

# ROZDZIAŁ 1

## Architektura danych

### 1.1. Wprowadzenie

Dane, informacja, wiedza i mądrość to bardzo silnie powiązane pojęcia, które pozwalają opisywać i zrozumieć otaczającą nas rzeczywistość. Mogą być one intuicyjnie kojarzone ze światem nauki i prowadzeniem badań naukowych. Badacze dokonują pomiarów, zbierają dane, a następnie je analizują. Naukowcy posiadają wiedzę, która pozwala teoretycznie lub praktycznie zrozumieć określone zagadnienia. Dzięki mądrości potrafią myśleć i działać, wykorzystując wiedzę, doświadczenie, zdrowy rozsądek i przeczucie. Dzielą się danymi i wiedzą, informując różnymi kanałami o swoich odkryciach. Nie są to jednak pojęcia hermetyczne i zamknięte wyłącznie dla świata nauki.

Jeżeli przez dane rozumiemy zbiór wartości ilościowych lub jakościowych, to nawet odkładając na bok ich naukowy lub informatyczny charakter, oznacza to, że w naszym otoczeniu znajduje się mnóstwo danych. Mogą one występować zarówno w postaci analogowej, jak i cyfrowej. Mogą mieć postać ciągu znaków zapisanych na kartce papieru lub sekwencji impulsów elektrycznych wysyłanych przez czujnik zamontowany na ścianie naszego pokoju. Faktura, wiadomość e-mail, lokalizacja nadajnika GPS, numer telefonu w kalendarzu – wszystko to są dane, które można składować, zbierać, przetwarzać i analizować. Dane mogą być zbierane przez indywidualnych ludzi, korporacje, rządy.

Celem rozdziału jest zapoznanie czytelnika z problemem dotyczącym zarządzania danymi, znaczeniem tego zagadnienia dla funkcjonowania i rozwoju organizacji oraz ogólnymi założeniami projektowania architektury danych i architektury korporacyjnej.

## 1.2. Dane i rozwój organizacji

Funkcjonowanie we współczesnym świecie jest związane z koniecznością przetwarzania gigantycznych ilości danych, które nieustannie napływają z naszego otoczenia. Rozwój technologii informacyjnych sprawił, że przez cały czas zwiększa się liczba potencjalnych źródeł danych, różnorodność form transmisji oraz łatwość, z jaką można się przyłączyć do globalnego „potoku informacyjnego”. Lawinowy wzrost ilości danych oznacza również wzmożoną potrzebę ich przetwarzania, identyfikacji i absorbowania lub odrzucania. Wygodny i efektywny dostęp do precyzyjnych danych umożliwia podejmowanie najszybszych decyzji przy jednoczesnym minimalizowaniu negatywnych ich skutków. Dla organizacji, której sukces jest uzależniony od efektywności procesów w niej zachodzących i koordynacji pracy jej uczestników, konieczność adaptacji do zmieniającego się otoczenia i możliwość usprawniania procesów przetwarzania danych może być nie lada problemem.

Istnieje bardzo wiele czynników, które skutecznie utrudniają szeroko rozumianą efektywną pracę z danymi w organizacji i adaptację procesów wykorzystujących dane do zmieniającego się otoczenia organizacji.

Jednym z takich czynników może być wzrost złożoności danych i procesów ich przetwarzania, związany z ewolucją i rozwojem organizacji. W przypadku małego – nawet jednoosobowego – przedsiębiorstwa właściciel posiada bardzo dużą wiedzę na temat swojego biznesu. Może nie wykorzystywać wszystkich szans, ale wie, jak jego firma funkcjonuje, a także zna przyczyny sukcesów i porażek. W sytuacji, kiedy organizacja się rozwija, zadania, które do tej pory wykonywał jeden człowiek, trzeba delegować pomiędzy wielu pracowników. Odpowiedzialność za procesy jest przekazywana do wyspecjalizowanych podmiotów (pracowników lub jednostek organizacyjnych), a wiedza na temat organizacji zostaje rozproszona. W takiej sytuacji może dochodzić do powstawania problemów komunikacyjnych i niewykorzystania maksymalnego potencjału przedsiębiorstwa. Potrzebne dane mogą znajdować się w organizacji, ale mogą być niedostępne, niekompletne, może być problem z ich szybkim odnalezieniem lub mogą być w nieprzydatnej postaci.

