aktivně využívaným objektům. Na rozdíl od datapedie, která slouží primárně k dokumentačním účelům a je závislá na správné synchronizaci, EPM reflektuje aktuální stav datových struktur a obsahuje nejaktuálnější informace používané v produkčním prostředí.

Součástí analýzy byla rovněž kontrola chybějících hodnot (definice, transformační pravidla, reference na byznys data), duplicitních záznamů, soudržnosti jmenných konvencí (především u názvu objektů). Chybějící hodnoty a duplicitní záznamy byly kontrolovány pomocí filtrů a podmíněného formátování a následné manuální kontroly.

Výsledky analýzy byly důkladně dokumentovány a komunikovány a lze je shrnout takto:

- Datapedie obsahuje zastaralé objekty a atributy, které se již v datamartu nepoužívají. Naopak v datapedii chybí některé novější záznamy. Obě tyto zjištění naznačují, že jde o problém v update a delete procedurách, které jsou použity pro synchronizaci dat do centralizovaného úložiště metadat. Tento problém byl eskalován vývojovému týmu.
- Definice jsou nedostatečné a nekonzistentní, což je pravděpodobně důsledkem toho, že vývoj probíhá ve vícečlenném týmu a pro definice nejsou pevně daná pravidla nebo nejsou respektována. Ve většině případů definice úplně chybí.
- Transformační pravidla jsou nedostatečně popsána. Často nejsou uvedena nebo jsou popsána pouze jako výpis programového kódu namísto srozumitelného vysvětlení logiky.
- Drobné nekonzistence v datových typech a názvosloví. Podobné sloupce mají rozdílné datové typy a tabulky či atributy mají občas názvy, které nejsou v souladu s definovanou jmennou konvencí.
- Přestože u některých tabulek lze pozorovat logické seskupení a pořadí sloupců, není vždy konzistentně respektováno.

- Neexistují reference na byznys data a byznys termíny na úrovni tabulek, ani sloupců. Mezi technickou a obchodní vrstvou dat neexistuje propojení, což je výraznou překážkou v budování důvěryhodné a přehledné dokumentace.

OLAP

Aby byla analýza kompletní, došlo v závěru této činnosti ještě k revizi stavu OLAP kostky, ke které lze přistupovat v Microsoft Excel pomocí SSAS spojení, čímž je vytvořena vazba mezi sešitem a databázovým serverem (viz kapitola 3.1). Obsah kostky byl porovnán s výpisem objektů a atributů datového tržiště EPM.

Tato kontrola ukázala, že kostka obsahuje velké množství záměrně skrytých metrik, přičemž, až na pár z nich, které slouží k mezi-výpočtům, jsou skryté z toho důvodu, že nejsou byznys uživateli využívány a jsou považovány za zbytečné. Skryté atributy jsou stále dostupné v datamartu, ale nejsou již nadále užitečné. Zjištěné rozdíly byly opět dokumentovány a nadbytečné metriky byly označeny za kandidáty ke smazání v EPM.

Kromě nedostatků byly rovněž při analýze získány poznatky, které jsou užitečné k lepšímu porozumění datovému prostředí a pro přípravu dalších částí řešení. Například pravidla jmenných konvencí nebo podrobná transformační pravidla ve formě programového kódu zdokumentována v systému RMS.

Parametr Source Table byl přidán až v průběhu analýzy, kdy bylo zjištěno, že u atributů zatím nejsou řádně dokumentovány zdroje, což znemožňuje sledovat datové toky. Jde tedy o příklad pomocného atributu, který zůstal při prvotní analýze prázdný.

3.2.2 Obchodní data

V rámci analýzy bylo nezbytné posoudit i stav obchodních dat, aby bylo možné učinit závěr o celkovém stavu dokumentace. Zejména bylo důležité porozumět jejich struktuře, způsobu správy a stávající kvalitě.

Data byla získána ze systému RMS (viz 3.1.2) ze specifické entity byznys dat, která je spravována byznys uživateli. Záznamy z RMS byly následně exportovány do MS Excel, kde probíhala externí analýza podobným způsobem, jako v případě aplikačních dat.

Původní soubor obsahoval celkem 18 parametrů. Pro účely analýzy bylo vybráno deset parametrů, které jsou uvedeny v tabulce 3.

Tabulka 3 Popis parametrů – obchodní data

Parametr	Popis	Příklad
ID	unikátní identifikační kód	1527
Name	název obchodního pojmu	Net Promoter Score
Definition	formální vymezení pojmu	A metric that measures customer loyalty by asking

		how likely they are to recommend a company.
Abbreviation	akronym nebo zkratka	NPS
Synonym	odkaz na pojem, který je synonymem	(1871) Customer Loyalty Index
Business domain	popis oblasti podniku, do které pojem spadá	Marketing department
Ownership	obchodní jednotka, tým či osoba odpovědná za správu pojmu	CX Team
Date	Datum publikace nebo aktualizace	2024-03-10
Status	status pojmu – Candidate, Validated, Deprecated	Validated
Replaced by	odkaz na pojem, který je náhradou za zastaralý pojem	-

Soubor obsahoval velké množství pojmů pokrývajících širokou škálu oblastí napříč celou organizací. Z tohoto důvodu nebyla provedena vyčerpávající analýza každého jednotlivého termínu, ale šlo především o to, identifikovat obecné vzorce, zákonitosti a opakující se prvky v rámci správy a kvality byznys dat. Pozornost byla věnována především klíčovým prvkům, jako jsou konzistence definic, existence duplicit, chybějící hodnoty, vazby mezi termíny a způsob jejich kategorizace. Tento přístup umožnil získat ucelený přehled o struktuře a fungování slovníku bez nutnosti podrobného zkoumání všech jednotlivých pojmů.

Klíčová zjištění byla dokumentována a komunikována a lze je shrnout takto:

- Některé pojmy jsou popsány důkladně, jiné pouze minimálně či vůbec, což snižuje jejich použitelnost.
- Lze narazit na duplicitní výskyt pojmů, ostatní parametry však nemusí být shodné a zpravidla má vlastní identifikační kód. To způsobuje riziko, že jeden pojem nemá napříč organizací stejnou definici. Nalezené duplicity byly dokumentovány, předány týmu orientovanému na řízení dat a nerelevantní záznamy byl označeny jako deprecated (vyřazené).
- Existuje oddělená entita od byznys dat zvaná byznys termíny nebo obchodní termíny. Ty představují nadřazenou množinu pojmů, které popisují obecnější koncepty používané v organizaci. Byznys pojmy tedy lze uvádět na úrovni sloupce a byznys termíny na úrovni tabulky. Přestože jde o oddělenou entitu, lze na termíny narazit i při analýze byznys dat a obě entity mají stejný formát ID, které se vzájemně překrývají a nejsou vůči sobě unikátní.

- Proces validace byznys definic je poměrně komplexní a časově náročný. Návrh na validaci provádí data management officer, který poskytne daný pojem k validaci skupině osob na manažerské úrovni. Pokud je definice vyhovující a skupina se shodne, potom lze definici považovat za věrohodnou (validated). V opačném případě je zahozena nebo vrácena s návrhy na změnu (candidate). Tento proces vzhledem ke své složitosti ztěžuje možnost pravidelného vývoje a úpravy definic a ovlivnil také ochotu zaměstnanců navrhovat změny v dokumentaci.

Na základě analýzy byznys a aplikačních dat se ukázalo, že samotné vytvoření vazeb mezi obchodními a aplikačními daty nemusí stačit ke zlepšení datové kvality. Kvůli nízké kvalitě metadat na obou stranách je nutné nejen vytvořit integraci, ale také navrhnout úpravy na zlepšení samotné dokumentace.

Tyto zjištěné nedostatky potvrdily nízkou kvalitu dokumentace a nedostatečnou údržbu datapedie, což vedlo k návrhu konkrétních opatření ke zlepšení, která jsou popsána v následujících kapitolách.

3.3 Návrh dokumentace

Na základě závěrů učiněným provedením analytické fáze bylo možné přistoupit k návrhu dokumentace, která by řešila identifikované nedostatky. Návrh je rozdělen do dvou hlavních částí.

Nejprve proběhla fáze metadata alignmentu, tedy vytvoření vazeb mezi byznys daty a aplikačními objekty, aby byla zajištěna konzistence a lepší dohledatelnost dat v organizaci. Tato část, mimo jiné, umožnila hlouběji posoudit současný stav datového a byznys slovníku a na základě toho přistoupit ke druhé části návrhu – metadata improvement. Tato část se zaměřuje především na narovnání nedostatků v dokumentaci, zejména přípravou chybějících popisů a dalších parametrů a sjednocením používané terminologie. Obě fáze byly realizovány v úzké spolupráci s odborníky ve společnosti na dané oblasti, aby zohledňovaly jak technické, tak obchodní požadavky.

3.3.1 Metadata alignment

Prvním krokem v návrhu dokumentace bylo vytvořit vazbu mezi aplikačními a obchodními daty. Cílem této aktivity bylo zajistit, aby každý objekt a atribut v datamartu měl svůj protějšek ve formě byznys pojmu či termínu. Sjednocení technického a obchodního pohledu je klíčové pro překlenutí rozdílu mezi obchodní logikou a skutečnou technickou implementací, což podporuje důvěryhodnost, srozumitelnost a efektivitu práce s daty uživateli, kteří s daty pracují.

Vytvoření vazeb probíhalo prostřednictvím dokumentů vytvořených ve fázi analýzy. Na straně aplikačních dat byl jako referenční zdroj využit report s metadaty z EPM datamartu, který obsahoval tabulky a sloupce včetně jejich technických parametrů – rozsah zvolených parametrů je popsán v předchozí části (viz Tabulka 1 a Tabulka 2).

Aby bylo zajištěno, že proces bude efektivní a nezahrnuje zbytečné datové elementy, které pro obchodní analytiku nejsou relevantní, byly nejprve vyřazeny záznamy, které nejsou součástí OLAP kostky a byly označeny ke smazání během analýzy.

V případě obchodních metadat byl rovněž použit soubor vytvořený ve fázi analýzy se zvoleným rozsahem parametrů (viz Tabulka 3). Soubor sloužil primárně pro vyhledávání pojmů a termínů. Do technického reportu byly z tohoto souboru přidány 4 sloupce – identifikační kód, název pojmu či termínu, definice a status

Pro vytvoření prvotních vazeb byly využity pouze definice, které měly status validated. Byznys termíny se statusem candidate nebyly na počátku využity, protože nebylo zajištěno, že odpovídají ověřeným definicím používaným v organizaci.

Prvním krokem v tomto procesu bylo identifikovat a vytvořit vazby mezi aplikačními a byznys daty tam, kde bylo možné potvrdit jednoznačnou souvislost. Pro každou tabulku a sloupec v datamartu bylo zkontrolováno, zda existuje odpovídající pojem nebo termín v byznys datech. Primárně šlo tedy o propojení na základě totožného nebo velmi podobného názvu položek. Proces byl několikrát zopakován, aby bylo dosaženo maximálního počtu spojení.

Po vytvoření jednoznačných vazeb bylo nutné postupovat důkladněji, protože nyní už nebylo možné jednoduše nalézt odpovídající obchodní termíny pouze na základě shody názvů nebo klíčových slov v definici. Proces vyžadoval intenzivní spolupráci s odborníky na řízení dat ve společnosti, kteří měli znalosti z obchodního prostředí a rovněž určitou znalost technické implementace.

Na základě komunikace s odborným týmem, s cílem získat širší kontext, bylo možné identifikovat další správná spojení u objektů a atributů, kde se na první pohled zdálo, že neexistuje přímý vztah. Kromě toho proběhla s týmem i celková validace dokumentovaných vazeb, jejíž výsledkem bylo schválení či nabídnutí alternativního termínu, který je vhodnější. Tím bylo zajištěno, že vytvořené vazby jsou relevantní a použitelné.

Posledním krokem v rámci alignmentu bylo prohledat termíny se statusem candidate, které představují neschválené návrhy. Každá definice začíná jako kandidátní, zároveň však mohou existovat takové záznamy, které v tomto stavu setrvávají z různých důvodů. Těmi může být například, že byl záznam odmítnut, protože je chybný a nahrazen jiným nebo je duplicitní či příliš podobný již schválenému záznamu nebo neodpovídá aktuální terminologii v organizaci. Zároveň však může jít o pojmy, které mohou být relevantní, a ještě čekají na schválení nebo byly schváleny ale ještě nebyly formálně přesunuty do validovaného stavu. Relevance a kvalita těchto záznamů není nutně zaručena a je třeba k nim přistupovat se značnou mírou skepse, na druhou stranu mohou poskytnout další cenné informace a potenciálně rozšířit dokumentaci, pokud budou relevantní a budou v budoucnu validovány, což umožní průběžné zlepšování metadat. Využití těchto záznamů je závislé na faktické kontrole z odborného hlediska a validaci v rámci organizace pro posouzení, zda odpovídají obchodní realitě.