Rozwój organizacji i separacja funkcji w organizacji mogą również przejawiać się eskalacją typowego problemu dotyczącego danych – nadmiarowości (*data redundancy*) treści w organizacji. Taka sytuacja ma miejsce, gdy pracownicy uzyskują dostęp do danych z tego samego lub z różnych źródeł i wielokrotnie wprowadzają je do systemu informacyjnego organizacji, np. pracownicy działu marketingu korzystają z syste-

mu CRM (*Customer Relationship Management*) i wprowadzają do niego dane adresowe klientów, a pracownicy odpowiedzialni za sprzedaż utrzymują w niepowiązanym programie służącym do wystawiania faktur oddzielną kopię danych identyfikujących klienta.

Problem z dostępem do danych może być również wynikiem błędnego procesu projektowania systemu informacyjnego organizacji. Projekt i wykonanie komponentów systemu informacyjnego mogą być prowadzone bez pełnej znajomości procesów i bez współpracy z wszystkimi interesariuszami. Można sobie łatwo wyobrazić sytuację, kiedy w organizacji powstaje wiele aplikacji, których tworzenie jest koordynowane przez różne działy. W założeniu aplikacje te powinny dzielić się danymi pomiędzy sobą, ale w wyniku złej współpracy i błędnego projektu do prawidłowej wymiany nie dochodzi. Takie niedziałające systemy nie są rzadkością i nawet posiadają branżowe określenie – „system, który gra i tańczy” (*stovepipe system*).

Z pewnością kwestia adaptacji procesów związanych z przetwarzaniem danych może stanowić kłopot dla starszych organizacji, które są w posiadaniu długodziałających i niemodyfikowanych od dłuższego czasu elementów systemu informacyjnego. Mowa o komponentach, które mogą mieć kilkanaście, a nawet kilkadziesiąt lat, które trwale zagnieżdżyły się w organizacji i z różnych powodów nie mogą być wyeliminowane lub zastąpione. Mogą to być bardzo wyspecjalizowane składniki, jak np. oprogramowanie sterujące pracą maszyn bez dalszego wsparcia producenta lub programy, których aktualizacja jest kosztowna, a sam proces nie przynosiłby wyraźnych korzyści (np. stary system obsługi studentów przechowujący archiwalne dane). W takim wypadku dostęp do danych może być utrudniony (np. opóźniony) ze względów technicznych lub z braku odpowiednich kwalifikacji pracowników organizacji (rozwiązanie jest tak stare, że tylko kilka osób je zna i wspiera).

Kłopotliwe dla organizacji może być utrzymywanie starych składników systemu, ale niefortunnie mogą się też zakończyć ich aktualizacja i rozwój. Technologie informacyjne rozwijają się w ostatnich latach wyjątkowo dynamicznie i menedżerowie dostają do dyspozycji bardzo szeroki wachlarz rozwiązań projektowych, nowych usług, modeli wdrożenia, technik tworzenia aplikacji i rozwiązań sprzętowych. Zróżnicowana gama możliwości i wariantów ich doboru oznacza również znaczny wzrost złożoności systemów i coraz większe wymagania związane z kwalifikacjami po stronie ich projektantów i wykonawców. W przypadku zmiany technologicznej stworzenie lub modernizacja systemu mogą zakończyć się pełnym lub częściowym niepowodzeniem w wyniku złego doboru technologii, nietrafionego projektu architektury

lub np. niewystarczających kompetencji do prawidłowego wykonania systemu. Nowe technologie oferują często obiecujące funkcjonalności, ale ich niedojrzałość może powodować konieczność radzenia sobie na etapie wdrażania z niespodziewanymi problemami, które zostały ujawnione na bardzo późnym etapie projektu (np. późno wykryta niepełna zgodność zaimplementowanych w organizacji mechanizmów uwierzytelniania z obsługiwanymi przez środowisko wykonawcze nowymi aplikacjami; ujawniona w końcowym etapie wdrożenia aplikacji niezadowolająca wydajność i awarie systemu baz danych w pełnym obciążeniu produkcyjnym). Może to, oczywiście, znacząco utrudnić i opóźnić budowę systemu informacyjnego.

Możliwość wystąpienia opisanych wyżej problemów i umiejętność radzenia sobie z nimi jest uzależniona m.in. od etapu rozwoju technologii informacyjnych w organizacji. Richard L. Nolan [Nolan, 1979] zaproponował model teoretyczny opisujący ten proces w sześciu etapach:

- wprowadzenie (*Initiation*),
- rozwój (*Contagion*),
- kontrola (*Control*),
- integracja (*Integration*),
- administracja danymi (*Data administration*),
- dojrzałość (*Maturity*).

W pierwszym etapie modelu Nolana IT wprowadza się do organizacji. Zostają wdrożone pierwsze aplikacje (zwłaszcza redukujące koszty), jak system płacowy, fakturowanie, magazyn, księgowość. Wraz z uruchamianiem aplikacji użytkownicy i menedżerowie zaczynają identyfikować zapotrzebowanie na dodatkowe funkcjonalności, które mają zaspokoić ich potrzeby biznesowe. Dział IT jest mały i wykonuje zleczone zadania.

Drugi etap charakteryzuje się wzmożonym zapotrzebowaniem na nowe aplikacje. Są one tworzone często bez planu i bez odwoływania się do innych aplikacji. Potrzeba zaspokojenia podaży na dane prowadzi do proliferacji powielanych danych i procesów. Etap ten, pomimo wzrostu znaczenia IT, nie oznacza wzrostu jakości rozwiązań – charakteryzuje się brakiem wspólnej wizji i planowania. Wynikiem braku planowania jest konieczność tworzenia nadmiarowych rozwiązań, które pozwalają łączyć i uzupełniać źle przygotowane aplikacje.

W etapie trzecim organizacja odczuwa konieczność wzmożonej kontroli nad IT. Bałagan wynikający z wdrożenia nieefektywnego systemu przekłada się na niezadowolenie jego użytkowników. Próbując zaplanować nad sytuacją, IT mniejszy nacisk kładzie na tworzenie nowego oprogramowania, a kieruje wysiłki w stronę restrukturyzacji istnieją-

cych aplikacji, powołania grupy administrującej bazami danych i sformalizowania procesów projektowania i wdrażania oprogramowania.

Na etapie integracji istniejące aplikacje zostają zmodernizowane, a wykorzystywanie modeli staje się głównym elementem metodyki tworzenia oprogramowania. Użytkownicy otrzymują coraz więcej informacji poprzez dostęp do danych, a przez to zwiększa się ich świadomość, czego i jak powinni oczekiwać od technologii informacyjnych. Dział IT dla zaspokojenia rosnącego zapotrzebowania musi się rozrastać. Nie wyeliminowano jednak wszystkich potencjalnych problemów. Brak analizy danych na poziomie korporacji i występująca przez cały czas redundancja danych uniemożliwia uzyskanie pełnej kontroli nad procesem tworzenia systemu informacyjnego. Na tym etapie osoby odpowiedzialne za rozwój IT w organizacji uświadamiają sobie, jak istotną (kluczową) rolę w procesie modernizacji i unowocześniania procesów pełnią dane. Zmienia to również filozofię podejścia do tworzenia aplikacji z prostego automatyzowania procedur do badania i konsolidowania danych w celu ich przetwarzania.

W etapie piątym jest wdrażane planowanie strategiczne na poziomie organizacji oraz zostaje podkreślona rola zarządzania zasobami informacyjnymi. Dział IT wprowadza zorientowaną na dane (*data-centric*) metodologię *top-down* do projektowania systemu informacyjnego, która zostaje oparta na uznanych i stabilnych modelach. Zostają wprowadzone znaczne usprawnienia w istniejącym systemie, a posiadane aplikacje zostają całkowicie zmodernizowane.

W etapie szóstym organizacja posiada wdrożone procesy analizy i modelowania danych na poziomie organizacji, uruchomione aplikacje odzwierciedlają procesy w niej zachodzące, a struktura organizacyjna zostaje dostosowana tak, by usprawnić proces projektowania architektury.