Nejvíce vazeb bylo vytvořeno v rámci prvotního prohledávání jednoznačných napojení. Při komunikaci s odborníky přibyly další spojení. Revize kandidátních pojmů se ukázalo být

dobrým rozhodnutím, protože potvrdilo zjištění z analýzy obchodních dat, týkající se složitého procesu validace definic. Kvůli tomu existovala velká spousta kandidátních termínů, z nichž celá řada z nich se jevila jako relevantní, ale nebyla schválena, pouze z důvodu náročnosti schvalovacího procesu. Tyto pojmy byly v dokumentu poznamenány jako potenciální vazby, ale nebylo možné je v současném stavu nasadit, dokud nebudou oficiálně přijaty. Naopak některé záznamy, které nebylo možné rovnou použít, posloužili jako vhodný základ v dalších fázích řešení.

V rámci aktivity metadata alignmentu bylo analyzováno celkem 12 tabulek, přičemž každá z nich obsahovala přibližně 15–30 atributů. Pro každý atribut i tabulku byla snaha nalézt odpovídající obchodní protějšek v byznys slovníku. Vazby nebyly uvažovány pro primární a cizí klíče. Výsledkem bylo, že se podařilo vytvořit vazbu přibližně u 70 % záznamů – tedy ve většině případů byl technický datový prvek úspěšně napojen na relevantní obchodní termín. Je však důležité podotknout, že úspěšné vytvoření vazby nutně neznamená, že daný obchodní termín je i kvalitně definován. I mezi termíny se statusem validated se nacházely definice, které sice byly oficiálně schváleny, ale z hlediska obsahové úplnosti nebo srozumitelnosti byly nedostatečné. Ukázka dokumentu s naznačenými vazbami mezi objekty a byznys termíny je k vidění na obrázku 17. Vazby jsou naznačeny pomocí sekce business data (označena zeleně). Kromě toho je kombinace identifikačního kódu a názvu zaznamenána ve sloupci „TERM_REFERENCE“ pomocí Excel funkce CONCAT ve formátu „<ID>#<NAME>“. Tento sloupec bude blíže popsán ve finální fázi řešení.

APPLICATION DATA					BUSINESS DATA			
OBJECT_NAME	TABLE_TYPE	COMMENT	TRANSFORMATION	TERM_REFERENCE	ID	NAME	DEFINITION	STATUS
fact_401_Claim	20_fact_m100	Contains records o		11#Claim	11	Claim	A request made	Validated
fact_401_Premium	20_fact_m100			3#Premium	3	Premium	Money charged	Validated
fact_401_Policy				7#Policy	7	Policy	Insurance policy	Validated
dim_401_Calendar	30_dim_m100	Calendar date dim		9#Calendar	9	Calendar	Calendar date d	Candidate

Obrázek 17 Metadata alignment mezi datovými objekty a byznys termíny (vlastní zpracování)

Metadata alignment představoval klíčový krok v propojení technických datových struktur s jejich obchodním významem. Díky této aktivitě vznikla první ucelená datová mapa, která představuje základ na cestě k lepšímu porozumění datům jak pro obchodní, tak pro technické uživatele. Současně tato fáze potvrdila slabá místa nebo zcela chybějící vazby, čímž vznikl přehled o tom, které části dokumentace je třeba doplnit, upravit nebo nově vytvořit, čímž se otevřela cesta pro další fázi řešení, zaměřenou na zlepšení dokumentace.

3.3.2 Metadata improvement

Proces metadata improvementu představoval nejnáročnější, ale zároveň nejdůležitější část celého řešení. Cílem této fáze bylo napravit identifikované nedostatky ve stávající dokumentaci a doplnit klíčové informace, které dosud chyběly nebo byly zpracovány pouze částečně. Na rozdíl od předchozí fáze, která se zaměřovala na vytváření vazeb mezi aplikačními a obchodními metadaty, zde došlo ke skutečné transformaci kvality dokumentace – a to jak na technické, tak na obchodní úrovni.

Tato etapa si vyžádala značné množství času, hluboké porozumění doménové i technické problematice a úzkou spolupráci s odborníky z různých oblastí podniku. Jednalo se o

detailní a iterativní práci, která zahrnovala nejen opravu a doplnění technických metadat, ale také tvorbu obchodních definic, a tím i zvýšení celkové srozumitelnosti a důvěryhodnosti dat v organizaci.

Aktivity provedené v rámci této kapitoly byly rovněž dokumentovány do souboru použitého v předchozích kapitolách.

Technická metadata

První aktivitou v rámci úpravy technických metadat bylo *odstranění nekonzistencí* u vybraných technických parametrů, zejména názvů tabulek a sloupců, pořadí atributů a narovnání datových typů. Nekonzistence byly identifikovány v rámci analýzy současného stavu dokumentace (viz kapitola 3.2.1). V případě názvů šlo pouze o drobné úpravy pro zajištění souladu s jmennými konvencemi organizace. Primárně šlo o přidání správných číselných prefixů k názvům objektů pro identifikaci úrovně v datovém skladu. Datamart vrstva datového skladu má označení 4XX – v případě EPM prefix odpovídá hodnotě 401. Celý název datamartu tedy zní 401_EPM. Úpravy odpovídají požadavkům definovaným podnikovou datovou architekturou. V případě sloupců byly provedena analýza dopadů pomocí WhereScape RED a na základě analýzy bylo rozhodnuto, že nebudou provedeny žádné změny, protože by to znamenalo značné následky na procedury, kde by bylo třeba opravit všechny výskyty daného sloupce.

Ani v případě datových typů nešlo o velké operace. Jednalo se pouze o standardizaci či optimalizaci rozsahu některých decimálních a číselných datových typů u vybraných metrik, například `decimal(18,6)` → `decimal(38,12)` nebo `bigint` → `int`. V některých případech měla stejná skupina metrik rozdílná nastavení datových typů, která byla prostřednictvím této aktivity harmonizována.

Poslední činností, která byla soustředěna na konzistenci bylo definování vhodného logického pořadí atributů ve faktových tabulkách. Atributy byly seskupeny podle toho, jak spolu souvisí.

Entity level		
sk_Partner	bigint	1
sk_Product	bigint	
sk_Risk	bigint	
sk_Cohort	bigint	
sk_Channel	bigint	
Country_Code	nvarchar(2)	2
Business_Period	int	3
Business_Last_Month_Date	date	
GWP_plus_IPT_incl_Estimates_Cardif_Share	decimal(38,12)	4
GWP_plus_IPT_incl_Estimates_Full_Share	decimal(38,12)	4.1
GWP_Policy_Active	decimal(38,12)	
GWP_Policy_Inactive	decimal(38,12)	
GWP_Policy_New	decimal(38,12)	4.2

Obrázek 18 Pořadí atributů ve faktových tabulkách (vlastní zpracování)

V každé tabulce bylo stanoveno jasné pořadí atributů (viz Obrázek 18):

1. primární klíče,

2. kód země,
3. časová dimenze,
4. metriky.

Metriky byly rozpadnuty na několik celků podle souvislosti a jejich pořadí bylo určeno pořadím v OLAP kostce.

Popis transformační logiky

Významnou součástí přípravy technické dokumentace byl popis transformačních pravidel, která představují soubor instrukcí nebo operací, jež se aplikují na data v průběhu jejich zpracování. Na rozdíl od popisů obchodních pojmů, jejichž cílem je obecné vysvětlení významu jednotlivých atributů, transformační pravidla slouží k objasnění logiky manipulace s daty, jejich zpracování a transformaci. V kontextu EPM datamartu šlo konkrétně o popis metodiky výpočtu klíčových ukazatelů výkonnosti (KPI) uvedených ve faktových tabulkách. Sloupce z dimenzionálních tabulek byly většinou přebírány přímo z jádra beze změn.

Motivací pro tuto aktivitu byla nízká úroveň kvality současného popisu transformační logiky zjištěná při analýze. Transformační pravidla byla zcela chybějící, nebo existovala pouze ve formě kódu, který není běžným uživatelům srozumitelný. Přesná pravidla pro kalkulaci všech sloupců jsou dokumentována v RMS, cílem této aktivity však bylo vytvořit popis v běžném, čitelném jazyce tak, aby daný uživatel byl schopen se v logice rychle zorientovat a umožnit tak porozumění i běžným uživatelům. V některých případech byl do dokumentace zařazen i relevantní úryvek kódu, pokud to napomohlo srozumitelnosti.

Proces dokumentace transformačních pravidel vyžadoval značnou míru času, úsilí a nutnou znalost standardizovaného strukturovaného dotazovacího jazyka (SQL). Kromě toho byla logika podrobně komunikována se členy vývojového týmu, kteří danou část vyvíjeli. Výpočty jednotlivých sloupců byly dohledávány pomocí reverzního inženýrství v prostředí WhereScape RED. Transformační logika je rozdělena do více tzv. „work“ tabulek (viz kapitola 3.1.1), které jsou navzájem vnořené – to znamená, že řešení se skládá z hierarchie SQL skriptů, které jsou na závěr sjednoceny do jedné tabulky, ze které je vytvořena závěrečná tabulka. Každou z těchto vrstev bylo nutné projít postupně až na nejnižší úroveň, kde se objevovaly konkrétní výpočty, operace nebo odkazy na sloupce z jádra datového skladu. Typickými operacemi byly agregace, matematické výpočty, filtrování podle datových podmínek, použití parametrických proměnných nebo překládání kategorií pomocí kódových seznamů.

Zjednodušený příklad souboru pracovních tabulek používaných v procesu transformace je uveden níže (viz Výpis 1). Ačkoli logika věrně (byť zjednodušeně) odráží strukturu a složitost, s níž se bylo možné setkat při skutečné implementaci, samotný kód je zcela fiktivní a byl vytvořen pouze pro ilustrační účely. Vzhledem k požadavkům na bezpečnost a důvěrnost dat nelze sdílet žádný skutečný produkční kód ani obchodní logiku. Bylo však vynaloženo veškeré úsilí, aby příklad zůstal relevantní pro daný případ použití – konkrétně pro zpracování a výpočet klíčových ukazatelů výkonnosti (KPI) pro pojistné události v prostředí datového skladu.

Výpis 1 Ukázka sady pracovních tabulek pro kalkulaci KPIs (vlastní zpracování)

```
INSERT INTO wrk_claims_details
SELECT
    c.claim_id,
    c.claim_date,
    c.claim_amount,
    c.policy_id,
    p.product_code,
    p.policy_start_date,
    p.policy_end_date,
    clm_type.description AS claim_type_desc,
    DATEDIFF(DAY, p.policy_start_date, c.claim_date) AS days_since_policy_start
FROM dw_claims c
JOIN (
    -- Subquery to get only the most recent version of each policy
    SELECT policy_id, product_code, policy_start_date, policy_end_date
    FROM (
        SELECT *,
            ROW_NUMBER() OVER (PARTITION BY policy_id ORDER BY version_date DESC) AS rn
        FROM dw_policies
    ) latest
    WHERE latest.rn = 1
) p ON c.policy_id = p.policy_id
LEFT JOIN ref_claim_type clm_type ON c.claim_type_code = clm_type.code
WHERE c.claim_status = 'APPROVED';

INSERT INTO wrk_claims_aggregated
SELECT
    product_code,
    FORMAT(claim_date, 'yyyy-MM') AS claim_month,
    COUNT(claim_id) AS claim_count,
    SUM(claim_amount) AS total_claim_amount,
    AVG(claim_amount) AS avg_claim_amount
FROM wrk_claims_details
GROUP BY product_code, FORMAT(claim_date, 'yyyy-MM');

INSERT INTO wrk_claims_with_flags
SELECT
    a.product_code,
    a.claim_month,
    a.claim_count,
    a.total_claim_amount,
    a.avg_claim_amount,
    CASE
        WHEN a.avg_claim_amount > r.threshold_amount THEN 'HIGH'
```

```

        ELSE 'NORMAL'
    END AS risk_flag
FROM wrk_claims_aggregated a
LEFT JOIN ref_product_risk_thresholds r ON a.product_code = r.product_code;

INSERT INTO fct_claims_kpis
SELECT
    product_code,
    claim_month,
    claim_count,
    total_claim_amount,
    avg_claim_amount,
    risk_flag
FROM wrk_claims_with_flags;

```

Každá tabulka postupně přidává obchodní logiku, obohacení metadat nebo agregace. V textovém výpisu programového kódu jsou všechny části vloženy pod sebou, ve skutečnosti však každá tabulka byla uložena ve vlastním skriptu, které bylo třeba postupně projít.