### 1.3. Architektura korporacyjna i architektura danych

Działy IT są tylko jednym z elementów struktury organizacyjnej typowej organizacji, ale ze względu na dostarczane funkcjonalności zazwyczaj są elementem niezwykle istotnym. Poprawne zaprojektowanie systemu informacyjnego wymaga nie tylko odpowiedniej wiedzy informatycznej,

ale również ścisłej współpracy z różnymi przedstawicielami organizacji, zapewnienia odpowiednich kanałów komunikacyjnych i zdolności przedstawienia odpowiednich celów. Wymagana jest jak najpełniejsza wiedza na temat organizacji i procesów w niej zachodzących. Konieczność zaangażowania wielu interesariuszy w opisanie organizacji wymaga wykorzystania sformalizowanych i jak najbardziej uniwersalnych mechanizmów współpracy. Powinno to umożliwić pozyskanie większej grupy podmiotów wnoszących wartościowy wkład do opisu organizacji oraz łatwiejsze wprowadzenie do zespołu nowych członków na późniejszym etapie projektu.

Aby sprostać przedstawionym wyzwaniom, opracowano wiele standardów, modeli i pojęć, które mają w tym pomóc. Należą do nich m.in. architektura korporacyjna (*enterprise architecture*) i architektura danych (*data architecture*). Architektura korporacyjna to zbiór właściwości danej korporacji (włącznie ze strukturą), które stanowią o zdolności do realizacji jej misji – czyli dokładny jej opis [Graves, 2016].

Podwaliny pod dziedzinę badań nad architekturą korporacyjną zostały położone poprzez pracę J.A. Zachmana, opublikowaną w 1987 roku [Zachman, 1987]. Jej głównym celem była próba sprostania złożoności zarządzania coraz bardziej rozproszonymi systemami. Według Zachmana najlepszym sposobem na uświadomienie sobie prawdziwej wartości biznesowej drzemącej w organizacji i sposobem na osiągnięcie odpowiedniego poziomu zwinności w opisywaniu, projektowaniu i modyfikowaniu zachodzących w niej procesów jest holistyczne podejście do całej architektury systemowej, które bierze pod uwagę każde ważne zagadnienie w organizacji z każdej istotnej perspektywy. Takie spojrzenie na architekturę systemu zostało przez niego opisane jako Model Architektury Systemów Informacyjnych (*Information Systems Architectural Framework*), który został później przemianowany na Model Architektury Korporacyjnej (*Enterprise Architecture Framework*).

System informacyjny wspierający nowoczesną organizację może być bardzo skomplikowany i jego zaprojektowanie oraz obsługa wymagają współpracy wielu wyspecjalizowanych interesariuszy. W zależności od posiadanej wiedzy i zakresu obowiązków spojrzenie na ten sam proces może oznaczać wykorzystanie innych narzędzi i kompetencji do jego opisanie i zaprojektowania. Menedżer, projektant baz danych i inżynier sieciowy będą mieli zupełnie inny punkt widzenia i wymagania np. na proces komunikacji z klientem, ale każda z tych perspektyw (pomimo innego poziomu szczegółowości, narzędzi czy języka służącego do jego opisanie) będzie istotna do osiągnięcia sukcesu organizacji i wymaga szczegółowego projektu. W ten sposób powstaje wiele dokumentów

opisujących architekturę systemu (artefaktów), które są porządkowane przez Model Zachmana (*Zachman's Enterprise Framework*).

Model Zachmana jest szablonem służącym do organizowania artefaktów architektury (dokumentów projektowych, specyfikacji, modeli), które biorą pod uwagę zarówno ich odbiorców (np. właściciele, projektantów), jak i poszczególne zagadnienia, które wymagają opisanie i zaprojektowania (np. dane i funkcjonalności). W przedstawionej na rysunku 1.1 strukturze Zachmana wymiar wierszy organizuje poziom artefaktów według interesariuszy. Zachman zaproponował sześć poziomów – planistę, właściciela, projektanta, budowniczego, podwykonawcę i korporację. Wymiar kolumn koncentruje się na opisie celu artefaktu w projekcie: co (dane), jak (funkcje), gdzie (sieci), kto (ludzie), kiedy, dlaczego.

Zarówno właściciel, jak i projektant muszą znać odpowiedzi na te same pytania, ale odpowiedzi na nie będą w zależności od pytającego inne.