Pro ilustraci tohoto procesu použijeme sloupec `risk_flag` z pracovní tabulky `wrk_claims_with_flags`. Sloupec je kalkulován na základě dvou různých sloupců, které pocházejí z různých zdrojů:

- `avg_claim_amount` – pochází z `wrk_claims_aggregated`, kde je spočítán jako průměrná výše pojistného plnění (`claim_amount`). `Claim_amount` pochází z tabulky `wrk_claims_details`, která tento sloupec přebírá z jádra datového skladu z tabulky `dw_claims`.
- `threshold_amount` – pochází z referenční tabulky `ref_product_risk_thresholds` z jádra skladu.

V tomto případě by popis transformačního pravidla vypadal takto:

„This column represents a business rule that classifies the risk level of a given product and month based on the average claim amount (avg_claim_amount). If the average amount of claims for a product in a given month exceeds the predefined threshold (threshold_amount) from the reference table, the product-month combination is flagged as HIGH risk. Otherwise, the risk level is marked as NORMAL.“

Jak již bylo zmíněno v kapitole 3.2, všechny výstupy byly připraveny v anglickém jazyce, vzhledem k tomu, že šlo o spolupráci v mezinárodním týmu. Veškeré poznatky získané během této fáze byly zaznamenány do sloupce `TRANSFORMATION_RULE` v dokumentačním souboru. Pokud byl sloupec výsledkem výpočtu, byl uveden stručný a srozumitelný popis logiky jeho kalkulace. V případech, kdy nedocházelo k žádné transformaci a sloupec byl pouze převzat z jiného zdroje, bylo v dokumentaci uvedeno, že sloupec byl převzat ze zdroje bez transformací. Tím bylo zajištěno, že každý sloupec měl jasně dokumentovaný způsob svého vzniku – ať už šlo o složitý výpočet, nebo jednoduchý

přenos hodnoty. Na úrovni tabulek byl připraven obecný popis, který byl následně přidán ke všem tabulkám v dokumentaci. Tento popis slouží jako orientační prostředek a definuje jakým způsobem porozumět jednotlivým sloupcům v systému RMS v entitě zaměřené na transformační logiku, která detailně technicky dokumentuje transformační pravidla každého atributu.

Na dokumentaci transformační logiky úzce navazovala také identifikace zdrojových objektů a sloupců, která je popsána v následující části.

Identifikace zdrojů

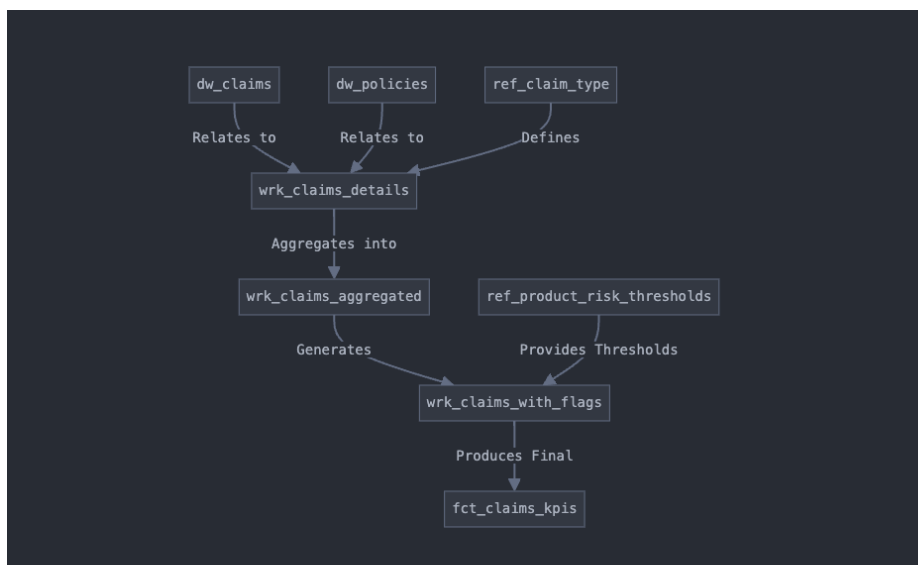
Obě aktivity jsou úzce propojené – aby bylo možné správně popsat výpočet, je zároveň nutné znát vstupy, ze kterých daný výpočet vychází. Výsledkem by měl být přehled, ve kterém má každý sloupec zaznamenán svůj výpočetní postup i zdroj, z něhož daná hodnota vznikla.

Identifikace zdrojů je aktivita, která probíhala současně s popisem transformační logiky, protože je výsledkem stejného reverzního inženýrství popsaného výše. V případě, že jde o sloupec, který je vypočítáván v datamartu, je nutné zaznamenat zdroje pro všechny části, které jsou potřebné k výpočtu. V případě, že jde o přenesení hodnoty bez transformací, je nutné zaznamenat odkud hodnota originálně pochází.

Pokud navážeme na příklad sloupce `risk_flag` popsaného dříve, pak by zaznamenání zdrojů vypadalo takto:

„dw_claims.claim_amount, ref_product_risk_thresholds.threshold_amount“

Přesné zaznamenání zdrojových tabulek a sloupců je nezbytné nejen pro pochopení způsobu manipulace s daty, ale také pro vytvoření přesného toku dat (viz Obrázek 19). Tyto informace hrály klíčovou roli při mapování toku dat z jádra datového skladu až ke konečným výstupům. Udržování jasného záznamu o těchto zdrojích zajistilo, že každou metriku a atribut v datovém skladu bylo možné dohledat až k jejímu původu, což je důležité zejména pro validaci, auditovatelnost a budoucí údržbu řešení.



Obrázek 19 Naznačení datového toku pro ilustrační tabulky (vlastní zpracování)

Všechny zdroje byly opět dokumentovány do původní souboru do sloupce SOURCE_TABLE a využity v další části řešení.

Popis aplikačních dat

Poslední fází v rámci technické části je popis technických elementů – objektů a jejich atributů. Jak již bylo zmíněno ve fázi metadata alignmentu (kapitola 3.3.1), samotná integrace obchodních a aplikačních dat není dostačující k zajištění kvalitní dokumentace, protože vazby nebylo možné vytvořit všude a v některých případech nebyly dostatečně kvalitní.

Popisem technických dat se rozumí příprava definic jednotlivých sloupců a tabulek v datamartu v takové podobě, aby bylo na první pohled zřejmé, který koncept nebo metriku daný atribut či entita zastupuje. Jinými slovy jde o přípravu obchodních definic. Definice byly připraveny v rámci parametru COMMENT v dokumentačním souboru. Pokud šlo o definici, která se váže k pojmu, který zatím neexistuje ve slovníku obchodních pojmů, pak byl vytvořen i název a přidělen unikátní identifikátor. Upravené či nově vytvořené definice ve formátu ID, název, definice, případně synonyma či zkratky byly v další fázi řešení předány týmu orientovanému na řízení dat ve společnosti, který po schválení a doplnění dalších parametrů, jako vlastnictví, obchodní doména atp. navrhl termín jako kandidátní, aby bylo možné ho zařadit do oficiálního schvalovacího procesu a následné formální validaci pro vytvoření dalších vazeb mezi aplikačními a obchodními daty.

Ne všechny technické záznamy jsou však pro obchodní slovník relevantní a nepotřebují nutně svůj protějšek v podobě byznys termínu – například primární a cizí klíče, auditní a systémové sloupce, datumové klíče a technické identifikátory či syntetické klíče. Definice pro tyto atributy byly vytvořeny s pomocí popisu transformační logiky nebo doplněním definic z jádra datové skladu, pokud byly dostupné. Tyto definice byly ponechány pouze na úrovni technických dat.

Vzhledem k tomu že jde v podstatě o obchodní popis, je tato etapa úzce spjata s další částí, která se věnuje přípravě dokumentace na obchodní úrovni. Samotný proces přípravy těchto definic je tedy popsán v další části.

Obchodní metadata

Tvorba definic není triviální záležitostí. Vyžaduje čas, celou řadu odborných znalostí a analytických či dokumentačních dovedností a spolupráci s odborníky z daných oblastí podniku, aby bylo zajištěno, že definice jsou nejen fakticky správně, ale rovněž jsou v souladu s realitou v dané společnosti. Kromě toho by se kvalitní definice měli řídit předem definovanými pravidly, aby bylo zajištěno, že budou srozumitelné a všeobecně přijaté napříč organizací. Nutné znalosti, dovednosti a pravidla pro přípravu kvalitních definic byly rozebrány v kapitole 1.4.3.

Pro zajištění faktické správnosti připravovaných popisů byly použity zdroje zmíněné ve fázi analýzy současného stavu dokumentace (viz kapitola 3.2) – kombinace interních materiálů, pojišťovacích slovníků a internetových materiálů. Podobně jako v případě analýzy i v tomto případě přípravu ztížila přísná bezpečnostní politika, díky které nebylo možné využít celou

řadu zdrojů ani žádných dostupných nástrojů umělé inteligence. Jednalo se o iterativní proces, který probíhal v úzké, pravidelné spolupráci s různými týmy napříč organizací (např. analytici, účetní, pojistní matematici, produktoví manažeři, risk manažeři atp.). Kromě toho pomohl také popis transformační logiky, díky kterému bylo možné lépe porozumět, jak jsou dané koncepty skutečně realizovány a jakým způsobem se s nimi v praxi pracuje, což umožnilo zpřesnit definice dle zvyklostí organizace.

V případě, že byla u technického záznamu vytvořená vazba na obchodní pojem či termín, který byl schválen a jeho definice byla dostatečně kvalitní, bylo možné ji využít. Ani v opačném případě však nebylo vždy nezbytné začínat s definicí od začátku, základem mohli být definice, které nebyly dostatečně detailní nebo kandidátní pojmy, které byly relevantní. Jejich kvalitu však bylo nutné pečlivě ověřit. Příprava definic pro tabulky byla zpravidla jednodušší než u jednotlivých atributů, protože tabulky představují logické celky odpovídající konkrétním oblastem pojišťovnictví, jako jsou například Premium, Claim, Policy nebo Risk. Každá tabulka tedy většinou reflektuje určitý obchodní koncept, který je možné jako celek popsat a propojit s příslušným byznys termínem. Tyto termíny už mimo jiné měli existující definice, které bylo ve většině případů možné použít.

Tabulka 4 ukazuje příklad atributu, který má vyplněné všechny parametry. Pro ilustraci byl opět použit příklad fiktivních tabulek prezentovaný při popisu transformační logiky (viz Výpis 1). Z tabulky je patrný rozdíl mezi obchodní definicí a popisem transformačního pravidla. Definice popisuje, k čemu daný sloupec slouží, zatímco transformační pravidlo uvádí způsob jeho kalkulace. Definice je v tabulce dvakrát – poslední čtyři sloupce představují sekci „BUSINESS DATA“, která byla využita ve fázi metadata alignmentu. Nové pojmy či termíny, které vznikly v této fázi byly označeny statusem New pro jasnou identifikaci. Všechny nové záznamy byly později předány data management týmu (viz kapitola 3.4).

Tabulka 4 Ukázka kompletně zdokumentovaného záznamu

Parametr	Data
ATTRIBUTE_NAME	risk_flag
TABLE_NAME	fct_claims_kpis
COMMENT	A classification that indicates whether the level of claims activity for a given product in a specific time period is within an expected range. It highlights products or periods where the volume or value of claims may signal elevated risk. This flag helps the business identify unusual patterns in claims behavior and supports decision-making in areas such as risk management, underwriting, and product performance monitoring.
DATA_TYPE	varchar

PRECISION	10
SCALE	
COLUMN_ORDER	6
IS_NULLABLE	True
IS_PRIMARY	False
TRANSFORMATION_RULE	This column represents a business rule that classifies the risk level of a given product and month based on the average claim amount (avg_claim_amount). If the average amount of claims for a product in a given month exceeds the predefined threshold (threshold_amount) from the reference table, the product-month combination is flagged as HIGH risk. Otherwise, the risk level is marked as NORMAL.“
SOURCE_TABLE	dw_claims.claim_amount, ref_product_risk_thresholds.threshold_amount
TERM_REFERENCE	2617#Risk Indicator
ID	2617
Name	Risk Indicator
Definition	A classification that indicates whether the level of claims activity for a given product in a specific time period is within an expected range. It highlights products or periods where the volume or value of claims may signal elevated risk. This flag helps the business identify unusual patterns in claims behavior and supports decision-making in areas such as risk management, underwriting, and product performance monitoring.
Status	New

V návrhu dokumentace byly připraveny definice, transformační pravidla a další parametry pro všechny entity a atributy v datamartu, což představuje přibližně tři až čtyři stovky jednotlivých položek, které bylo nutné postupně projít a zdokumentovat. Výstupem všech popsaných předchozích kroků byl komplexní dokument, který obsahoval nejen samotnou technickou dokumentaci datamartu EPM, ale také definice a vazby na obchodní pojmy a termíny. Tento dokument tak představoval úplnou datovou mapu, která sjednocovala technický a obchodní pohled na data.

Na závěr byl vytvořen nový objekt, který byl pojmenován EPM OLAP a reprezentuje skutečný stav OLAP kostky. Pro objekt nebyly vytvořeny žádné sloupce, ty byly přidány ve fázi implementace. Tento objekt sice nebude vytvořen ve formě tabulky ve skutečné databázi, slouží však pro dokumentační účely. Na úrovni objektu byla připravena vazba na EPM_OLAP, což je již existující validovaný obchodní pojem typu report, který popisuje jeho význam a způsob, jakým k němu lze přistupovat.

Na základě tohoto výstupu bylo možné přistoupit k poslední fázi řešení – implementaci dokumentace do prostředí společnosti a synchronizaci metadat s centrálním úložištěm metadat – umožnění publikace připravených informací do metadata portálu Collibra, kde budou dostupné široké skupině uživatelů byznys a datových slovníků a umožní naplno využít přínos kvalitní dokumentace v každodenní práci s daty.

3.4 Implementace návrhu

Poslední etapou v rámci implementace byznys a datových slovníků bylo přenesení připraveného obsahu, který je výsledkem všech předchozích fází, do prostředí společnosti BNP Paribas Cardif a zajistit jeho dostupnost v rámci Datapedie. Tato fáze spočívala v převodu obsahu dokumentu do struktury odpovídající nástrojům, které firma používá pro správu datových aktiv, a jeho synchronizaci s reálným prostředím. Implementace byla klíčová pro zajištění využitelnosti celé připravené dokumentace – teprve po jejím nasazení se informace staly dostupné širokému spektru uživatelů v rámci organizace.

V této části jsou popsány jednotlivé kroky provedené při aplikaci dokumentace do systému a shrnuty výsledky této aktivity. Aby byla zajištěna konzistence v dokumentaci dostupné v datapedii a jejích zdrojích, je nutné změny provádět přímo ve zdrojích a následně nahrát do centrálního úložiště metadat, neboť v době řešení neexistovala zpětná synchronizace z datapedie do jednotlivých systémů, takže nebylo možné změny provést prostřednictvím Collibra a aplikovat do odpovídacích systémů. V praxi to znamená, že dokumentaci je třeba aplikovat přímo uvnitř EPM datamartu a RMS a poté publikovat do Datapedie.

3.4.1 Datový slovník

Provedením změn v EPM datamartu se rozumí úprava datového modelu tak, aby odpovídal požadovanému stavu v navržené dokumentaci. Jinými slovy, dostat datamart do cílového stavu, který reflektuje všechny provedené změny a zajišťuje konzistenci mezi technickým řešením a připraveným návrhem. Teprve takto upravený model mohl být následně nahrán do databáze Datapedie, čímž bylo zajištěno, že dokumentace v nástroji Collibra odpovídá realitě a zároveň že samotný datamart reflektuje požadavky ze strany byznysu i data governance.

Datové modely jsou v BNPP Cardif uloženy a spravovány v nástroji WhereScape 3D (viz kapitola 3.1.1). Existují dva způsoby, jakými lze provést adjustaci datového modelu:

- a) Manuálně – ruční úprava jednotlivých entit a atributů postupným procházením položek ve vytvořené dokumentaci.

- b) Importem – využití importní šablony z RMS a metadata loader komponenty ve WS 3D.

Přístupovat k takto obsáhlým úpravám manuálně je riskantní, protože se tím otvírají příležitosti k chybám plynoucím z nepozornosti. Zároveň by taková aktivita byla extrémně časově náročná a tím pádem jde o neefektivní způsob práce. Z tohoto důvodu byl zvolen efektivnější přístup v podobě importu.

Import datového modelu

Import metadat probíhal pomocí importních šablon, které byly připraveny v systému pro správu kmenových dat RMS a zpřístupněny prostřednictvím připojení v Microsoft Excel. Existovaly dvě oddělené šablony – jedna pro tabulky a druhá pro jejich sloupce.

Před samotným importem bylo třeba převést původní dokument do stavu, který odpovídal šablonám. Především šlo o přizpůsobení pořadí sloupců šabloně. Po naformátování bylo možné přistoupit k naplnění šablon. Import entit probíhal jako první. Každý záznam v šabloně musí disponovat dvěma dodatečnými parametry – unikátní ordinální identifikátor a název položky která se importuje. Pro název položky lze použít původní názvy objektů, ID se generuje v šabloně automaticky.

Unikátní identifikátor tabulky byl klíčovým údajem pro následné napojení sloupců. Podobně jako v šabloně pro tabulky, byly i zde přidány 2 parametry – ID tabulky z předchozí šablony a název položky, který byl vytvořen ve formátu <název tabulky>-<název sloupce>. Po naplnění obou šablon bylo možné připravené změny publikovat pomocí Excel formuláře do systému RMS.

Po úspěšné publikaci metadat do RMS bylo možné přistoupit k jejich importu do prostředí WhereScape 3D pomocí komponenty Metadata Loader. Tato komponenta slouží k automatizaci procesu načítání metadat z externích zdrojů do prostředí WhereScape 3D. Tento proces umožňuje uživatelům naplnit modely požadovanými datovými strukturami a sladit technická metadata s obchodním kontextem pro účely řízení a návrhu. Metadata loader používá předdefinované šablony k mapování zdrojových metadat na vnitřní strukturu WS 3D. Externím zdrojem metadat bylo v tomto případě právě prostředí RMS (Microsoft Master Data Services), ve kterém byla metadata připravena dle importních šablon popsanych v předchozím kroku.

V rámci komponenty Metadata Loader byl nejprve vytvořen nový objekt typu template (šablona). Tento template slouží jako základ pro objevování dat ze zdrojových systémů (tzv. data discovery), tedy proces vyhledání a načtení struktury metadat ze zvoleného zdroje. Vytvořený template vycházel z obecné šablony připravené pro tyto účely, která je namapována přímo na importní šablony v RMS. Po jeho definování proběhla fáze data discovery, během níž WS 3D načtl všechna dostupná metadata z importních šablon a vytvořil jejich reprezentaci v interní struktuře – konkrétně ve formě nového modelu obsahujícího všechny tabulky a jejich atributy podle připraveného návrhu dokumentace. Tímto způsobem byl importován nový datový model, který disponoval všemi potřebnými parametry, které byly definovány v předchozích částech řešení.

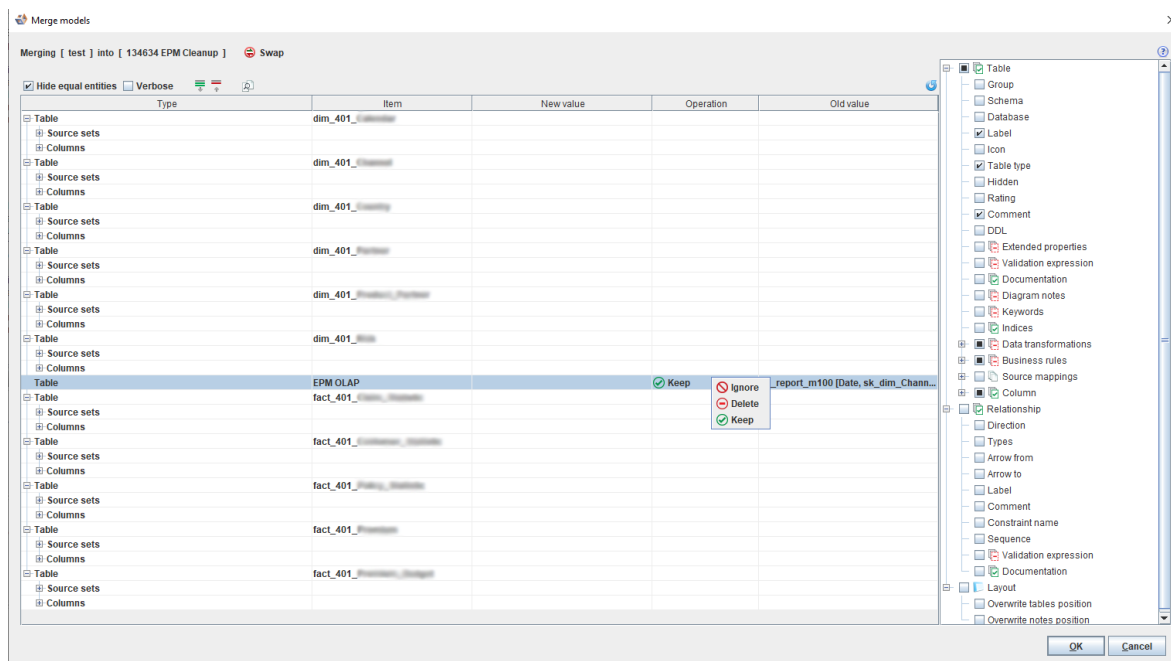
The diagram illustrates the data flow between three components:

- Microsoft Excel**: Represented by the Excel logo.
- Master Data Services (RMS)**: Contains two templates:
 - Import template - entities
 - Import template - attributes
- WhereScape 3D**: Contains a **Metadata loader template** which leads to a **merge models** step, resulting in the **401_EPM** model.

The connections are as follows:

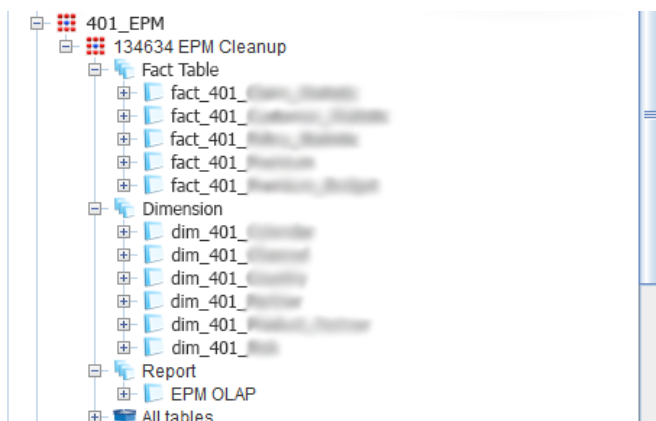
- A dashed arrow labeled **connection** points from Microsoft Excel to Master Data Services (RMS).
- A solid arrow labeled **publish changes** points from Master Data Services (RMS) to Microsoft Excel.
- A dashed arrow labeled **data discovery** points from Master Data Services (RMS) to WhereScape 3D.
- A solid arrow labeled **load metadata** points from Master Data Services (RMS) to WhereScape 3D.

Formulář zobrazený při slučování modelů včetně parametrů, u kterých byly sledovány změny je k vidění na obrázku 21. Při slučování byly filtrovány pouze změny, které byly předmětem řešení, zbylé možnosti nebyly uvažovány. Aby bylo zajištěno, že všechny změny budou provedeny dle očekávání, bylo nutné projít celý formulář a vyznačit u každé položky operaci, která má být provedena. Formulář se od začátku pokusí tyto operace odhadnout, ne vždy však jde o požadovaný stav, například pokud jde o smazání irelevantních atributů, je třeba to explicitně vyznačit.