Bardziej rozbudowane podejście do zagadnienia architektury korporacyjnej zostało zaproponowane w modelu TOGAF (*The Open Group Architecture Framework*) [Open Group, 2011]. Open Group w swoim modelu definiuje architekturę korporacyjną jako spójny model organizacji (rozumianej zarówno jako jednostki administracji publicznej, jak i przedsiębiorstwa) integrującej: cele i zadania strategiczne organizacji, procesy biznesowe oraz systemy informacyjne i technologiczne niezbędne do realizacji jej celów [Goikoetxea, 2014].

Architektura korporacyjna TOGAF składa się z następujących elementów [Open Group, 2005]:

- pryncypia architektury korporacyjnej – zbiór trwałych zasad bazujących na strategii rozwoju organizacji, które stanowią reprezentację całościowych potrzeb organizacji w zakresie tworzenia rozwiązań informatycznych,
- architektura biznesowa – określa strategię biznesową i sposoby zarządzania organizacją, jej strukturę organizacyjną oraz główne procesy biznesowe, a także relacje pomiędzy tymi elementami,
- architektura danych,
- architektura aplikacji – opisuje poszczególne systemy oprogramowania, ich rozlokowanie, wzajemne współdziałanie oraz relacje pomiędzy tymi systemami a głównymi procesami biznesowymi,
- architektura technologiczna – opisuje infrastrukturę techniczną, która stanowi podstawę funkcjonowania kluczowych systemów oprogramowania (obejmuje ona m.in.: systemy operacyjne, systemy zarządzania bazami danych, serwery aplikacyjne, sprzęt komputerowy oraz infrastrukturę komunikacyjną).

**Rys. 1.1.1.** Warstwa danych w modelu Zachmana

	Dane Co	Funkcje Jak	Sieć Gdzie	Ludzie Kto	Czas Kiedy	Motywacja Dlaczego
Zakres (kontekstualny)	Lista bytów istotna dla biznesu	Lista procesów realizowanych przez organizację	Lista lokalizacji, w których prowadzone są operacje biznesowe	Lista jednostek organizacyjnych	Lista zdarzeń istotnych z punktu widzenia biznesu	Lista celów strategicznych
Planista Model biznesu (kontekstualny)	Model semantyczny	Model procesów biznesowych	Biznesowa logika systemu	Model przepływu pracy	Harmonogram nadrzędny	Plan biznesowy
Właściciel Model systemu (logiczny)	Logiczny model danych	Architektura aplikacji	Architektura systemu rozproszonego	Architektura interfejsu użytkownika	Struktura przetwarzania	Model reguł biznesowych
Projektant Model techniczny (fizyczny)	Fizyczny model danych	Projekt systemu	Architektura techniczna	Architektura prezentacji	Struktura sterowania	Projekty reguł
Budowniczy Szczegółowa reprezentacja	Definicja danych	Program	Architektura sieci	Architektura bezpieczeństwa	Definicje zdarzeń czasowych	Specyfikacja reguł
Podwykonawca Działająca korporacja	Dane	Funkcje	Sieć	Organizacja	Harmonogram	Strategia

**Źródło:** [Sobczak, 2013].



Najważniejszym komponentem standardu TOGAF jest proces tworzenia architektury – ADM (*Architecture Development Method*), który opisuje metodę konstruowania i zarządzania architekturą korporacyjną. Określa ona w ustandaryzowany sposób cele, założenia, zbiór niezbędnych danych wejściowych, poszczególne kroki projektowe i dane wyjściowe niezbędne do opisanie opracowywanej architektury. Struktura Zachmana pozwalała określić miejsce artefaktów niezbędnych dla architektury korporacyjnej, a ADM jest narzędziem do ich tworzenia.

Jednym z obszarów zainteresowań, które wchodzi w skład architektury korporacyjnej, jest architektura danych. Według definicji TOGAF jest to opis struktury i interakcji podstawowych typów i źródeł danych przedsiębiorstwa, logicznych i fizycznych zasobów danych oraz mechanizmów zarządzających danymi [Harrison, 2013]. Razem z architekturą aplikacyjną (*application architecture*) tworzą one kluczową dla organizacji architekturę systemów informacyjnych. Zgodnie z założeniami ADM architektura danych jest przedmiotem ciągłego procesu projektowego, uzależnionego od ewolucji pozostałych elementów architektury korporacyjnej i zmian wymuszających dostosowanie projektu architektury.