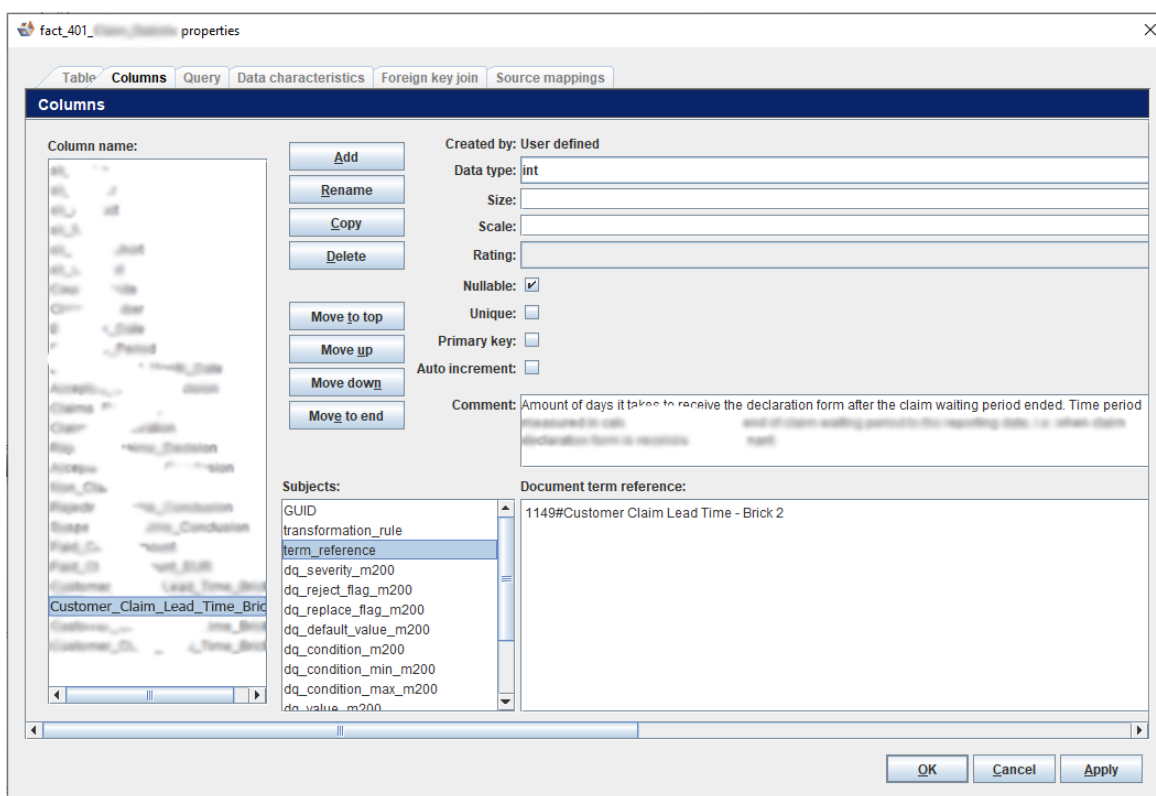
Po definování očekávaných operací pro všechny „konflikty“ bylo možné modely sloučit a tím aplikovat nový datový model dle připravených parametrů. Ukázka nového datového modelu je k vidění na obrázku 22, skutečné názvy jsou cenzurovány, z obrázku je však viditelné, že

objekty byly rozděleny do logických skupin pro zajištění přehlednosti a mimo jiné je zde vidět nově vytvořený objekt EPM OLAP. Nový model obsahuje veškeré definice, popis transformačních pravidel, splňuje nároky na jmenné konvence a disponuje logickým pořadím atributů, které je konzistentní napříč objekty. Kromě toho jsou vyznačeny obchodní vazby tam, kde bylo možné nalézt protějšek a datové typy jsou konzistentně zvoleny a optimalizovány. Vazba na byznys data, vyplněná obchodní definice a další parametry pro vybraný sloupec jsou k vidění na obrázku 23. Některé informace byly z bezpečnostních důvodů cenzurovány. Transformační logika je ve WS 3D zobrazena totožným způsobem jako reference na byznys data – v sekci „Subjects“.



Obrázek 22 WhereScape 3D – Nový datový model (vlastní zpracování)

Na úrovni tabulek byl uveden navíc ještě parametr table type, který určuje, jak jsou tabulky strukturovány a zpracovávány při převodu a exportu modelu.



Obrázek 23 WhereScape 3D – Zobrazení parametrů vybraného sloupce (vlastní zpracování)

Vazba na byznys data v obrázku 23 se skládá ze dvou částí – unikátního identifikátoru a názvu pojmu či termínu. Název slouží spíše pro referenci a na skutečné vytvoření vazby nemá vliv. Unikátní identifikátor je odkazem na obchodní pojem v byznys slovníku. Při synchronizaci dat do datapedie je zkontrolováno ID a pokud je nalezena shoda s byznys pojmem či termínem, je vytvořena vazba.

Poslední aktivitou, kterou bylo v novém datovém modelu nutné podstoupit bylo mapování zdrojů dat, aby bylo možné sledovat datové toky.

Mapování datových zdrojů

V rámci navržené dokumentace byl pro každý atribut v datamartu definován parametr `SOURCE_TABLE`, který popisuje původ daného datového prvku – tedy zdrojovou tabulku nebo výčet objektů použitých při jeho výpočtu (viz kapitola 3.3.2 – Identifikace zdrojů). Tento parametr slouží jako podklad pro vytvoření tzv. source mappings (mapování zdrojů), které jsou klíčovým prvkem pro pochopení datové lineage, tedy sledovatelnosti dat od jejich vzniku až po jejich využití.

Na rozdíl od ostatních parametrů, které bylo možné do WhereScape 3D importovat automatizovaně prostřednictvím šablon, zdroje dat nelze přenést tímto způsobem, protože se nejedná o standardní pole v rámci modelu. Vazby mezi zdrojovými a cílovými objekty se ve WhereScape 3D definují manuálně pomocí grafického rozhraní. Z tohoto důvodu nebylo možné celý proces mapování datových zdrojů zautomatizovat, a proto bylo nutné vytvořit jednotlivé vazby ručně.

Tato činnost znamenala pečlivý průchod jednotlivých objektů v novém datovém modelu a postupné vytváření vazeb mezi zdrojovými a cílovými sloupci na základě připravené dokumentace. Mapování ve WhereScape 3D probíhá v záložce „Source mappings“ u každé tabulky, kde je nejprve nutné přidat všechny zdrojové tabulky a pak propojit zdrojové sloupce s cílovými sloupci v EPM datamartu.

V případě faktových tabulek byla situace často složitější – výpočet metrik zpravidla vychází z více zdrojových sloupců, pocházejících z různých tabulek. Oproti tomu u dimenzionálních tabulek bylo mapování výrazně jednodušší. Většina dimenzionálních atributů není transformována a je přebírána z jádra datového skladu v nezměněné podobě. V těchto případech tedy šlo o přímé přiřazení 1:1 mezi zdrojovým a cílovým sloupcem (viz Obrázek 24).

Výsledkem této aktivity bylo, že každý atribut v datamartu EPM má nyní určený svůj zdrojový sloupec – ať už šlo o přímý převod nebo výpočet z více hodnot. Tento krok byl zásadní pro zajištění plné datové lineage. Byznys uživatelům tak pomáhá vybudovat důvěru v data, urychluje rozhodování a pomáhá s dodržováním předpisů tím, že ukazuje, odkud data pocházejí a jak jsou používána. Naopak technickým uživatelům pomáhá sledovat problémy s daty, vyhodnocovat dopad změn a zlepšuje spolupráci a dokumentaci tím, že zobrazuje celý tok dat.

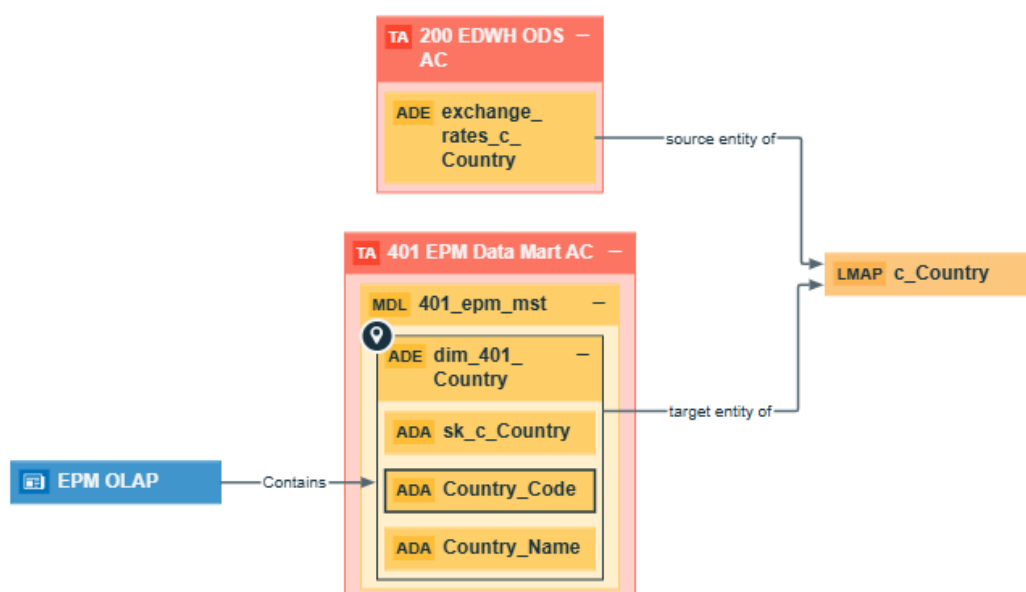
Mapování zdrojů dat bylo poslední aktivitou, která byla provedena v rámci implementace datového slovníku. Aby bylo možné čerpat z nově připravené dokumentace i v datapedii, je nutné nahrát nový datový model do databáze datapedie.

Synchronizace s Datapedií

Synchronizace metadat probíhá na bázi pravidelného přehrávání metadat z WhereScape 3D do databáze datapedie, která slouží jako technický základ pro data governance platformu Collibra. Tento proces zajišťuje, že dokumentace vytvořená v návrhovém a vývojovém prostředí je následně dostupná i koncovým uživatelům prostřednictvím jednotného rozhraní.

Samotný mechanismus synchronizace je již ve společnosti zaveden a je plně v gesci interního datového vývojového týmu, a proto není předmětem této práce.

Po dokončení synchronizace je možné procházet výsledný datový slovník v prostředí Datapedie. Na obrázku 26 je vyznačen datový tok vybraného atributu v nástroji Collibra.



Obrázek 26 Zobrazení datového toku vybraného atributu v nástroji Collibra (vlastní zpracování)

3.4.2 Byznys slovník

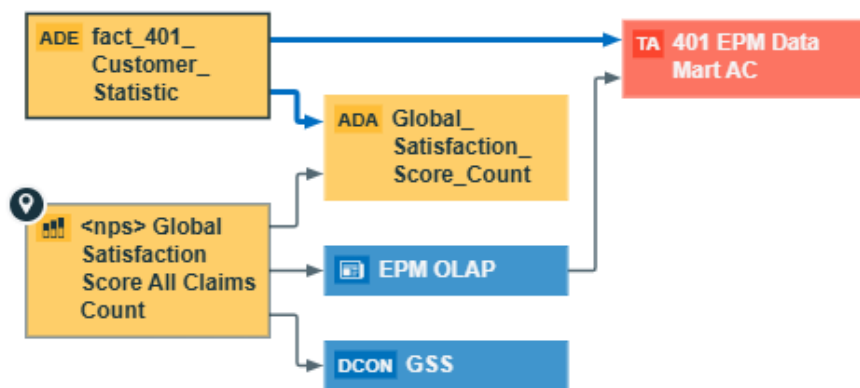
Na rozdíl od technického slovníku nebyla implementace byznys slovníku v rozsahu mého přímého působení. Ve fázi návrhu byly připraveny nové nebo upravené definice obchodních dat a termínů, které byly v rámci dokumentace označeny statusem New. Tyto definice byly zároveň použity při popisu aplikačních dat po schválení byznys uživateli a data management týmem.

Pro účely implementace byly všechny nově vytvořené či upravené záznamy soustředěny do samostatného dokumentu. Ten obsahoval název pojmu, jeho identifikační kód, definici, status (New), případně zkratky a synonyma. Tento dokument byl následně předán týmu zodpovědnému za řízení dat ve společnosti, který doplnil zbývající parametry, jako je příslušná byznys doména, vlastník (ownership) a další, a vložil tyto záznamy do systému RMS.

Obchodní data a termíny spravované v RMS jsou stejně jako technická metadata pravidelně synchronizována do databáze Datapedie a následně publikována v nástroji Collibra. Všechny návrhy jsou na počátku svého životního cyklu označeny statusem Candidate. Aby mohly být považovány za věrohodné (Validated) a bylo možné je využít k vytváření vazeb na technická data, musí projít formálním schvalovacím procesem.

Tímto předáním dokumentace přebírá odpovědnost za další postup data management tým. Z pohledu mého řešení tím implementace business slovníku končí. Jakmile jsou termíny schváleny, lze je dále využívat pro rozšiřování vazeb a tím posouvat úroveň dokumentace dále. Tato průběžná validace a doplňování je důležitým prvkem kontinuálního zlepšování kvality metadat a dokumentace.

Ukázka vazeb mezi aplikačními a byznys daty v nástroji Collibra je dostupná na obrázku 27. Vybranou metrikou (ikona sloupcového grafu) je v tomto případě <nps> Global Satisfaction Score All Claims Count, jejíž zdrojem je atribut (ADA) Global_Satisfaction_Score_Count. Atribut pochází z entity (ADE) fact_401_Customer_Statistic. Metrika je spojená s byznys pojmem (DCON) GSS a je součástí reportu (ikona dashboardu) EPM OLAP. Faktová tabulka i report EPM OLAP jsou součástí aplikace (TA) 401 EPM Data Mart AC.



Obrázek 27 Zobrazení vazeb mezi aplikačními a byznys daty v nástroji Collibra (vlastní zpracování)

Tímto byla dokončena poslední etapa praktické implementace obchodního a datového slovníku ve společnosti BNP Paribas Cardif. V rámci jednotlivých fází byly provedeny všechny klíčové kroky od analýzy současného stavu, přes návrh řešení až po jeho zavedení do praxe. V následující kapitole jsou shrnuty hlavní výsledky realizovaného řešení a navrženy doporučené kroky pro další rozvoj v oblasti správy metadat a datové dokumentace.

3.5 Zhodnocení přínosů a doporučení pro další rozvoj

Realizované aktivity v rámci této práce vedly k významnému posunu ve správě dat a metadat ve společnosti BNP Paribas Cardif. I přes existující procesy a zavedené nástroje pro správu metadat, nebylo data governance dosud plně implementováno a jeho využití v praxi bylo omezené. Navržené a realizované řešení přineslo eminentní zlepšení kvality dokumentace a vytvořilo pevný základ pro další rozvoj.

3.5.1 Přínosy

Klíčovým přínosem bylo vytvoření komplexní dokumentace pro EPM datamart, která umožnila propojení obchodních a technických dat, zvýšila srozumitelnost datových struktur a posílila důvěru v data napříč organizací. Díky nové dokumentaci je nyní možné jasně identifikovat obchodní význam jednotlivých datových prvků, jejich zdroj a transformační logiku. To přispívá nejen k lepší orientaci technických i byznys uživatelů, ale rovněž podporuje spolupráci napříč týmy.

Zároveň byly navrženy a zčásti implementovány nové obchodní pojmy a termíny, jejichž dokumentace byla předána datovému týmu k dalšímu zpracování a schválení. Tím byl nastartován důležitý proces postupného rozšiřování a zkvalitňování byznys slovníku, který v budoucnu umožní další propojování datových entit a hlubší porozumění firemním datům.

Výstupy řešení lze shrnout takto:

- Všechny datové prvky v EPM datamartu odpovídají high-level architektuře v rámci požadavků na jmenné konvence, datové typy a pořadí atributů.
- Každý relevantní element disponuje obchodním popisem.
- Každý atribut má jasně definovanou transformační logiku a je namapován na zdrojové sloupce, které jsou použity pro jeho definici.
- Většina záznamů má vytvořenou vazbu na obchodní data.
- Jsou navrženy nové obchodní definice, které dále přispějí ke kvalitě byznys slovníku a integraci mezi obchodními a aplikačními daty.

Význam těchto výsledků spočívá především v tom, že představují první skutečný krok k vytvoření použitelného rámce data governance ve společnosti. Ačkoliv byly v minulosti podniknuty dílčí kroky k zavedení nástrojů a procesů pro správu dat, jejich využití bylo omezené a dokumentace často nesloužila jako spolehlivý nástroj pro orientaci v datech. Navržené řešení tento stav zásadním způsobem posunulo – vytvořená dokumentace není pouze technickým výpisem, ale propojuje obchodní význam s technickou implementací, čímž vytváří důvěryhodný zdroj informací využitelný napříč týmy.

Tím, že řešení ukázalo praktickou cestu, jak lze dokumentaci systematicky vylepšovat a jak z ní mohou uživatelé benefitovat v každodenní práci, poskytuje zároveň silný argument pro rozšíření podobného přístupu do dalších oblastí. Může sloužit jako „proof of concept“, který dokazuje, že zlepšení je možné, realizovatelné a přináší konkrétní výsledky. Práce tedy vytváří nejen konkrétní zlepšení v oblasti EPM datamartu, ale i opakovatelný vzor pro další datamarty, systémy nebo dokonce jiné geografické regiony, kde společnost působí.

Významným přínosem je i to, že vytvořená dokumentace zlepšuje komunikaci mezi technickými a obchodními rolemi, což je zásadní pro dlouhodobý rozvoj datové kultury ve firmě. Pokud se dokumentace stane běžnou součástí práce s daty, zvyšuje se šance, že se zaměstnanci do její tvorby a údržby budou aktivně zapojovat. V tomto směru tedy práce neřeší jen konkrétní problém, ale zároveň pomáhá budovat prostředí, ve kterém je správe dat věnována patřičná pozornost – což je základní předpoklad pro úspěšnou implementaci data governance.

3.5.2 Překážky a omezení

Přestože se podařilo dosáhnout významného pokroku, naplnění tohoto cíle se neobešlo bez překážek. V rámci řešení se vyskytla některá omezení a překážky, které bylo nutné odstranit či eliminovat.

V rámci omezení jde především o integraci aplikačních a obchodních dat. Ne všechny atributy bylo možné provázat s obchodními pojmy, zejména kvůli absenci validovaných definic ve slovníku. Kromě toho nebylo možné napojit byznys termíny a entity v Collibře, protože v době řešení nebyla vyvinuta tato logika. Je však vhodné zmínit, že i v případě, že by logika pro napojení byla připravena, i přesto by byl problém integraci vytvořit, což je důsledkem špatně definovaných identifikátorů v entitě byznys termínů. Jak bylo zjištěno během fáze analýzy, existují dvě rozdílné entity – byznys data a byznys termíny a obě entity mají stejný formát ID, které se vzájemně překrývají a nejsou vůči sobě unikátní. Vzhledem k tomu, že propojení funguje pouze na základě identifikátorů, které vůči sobě nejsou jedinečné tak obchodní termíny na úrovni tabulek by byly zaměněny za obchodní pojmy. Požadavek na vývoj potřebné logiky byl eskalován vývojovému týmu, naopak návrh na vytvoření nového formátu identifikátorů pro obchodní termíny byl předán týmu řízení dat.

Tyto skutečnosti potvrzují klíčový aspekt data governance programu, který byl několikrát zdůrazněn v této práci – správa dat je neustále se vyvíjející proces, který vyžaduje pravidelnou údržbu, aktivní správu a spolupráci napříč organizací. Nestačí pouze vytvořit počáteční dokumentaci, klíčové je zajistit její neustálou aktualizaci a pravidelné rozšiřování obsahu. To s sebou přináší potřebu kontinuálních investic – jak časových, tak personálních.

Výzvy, které v průběhu řešení vyvstaly, byly především kombinací lidské a procesní složky DG. Zaměstnanci nebyli motivováni se na správě dat podílet, ať už z důvodu znalostních nároků nebo z různých sociologických důvodů (viz kapitola 2). Na této neochotě měla rovněž podíl právě procesní komponenta. Přísné nároky na ochranu dat vytvořili nepříznivou kulturu pro sdílení informací a zároveň výrazně ztížily možnost konzumace externích zdrojů. Kromě toho, složitý proces schvalování navržených obchodních pojmů a

termínů zapříčinil, že zaměstnanci ztratili zájem o navrhování a vylepšování dokumentace, protože proces byl velmi zdlouhavý a často bylo náročné definice prosadit.

Nabytí potřebných odborných znalostí a seznámení s firemní kulturou bylo nezbytnou prerekvizitou pro řešení diplomové práce. Vyřešení těchto překážek znamenalo proaktivní přístup ke studiu dostupných interních i externích materiálů a celou řadu formálních i neformálních rozhovorů s odborníky ve společnosti (viz kapitola 3.2).

Odstranění lidských a procesních bariér je dlouhodobý proces, který vyžaduje kombinaci systematického vedení, kultivace firemního prostředí a vhodných nástrojů. Tyto překážky nelze vyřešit jednorázovým opatřením nebo technickým řešením – vyžadují aktivní zapojení vedení společnosti, rozvoj datové kultury a vytvoření prostředí, kde je dokumentace chápána jako přirozená součást práce s daty, nikoliv jako obtěžující povinnost.

Z pohledu procesů se jedná především o otázky optimalizace a zjednodušení interních schvalovacích mechanismů. Přestože se tyto aspekty nacházejí mimo rámec této práce, je zřejmé, že jejich zefektivnění by přineslo zásadní posun vpřed.

V oblasti lidského faktoru je klíčové budovat datovou kulturu prostřednictvím pravidelné edukace, motivace a srozumitelné komunikace. Firma musí aktivně pracovat na tom, aby zaměstnanci chápali hodnotu kvalitní dokumentace – nejen pro organizaci jako celek, ale i pro jejich vlastní práci. Když zaměstnanci uvidí přínosy v podobě snadnější orientace v datech, efektivnější analýzy nebo snížení chyb, budou ochotnější se na dokumentaci podílet. Řešení prezentované v této práci k tomu přispívá – zpřístupněním dobře dokumentovaného EPM datamartu vznikl funkční příklad, jak by kvalitní dokumentace mohla vypadat, což může motivovat k rozšíření podobného přístupu do dalších oblastí.

3.5.3 AI jako nástroj pro překlenutí lidských bariér

Jedním ze způsobů, jak redukovat bariéry způsobené lidskou komponentou data governance, je využití nástrojů umělé inteligence (dále také „AI“). Během řešení práce se ukázalo, že mnoho časově náročných činností – jako analýza struktury, pochopení domény, návrh vazeb, popis transformačních pravidel nebo obchodních definic – by mohlo být alespoň částečně automatizováno nebo podpořeno za pomoci AI nástrojů.

Nástroje umělé inteligence mohou pomoci především v těchto oblastech:

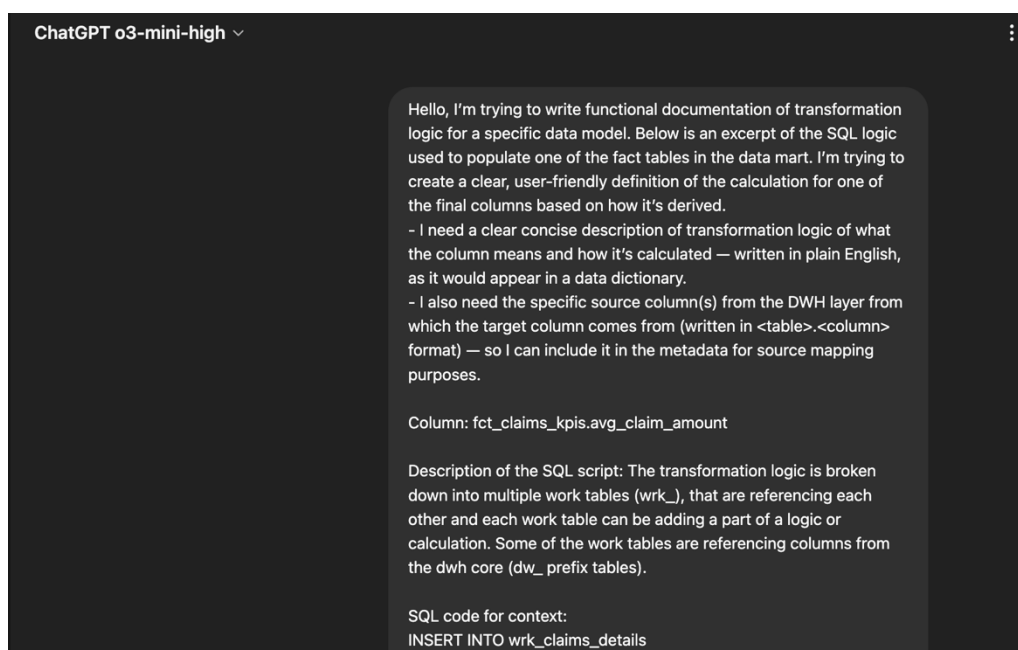
- porozumění technickému obsahu,
- analýza stavu dokumentace,
- generování návrhů definic,
- zlepšení přístupnosti a porozumění.