W dziedzinie technologii informacyjnych architektura danych jest zbiorem modeli, polityk, reguł i standardów, które są stosowane w procesie zarządzania sposobem zbierania, składowania, integracji i używania danych w przedsiębiorstwie [Sobczak, 2013].

Jak podkreślono we wstępie rozdziału, dane, którymi powinna zarządzać organizacja, mogą mieć różne źródła i charakter. Przykładem danych użytecznych dla organizacji mogą być:

- stany magazynowe przechowywane w bazach danych programów ERP,
- dane kontaktowe klientów składowane w plikach Excela,
- poczta elektroniczna użytkowników organizacji,
- zapis lokalizacji pojazdów organizacji z nadajnikami GPS,
- zapisy wideo z kamer monitoringu,
- skany faktur,
- dane dotyczące organizacji lub klientów zamieszczane na portalach społecznościowych,
- wiadomości na branżowych portalach internetowych,
- wpisy z wypełnionych przez klientów formularzy kontaktowych,
- systemy przechowujące dane o poświadczeniach, pozwalające na uwierzytelnienie i autoryzację użytkowników.

Różnorodność danych, które są w posiadaniu organizacji lub w zasięgu jej zainteresowania, może być bardzo duża, a w sytuacji specjalistycznych organizacji może mieć charakter unikatowy. Część danych będzie

jednak miała charakter uniwersalny, np. dane z systemów obsługujących dział zasobów ludzkich (*Human Resources*). Pomimo uniwersalnego charakteru system zarządzania danymi może być jednak zupełnie inaczej zaimplementowany, np. oferować inny *interface* użytkownika, zapewniać różny poziom szczegółowości danych lub przechowywać dane w różnych systemach bazodanowych (*Database Management System*).

Do poprawnego zdefiniowania architektury danych istotne jest nie tylko zdefiniowanie oczekiwań interesariuszy systemu. W przypadku systemu informacyjnego to dane są jego kluczowym elementem i należy również z perspektywy danych określić ich cechy, wymagania w stosunku do nich i właściwości środowiska, w którym będą składowane i przetwarzane.

Według modelu TOGAF można wymienić sześć pryncypiów kluczowych dla danych:

- dane są wartościowym zasobem,
- dane są udostępniane,
- dane mają być dostępne,
- dane wymagają powiernika (*data trustee*),
- dane wymagają zdefiniowania i wspólnego słownika pojęć,
- dane muszą być bezpieczne.

Celem danych jest wspomaganie procesów decyzyjnych. Dokładne, aktualne dane mają kluczowe znaczenie dla dokładnych, terminowych decyzji. Większość aktywów firmy jest starannie zarządzana, a dane nie są wyjątkiem. Dane są podstawą procesu decyzyjnego, dlatego też należy starannie nimi zarządzać, aby upewnić się co do ich lokalizacji, wiarygodności i dokładności oraz możliwości uzyskania do nich dostępu w sytuacji, kiedy jest to potrzebne.

Według TOGAF zagwarantowanie jakości danych i właściwe zarządzanie danymi wymaga od organizacji spełnienia następujących wymogów:

- dzięki odpowiedniej edukacji wszystkie organizacje w korporacji rozumieją relacje pomiędzy wartością danych, ich udostępnianiem i dostępem do nich;
- należy odejść od myślenia w kategoriach właściciela danych (*Data ownership*) na rzecz zarządcy danych (*Data stewardship*);
- zarządcy muszą mieć władzę i środki do zarządzania danymi, za które są odpowiedzialni;
- rola zarządcy danych jest krytyczna, ponieważ w wyniku jego nieprawidłowej pracy może dojść do przekazania personelowi przedsiębiorstwa przestarzałych, nieprawidłowych lub niespójnych danych, co może mieć negatywny wpływ na proces podejmowania decyzji w przedsiębiorstwie;

- zadaniem zarządcy danych jest zapewnienie jakości danych; należy opracowywać i wykorzystywać procedury w celu zapobiegania występowaniu błędów w informacjach i ich korygowania oraz usprawnienia tych procesów, które są źródłem błędnych informacji; wymagane jest opracowanie i wdrożenie polityk i procedur pozwalających na pomiar jakości danych i podejmowanie kroków w celu poprawy ich jakości;
- o zmianach procesów proponowanych przez zarządcę danych powinno decydować forum z szeroką reprezentacją w skali całego przedsiębiorstwa;
- ponieważ dane są wartością dla całego przedsiębiorstwa, odpowiedzialni za prawidłowe zarządzanie danymi muszą być asygnowani na szczeblu przedsiębiorstwa.