Porozumění technickému obsahu

V oblasti popisu transformační logiky a nalezení zdrojů pro dané sloupce může AI hrát zásadní roli. Jak bylo prezentováno v kapitole 3.3.2, překlad technického kódu do běžného jazyka, kterému rozumí běžní uživatelé, a nalezení a zaznamenání všech zdrojů je velmi časově náročný proces, který vyžaduje kombinaci znalostí vývojových rámců a odborné

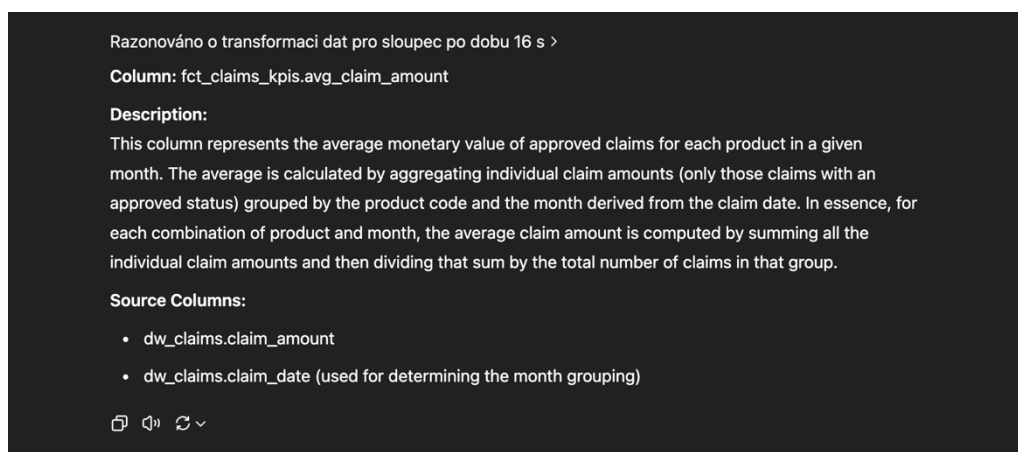
domény. Moderní AI nástroje dokáží tento proces zásadně usnadnit. Důvodem je jejich schopnost „porozumět“ celé řadě programovacích jazyků, včetně SQL, a zároveň schopnost vysvětlovat pojmy a struktury ve srozumitelném a kontextově přizpůsobeném jazyce. Modely byly trénovány na rozsáhlých množstvích textů – včetně technické dokumentace a programovém kódu – což jim umožňuje velmi přesně interpretovat funkci jednotlivých výpočtů a následně ji popsat srozumitelně i pro netechnické uživatele.

Pro ilustraci byl použit model ChatGPT o3-mini-high, který byl seznámen s výpisem kódu, použitým jako příklad při popisu transformační logiky v kapitole 3.3.2 (viz Výpis 1). Obrázek 28 ukazuje použitý prompt, který nejprve popisuje problematiku a požadavky, poté jmenuje konkrétní sloupec zájmu, následně popisuje logiku skriptu a na závěr uvádí konkrétní výpis kódu (výpis kódu je oříznutý pro lepší přehlednost).



Obrázek 28 Prompt pro popis transformační logiky vybraného sloupce (vlastní zpracování)

Na obrázku 29 je zřetelná vygenerovaná odpověď, která shrnuje transformační logiku a uvádí zdrojové sloupce.



Obrázek 29 Popis transformační logiky vybraného sloupce pomocí ChatGPT (vlastní zpracování)

Výsledky AI modelu je samozřejmě vhodné důkladně zkontrolovat, tento přístup však může výrazně přispět ke snížení bariéry mezi IT a byznys týmy – překlenout tradiční propast mezi tím, kdo data vytváří, a tím, kdo s nimi potřebuje pracovat. AI se tak může stát prostředníkem, který umožní lepší porozumění technickému obsahu a podpoří spolupráci mezi datovými inženýry a byznys uživateli. V praxi, pokud by šlo o firemní model, by bylo možné instrukce uvedené v promptu, model naučit a samotný prompt by mohl obsahovat pouze název sloupce a ukázkou kódu s jednoduchým popisem.

Analýza stavu dokumentace

Další významnou oblastí je fáze *metadata alignmentu*, tedy vytváření vazeb mezi aplikačními a obchodními daty. Tato činnost je nejen časově náročná, ale vyžaduje i značné porozumění doménové oblasti. V této fázi mohou AI nástroje představovat zásadní urychlení práce – díky své schopnosti analyzovat rozsáhlé množství strukturovaných i nestrukturovaných informací.

Pokud má model k dispozici výstupy technické dokumentace (např. reporty aplikačních metadat obsahující názvy tabulek, sloupců, datové typy, transformační logiku) a zároveň odpovídající sadu obchodních metadat (např. popisy KPI, názvy obchodních termínů nebo slovník pojmů), je schopen navrhnout první sadu potenciálních vazeb. Tyto vazby může odvozovat na základě jazykové podobnosti, shod kontextu, datových typů, nebo i na základě zjištěných vzorců mezi různými doménami.

Takový návrh může následně sloužit jako pracovní základ, který datový tým pouze validuje a doplňuje. AI tím přebírá roli asistenta, který zkracuje úvodní analytickou fázi a pomáhá rychleji identifikovat relevantní spojení mezi byznys a IT pohledem. Podobně může pomoci odhalit nekonzistence, chyby v jmenných konvencích nebo datových typech a podobně.

Generování návrhů definic

Nástroje umělé inteligence lze využít pro generování návrhů definic technických a obchodních termínů na základě interní dokumentace, dostupných datových struktur nebo veřejně dostupných slovníků. Přestože výstupy nemusí být vždy perfektní, mohou opět posloužit jako výchozí bod, který pak odborník upraví podle požadavků společnosti. Modelu lze poskytnout pravidla, kterými se řídit při generování definic, způsob jazyka, který má využít a formulovat srozumitelný výstup. To umožňuje snížit bariéry spojené s dokumentačními dovednostmi a odbornými znalostmi.

Pro ukázkou byl použit pojem Average Claim Amount, který vychází z příkladu uvedeného na obrázcích 28 a 29. Konverzace navazuje na předchozí konverzaci týkající se transformační logiky a využívá tedy i stejný model. Prompt použitý k získání definice je znázorněn na obrázku 30. Požadavek by mohl být dále doplněn o specifická pravidla, kterými se model má řídit při přípravě definic (viz kapitola 1.4.3).

Based on your knowledge of the insurance industry and insurance claims and the provided transformation logic of avg_claim_amount, prepare a business-facing definition in plain-english of a term called Average Claim Amount that could be used in a business glossary for all the business users to understand this business concept.

Obrázek 30 Prompt přípravu definice pojmu Average Claim Amount (vlastní zpracování)

Výsledek v podobně obchodní definice průměrné výše pojistného plnění ukazuje je prezentován na obrázku 31.

Term: Average Claim Amount

Business Definition:

The Average Claim Amount represents the typical cost of an insurance claim for a specific product over a given period, usually measured monthly. It shows how much, on average, the company pays out for each approved claim. This helps the business understand trends in claim costs, compare the performance of different insurance products, and identify any unusual spikes that might indicate risks or changes in customer behavior.

How it's calculated:

To calculate the average claim amount, all approved claim payouts for a specific product in a given month are added together, and that total is divided by the number of approved claims in that same product and month group.

Example:

If a product had 5 approved claims in March, with payouts of \$200, \$300, \$250, \$400, and \$350, the average claim amount would be:

$$(\$200 + \$300 + \$250 + \$400 + \$350) \div 5 = \$300$$

Obrázek 31 Obchodní definice pojmu Average Claim Amount pomocí ChatGPT (vlastní zpracování)

Výsledky lze dále zpřesnit, pokud by model měl přístup k interním dokumentům společnosti a mohl by tak využít specifické termíny, zkratky a výrazy, které jsou ve společnosti všeobecně známy a využívány.

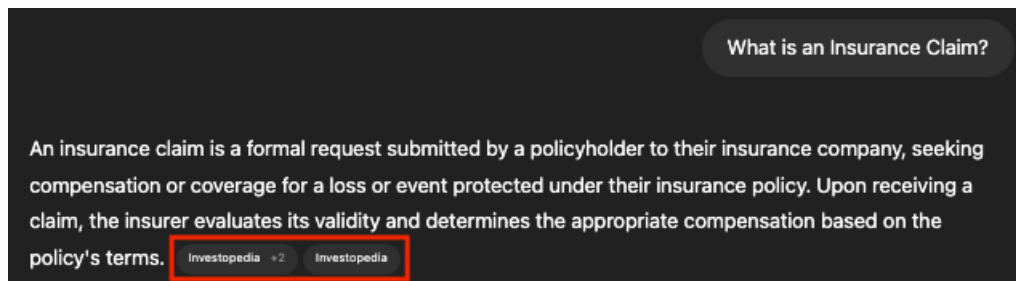
Zlepšení přístupnosti a porozumění

Umělá inteligence se může stát užitečným nástrojem pro rychlou orientaci v nové odborné oblasti, zejména v úvodních fázích projektu, kdy je třeba pochopit základní pojmy, klíčové koncepty a terminologii. Modely typu GPT dokáží srozumitelně vysvětlit i komplexní témata, přeložit technické termíny do běžného jazyka a vytvořit přehled, který uživateli pomůže získat základní povědomí o dané doméně. Tím usnadňují začátečníkům vstup do problematiky a zároveň zvyšují efektivitu i u zkušenějších uživatelů, kteří si potřebují rychle připomenout nebo ověřit související pojmy.

Při využívání AI je však nutné mít na paměti, že tyto modely nejsou neomylné. Zvláště při dotazech zaměřených na konkrétní fakta může docházet k tzv. „halucinacím“ – tedy ke generování smyšlených nebo nepřesných informací. Z tohoto důvodu je důležité ke

každému výstupu přistupovat kriticky a ověřovat jeho správnost z dalších důvěryhodných zdrojů.

Na druhou stranu, moderní jazykové modely často umožňují přímo uvést zdroje, ze kterých při odpovědi čerpají (viz Obrázek 32).



Obrázek 32 Uvedené zdroje v odpovědi modelu ChatGPT

To může významně přispět nejen k ověření správnosti tvrzení, ale také k nalezení dalších relevantních materiálů pro hlubší studium daného tématu.

Jak je z příkladů patrné, využití umělé inteligence při přípravě kvalitní a soudržné dokumentace může sloužit jako významný akcelerátor v oblasti správy metadat. Aby bylo možné AI skutečně efektivně využít, je nutné vytvořit vhodné prostředí. Velké organizace často volí cestu před-trénovaných modelů, které jsou dále doladěny na interních datech a provozovány v bezpečném prostředí, což umožňuje využití odborného firemního jazyka a zajištění souladu s bezpečnostními pravidly. Společnosti, které zvažují využití umělé inteligence pro podporu datové správy, se nemusí hned pouštět do tvorby komplexního a všeobecně trénovaného modelu. V mnoha případech může být efektivnější začít s menšími, cíleně zaměřenými modely nebo specializovanými AI agenty. Tyto tzv. účelové modely jsou navrženy pro konkrétní úlohy a oblast použití (např. specializovaný agent pro SQL nebo model specializovaný na pojmy z pojišťovnictví atp.)

Výhodou těchto menších, úzce specializovaných řešení je jednak vyšší přesnost díky zúženému zaměření, ale také nižší náročnost na implementaci a správu. Tyto modely lze navíc snadno aktualizovat nebo vyměnit podle aktuálních potřeb organizace. Využití účelových modelů tak představuje realistický a škálovatelný přístup k začlenění umělé inteligence do procesu správy dat, bez nutnosti enormních investic či zásahů do infrastruktury.

Je však třeba zmínit i omezení. Přestože AI dokáže nabídnout kvalitní návrhy, výstupy je třeba validovat a interpretovat v kontextu organizace. Pokud má být výstup opravdu přesný a použitelný, může se stát, že kontrola často zabere stejně času jako ruční vytvoření dokumentace – kvalita generovaného obsahu se sice zlepšuje, ale zatím stále neumožňuje plnou autonomii. Kromě toho, je kvalita výsledku daného modelu závislá na kvalitě vstupu od uživatele, což klade na uživatele určité znalostní nároky – např. základy prompt engineeringu, znalost kontextu domény, kritické myšlení a schopnost validace výstupu apod.