Szybki dostęp do dokładnych danych ma podstawowe znaczenie dla poprawy jakości i skuteczności podejmowania decyzji przez przedsiębiorstwa. Łatwiej i taniej jest utrzymywać aktualne, dokładne dane w pojedynczej aplikacji, a następnie dzielić się nimi, niż utrzymywać duplikaty danych w wielu aplikacjach. Przedsiębiorstwa posiadają duże ilości danych, ale są one często przechowywane w setkach niezgodnych baz danych. Prędkość gromadzenia danych, ich tworzenia, przekazywania i asymilacji zależy od zdolności organizacji do efektywnego współużytkowania tych wysp danych w ramach całej organizacji.

Według TOGAF do zagwarantowania możliwości udostępniania danych, a co za tym idzie ich wymiany pomiędzy różnymi systemami, istotne jest wdrożenie następujących wytycznych:

- aby umożliwić wymianę danych, należy opracować i przestrzegać wspólnych zasad, procedur i standardów dotyczących zarządzania danymi i dostępu do nich zarówno w krótkim, jak i w długim okresie;
- w perspektywie krótkoterminowej, aby nie zaprzepaścić inwestycji w starsze systemy, należy zainwestować w oprogramowanie zdolne do migracji danych ze starszych systemów do wspólnego środowiska danych;
- należy opracować standardowe modele danych, elementy danych i inne metadane definiujące współużytkowane środowisko i opracować system repozytoriów służący do przechowywania tych metadanych;
- w perspektywie długoterminowej, gdy przestarzałe systemy zostaną zastąpione nowymi rozwiązaniami, należy przyjąć i egzekwować wspólne zasady dostępu do danych i wytyczne dla nowych twórców aplikacji, aby zagwarantować, że dane w nowych

aplikacjach są dostępne dla wspólnego środowiska, a dane w środowisku współużytkowanym mogą nadal być wykorzystywane przez nowe aplikacje;

- w perspektywie krótko- i długoterminowej należy przyjąć wspólne metody i narzędzia służące do tworzenia, przechowywania i uzyskiwania dostępu do danych udostępnianych w przedsiębiorstwie;
- praca we wspólnym środowisku danych będzie wymagać znacznej zmiany kulturowej;
- zasada udostępniania danych będzie stale w sprzeczności z zasadą bezpieczeństwa danych; w żadnym wypadku zasada udostępniania danych nie może doprowadzić do naruszenia poufności danych;
- udostępnione dane powinny być wykorzystywane przez wszystkich użytkowników do wykonywania powierzonych im czynności. Dzięki jednemu źródłu danych tylko najbardziej aktualne i dokładne dane zostaną użyte w procesie podejmowania decyzji.

Szeroki dostęp do danych prowadzi do poprawienia efektywności i skuteczności w podejmowaniu decyzji oraz pozwala na dostarczenie aktualnych odpowiedzi na ządania o dane. Korzystanie z informacji musi być rozpatrywane z punktu widzenia przedsiębiorstwa, aby umożliwić dostęp szerokiej gamie użytkowników. Można zaoszczędzić czas pracy personelu i zwiększyć spójność danych.