Řada bariér ve správě dat je způsobena lidskou složkou – nízká motivace, obavy ze sdílení informací nebo nedostatečné znalosti. AI tuto složku sice v současnosti nemůže plně

nahradit, ale může ji významně podpořit. Pokud uživatel nemusí dokumentaci tvořit „od nuly“, ale dostane připravený návrh, může vstoupit spíše do role editora než autora – což odbourává psychologickou bariéru a snižuje tlak a zvyšuje šance, že se lidé do tvorby zapojí.

3.5.4 Doporučení pro další rozvoj

Realizované aktivity ukázaly, že efektivní správa dat není jednorázový projekt, ale kontinuální proces, který vyžaduje pravidelnou údržbu, vytvoření firemní kultury, kde se zúčastněné strany budou aktivně podílet na budování společného datového prostředí a schopnost přizpůsobovat se měnícím se podmínkám. Taková mentalita je pro úspěch této iniciativy naprosto klíčová a přináší s sebou závazek v podobě nutné investice, a to jak časové, tak personální a finanční. Pro zajištění dlouhodobé udržitelnosti a přínosu vytvořené dokumentace je nutné navázat na započatou práci a rozvíjet ji. Rozvoj lze rozdělit do dvou klíčových oblastí – operativní doporučení a strategická.

Operativní oblast

Doporučeními v rámci operativní oblasti se rozumí především aktivity, které je potřeba přímo vykonat, mají jasné výstupy a dají se jednoduše naplánovat. V zásadě jde o činnosti, které přispějí k dalšímu zvýšení kvality dokumentace. V rámci operativních činností lze doporučit následující:

1. Formální validace a rozvoj byznys slovníků – Provést schválení navržených obchodních termínů, které byly označeny jako kandidátní. Tím se otevře cesta k doplnění chybějících vazeb mezi aplikačními daty a obchodními pojmy v EPM datamartu. Kromě toho je třeba implementovat technickou logiku, která umožní vytvářet a zobrazit tyto vazby v Datapedii.
2. Rozšíření dokumentace do dalších částí datové architektury a dalších regionů – Práce provedená na EPM datamartu představuje výchozí bod pro další rozvoj správy dat. Proto je vhodné rozšířit stejný postup – horizontálně na další datamarty a vertikálně na další vrstvy datového skladu v rámci CE regionu. Následně lze uvažovat o škálování procesu do dalších regionů, čímž bude podpořena unifikace správy dat v rámci celé skupiny.
3. Zavedení mechanismu pro správu a údržbu dokumentace – Zajistit pravidelný cyklus revizí a nastavit metriky, které umožní sledovat aktuálnost a úplnost dokumentace a postupně ji vylepšovat. Zároveň definovat jasnou strukturu odpovědností, tj. kdo má na starosti revize, aktualizace, schvalování a komunikaci změn.

Strategická oblast

Na rozdíl od operativní oblasti, doporučení v rámci strategické složky představují dlouhodobé změny v přístupu, kultuře nebo procesech. Vyžadují vedení, změnu mindsetu a spolupráci.

1. Posílení datové kultury a motivace zaměstnanců – Podporovat otevřenou a spolupracující firemní kulturu, kde je sdílení znalostí vítané a kde jsou přínosy dobře komunikované. Zaměřit se na školení a vzdělávání uživatelů – aby rozuměli

významu správy dat, ale také věděli, jak s dokumentací pracovat a měli potřebné znalosti a dovednosti na kontinuální vylepšování celého programu.

2. Revize a optimalizace schvalovacích a bezpečnostních procesů – Zvážit úpravy rigidních schvalovacích procesů, které zpomalují nebo znemožňují validaci nových definic. Užitečné by mohlo být například rozprostit schvalování podle obchodních domén. Kromě toho je žádoucí přehodnotit některé bezpečnostní požadavky, které brání efektivnímu sdílení a rozvoji znalostí, a hledat cesty, jak zároveň ochránit citlivé informace a podpořit rozvoj.
3. Využití umělé inteligence jako podpůrného nástroje – Jak bylo prezentováno v předchozí kapitole (viz kapitola 3.5.3), využití moderních jazykových modelů může významně podpořit aktivity spojené s implementací kvalitní dokumentace. Z tohoto důvodu by mohlo být výhodné hlouběji prozkoumat možnosti, které umělá inteligence nabízí a zvážit postupné zavádění účelových AI modelů, které mohou být trénovány na interních datech a reflektovat specifické potřeby společnosti.

Zavedením výše uvedených kroků může společnost BNP Paribas Cardif postupně vybudovat udržitelný a efektivní rámec správy dat, který bude nejen v souladu s regulačními požadavky, ale také skutečně přínosný pro každodenní práci s daty. Výsledky této práce ukázaly, že i cílené aktivity na menším rozsahu mohou vést k měřitelnému zlepšení a stát se základem pro škálování data governance napříč organizací. Doporučení formulovaná na základě provedených aktivit tak představují nejen logické pokračování dosavadních kroků, ale zároveň konkrétní směr, jak dosáhnout vyšší datové vyspělosti, větší transparentnosti a lepší využitelnosti dat napříč všemi odděleními firmy.

3.6 Shrnutí kapitoly implementace byznys a datových slovníků

Přestože se podařilo navrhnout komplexní a kvalitní dokumentaci, tím práce na zlepšování správy dat zdaleka nekončí. Metadata, stejně jako samotná data, podléhají změnám – je proto nezbytné zajistit jejich průběžnou aktualizaci, údržbu a validaci. Výsledky dosažené v rámci této práce však mohou sloužit jako praktický základ, případně referenční model, pro další rozvoj správy metadat ve společnosti.

Řešení identifikovalo a popisuje i řadu překážek – zejména v oblasti lidské a procesní složky. Tyto výzvy se nepodařilo v rámci řešení zcela odstranit, protože přesahují rámec jednotlivého projektu a vyžadují širší organizační úsilí. Z pohledu lidského faktoru je nutné dlouhodobě investovat do zvyšování povědomí, edukace, komunikace přínosů a budování kultury spolupráce a otevřenosti. Právě v této oblasti mohou výsledky práce sloužit jako pozitivní impuls: kvalitní dokumentace zvyšuje důvěru v data, usnadňuje práci a může motivovat uživatele k větší participaci. Dále, jak bylo ukázáno v kapitole 3.5.3, mohou být lidské bariéry částečně zmírněny využitím umělé inteligence – například jako podpora při porozumění technickým výpočtům, návrzích definic nebo vysvětlení složitějších konceptů.

Procesní složka – zejména striktní bezpečnostní politika, složitý schvalovací proces nebo rigidní interní pravidla – je výrazně obtížnější na změnu, protože její řešení leží v gesci jiných týmů a úrovní řízení. Adresování těchto nedostatků je tedy mimo rozsah této práce,

nicméně na základě identifikovaných problémů lze doporučit její revizi a zjednodušení tak, aby lépe podporovala rozvoj dokumentace a spolupráci napříč odděleními.

Výsledky tak představují první krok směrem k systematickému a dlouhodobě udržitelnému rozvoji data governance. Zvolený postup může zároveň posloužit jako inspirace pro ostatní organizace čelící podobné výzvě – a to nejen svými výstupy, ale také systematickým způsobem práce, zapojením uživatelů a důrazem na propojení technických a obchodních pohledů.

Závěr

Hlavním cílem diplomové práce bylo analyzovat a zmapovat proces implementace byznys a datových slovníků a veškerá jejich úskalí a nabídnout realizovatelná doporučení, jak efektivněji čelit lidským, technickým a procesním výzvám při zavádění konceptu data governance. Tento cíl byl naplněn prostřednictvím propojení teoretických poznatků s praktickou aplikací ve společnosti BNP Paribas Cardif.

Hlavní cíl byl zároveň strukturován do několika dílčích cílů, jejichž naplnění je shrnuto v následující tabulce (viz Tabulka 5). Všechny definované dílčí cíle byly splněny – od literární rešerše, přes shrnutí současného stavu správy dat ve společnosti, analýzu a návrh řešení, až po implementaci a vyhodnocení přínosů.

Tabulka 5 Přehled naplnění dílčích cílů diplomové práce

Dílčí cíl	Kapitola
Popis všech relevantních oblastí data governance, které jsou klíčové pro realizaci hlavního cíle.	1. Rešerše literatury
Shrnutí současného stavu společnosti a identifikace specifických požadavků na data governance.	2. Současný stav
Dokumentace procesu analýzy, návrhu a implementace byznys a datových slovníků v konkrétní společnosti.	3. Implementace byznys a datových slovníků v BNP Paribas Cardif
Identifikace problémů odhalených při řešení a formulace realizovatelných doporučení.	3. Implementace byznys a datových slovníků v BNP Paribas Cardif

Práce byla rozdělena do teoretické a praktické části. V teoretické části (viz kapitola 1) byla formou rešerše zpracována problematika správy dat, včetně definice klíčových znalostních oblastí a hlavních komponent. Zvláštní pozornost byla věnována byznys a datovým slovníkům a typickým výzvám a nezbytným znalostem spojeným s jejich implementací. Tato část poskytla nezbytný kontext a metodologické ukotvení pro praktické řešení.

Na základě teoretického rámce, interní dokumentace a konzultací s odborníky byl popsán výchozí stav správy dat v BNP Paribas Cardif (viz kapitola 2), včetně hlavních problémových oblastí, které ovlivnily návrh řešení.

Následně byl navržen a realizován konkrétní postup pro zlepšení dokumentace. Praktická část detailně dokumentuje jednotlivé fáze řešení – od seznámení s firemním a doménovým

prostředím, přes návrh a realizaci dokumentace až po nasazení do prostředí společnosti. Zahrnuje také využití metody, identifikaci potřebných kompetencí a práci s uživateli a odbornými týmy (viz kapitoly 3.1 až 3.4).

V závěrečné části praktické implementace byly shrnuty překážky, které se během řešení objevily – zejména v oblasti lidských a procesních bariér – a navrženy konkrétní kroky, jak tyto problémy řešit v rámci dalšího rozvoje data governance (viz kapitoly 3.5 a 3.6). Kromě toho předkládá kapitola 3.5 také stručné zamyšlení nad využitím nástrojů umělé inteligence pro podporu lidské komponenty správy dat.

V diplomové práci byla využita kombinace metod kvalitativního výzkumu a případové studie, doplněná o semistrukturované rozhovory, analýzu dokumentace, pozorování a validaci jednotlivých kroků s uživateli. Nedílnou součástí byl i návrh doporučení pro další rozvoj správy dat a reflexe překážek, které během řešení vyvstaly.

Za hlavní přínos práce lze považovat to, že propojuje teoretický přehled relevantních oblastí s praktickou zkušeností z reálného projektu. Vzniknutý dokument může posloužit jako referenční rámec pro další organizace – ať už ve fázi úvodního porozumění problematice, nebo během samotné implementace. Práce se zároveň věnuje lidským a procesním aspektům, které bývají často podceňovány, přestože představují klíčové faktory úspěchu v oblasti data governance.

Závěrem lze říci, že úspěšná správa dat není jednorázovou iniciativou, ale kontinuálním úsilím. Je to proces, který vyžaduje nejen technická řešení, ale především trvalý zájem a ochotu organizace do oblasti dat investovat – jak finančně, tak lidsky. Výsledná práce tak nabízí nejen konkrétní výstupy, ale i důležitý příklad toho, že i postupnými kroky je možné posunout datovou kulturu směrem k udržitelné a funkční správě dat.

Navzdory tomu, že některé výzvy nebyly v rámci této práce zcela překonány, nelze výsledky hodnotit jako neúspěch. Naopak – právě identifikace konkrétních bariér a úzkých míst je nezbytným krokem na cestě k dlouhodobě udržitelnému data governance programu. Cílem této práce nebylo jednorázově vyřešit všechny aspekty data governance, ale vytvořit pevný a funkční základ, na kterém lze stavět.

Použitá literatura

- 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 Pojiťovně. *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. *EW Solutions* [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

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/top-reasons-why-data-governance-initiatives-fail-generate-hashmi/>

MOUTINHO, Adriana, 2024. Glossary of Business Terms - Challenges and strategies for successful documentation. *LinkedIn* [online] [vid. 2025-02-09]. Dostupné z: <https://www.linkedin.com/pulse/glossary-business-terms-challenges-strategies-adriana-moutinho-5vyrf/>

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

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-slovniku-ve-financni-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=gbs_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