Według TOGAF, aby zagwarantować użytkownikom efektywny dostęp do danych, powinny być przestrzegane następujące zalecenia:

- dostępność obejmuje łatwość uzyskiwania informacji przez użytkowników;
- sposób uzyskiwania dostępu do informacji i wyświetlenia musi być wystarczająco elastyczny, aby sprostać wymaganiom użytkowników korporacyjnych pełniących różne funkcje, co może się przekładać na inne metody dostępu (różne urządzenia dostępowe, dostęp z urządzeń mobilnych) i zapotrzebowanie na dane o różnorodnym charakterze (zaawansowane raportowanie lub wykonywanie bardzo prostych i powtarzalnych operacji);
- dostęp do danych nie oznacza zrozumienia danych; personel powinien zachować ostrożność, aby nie nadinterpretować informacji;
- dostęp do danych niekoniecznie daje prawa dostępu użytkownika do modyfikowania lub ujawniania danych. Uświadomienie pracownikom organizacji znaczenia danych, zasad kontroli dostępu do danych i konsekwencji ujawnienia informacji wymaga procesu edukacyjnego i często zmiany w kulturze organizacyjnej, która obecnie popiera przekonanie o „własności” danych przez jednostki funkcjonalne.

Jedną z zalet projektowanego środowiska jest możliwość dzielenia się danymi (np. tekstem, wideo, dźwiękiem itp.) w przedsiębiorstwie z wieloma użytkownikami. Wraz ze wzrastającym stopniem udostępniania danych zwiększa się stopień, w którym różne jednostki biznesowe opierają się na tych samych zasobach danych organizacji – w takim przypadku istotne staje się, aby za zawartość danych odpowiadał tylko jeden podmiot – powiernik danych (*data trustee*). W sytuacji, kiedy te same dane (np. dane klienta) są powielane i modyfikowane w wielu miejscach lub przechowywane w jednej kopii, ale modyfikowane przez wielu użytkowników, może dojść do utraty integralności (*integrity*) danych. Utworzenie roli powiernika danych i powierzenie jej wyłącznej odpowiedzialności za wprowadzanie i modyfikacje konkretnych danych eliminuje powielanie czynności wykonywanej przez pracowników i zmniejsza zapotrzebowanie na zasoby związane z przechowywaniem danych. W procesie zarządzania danymi można obok roli powiernika wymienić rolę stewarda danych (*data steward*). Powiernik różni się od zarządcy – jest odpowiedzialny za dokładność i jakość danych, a obowiązki stewarda mogą być szersze i obejmują:

- zarządzanie definicjami danych, rozpoczynając od ich znaczenia biznesowego aż po techniczną specyfikację potrzebną podczas implementacji (np. struktura i atrybuty obiektów biznesowych; zakresy dopuszczalnych wartości);
- tworzenie i realizowanie polityk oraz procedur związanych z zarządzaniem i uzgadnianiem zmian w definicjach danych oraz zmian w systemach informatycznych przetwarzających dane;
- tworzenie zasad zarządzania jakością danych;
- wspieranie właściwego i skutecznego wykorzystania danych.

Według TOGAF dla zagwarantowania odpowiedniej jakości danych proces zarządzania i dostępu do danych powinien charakteryzować się następującymi cechami:

- prawdziwe powiernictwo rozwiązuje kwestie „własności” danych i umożliwia dostęp do danych w celu zaspokojenia potrzeb wszystkich użytkowników, może być wymagana zmiana kulturowa z „własności” danych na „powiernictwo” danych;
- powiernik danych będzie odpowiedzialny za spełnienie wymagań jakościowych nakładanych na dane, za które jest odpowiedzialny;
- istotne jest, aby powiernik miał możliwość zapewnienia użytkownikom zaufania do danych opartych na atrybutach, takich jak „źródło danych”;
- konieczne jest zidentyfikowanie prawdziwego źródła danych, aby organowi ds. danych można było przypisać odpowiedzialność za jego kontrolowanie;

- informacje powinny być rejestrowane elektronicznie raz i natychmiast zatwierdzone tak blisko źródła, jak to możliwe; należy wdrożyć odpowiednie środki kontroli jakości, aby zapewnić integralność danych;
- w wyniku udostępniania danych w przedsiębiorstwie powiernik jest odpowiedzialny za dokładność i aktualność wyznaczonego elementu danych, a następnie musi uznać znaczenie tej odpowiedzialności powierniczej.

Dane, które będą wykorzystywane przy opracowywaniu aplikacji, muszą mieć wspólną definicję w całej centrali, aby umożliwić udostępnianie danych. Wspólne słownictwo ułatwi komunikację i umożliwi skuteczne dialogowanie. Dodatkowo wymagane jest powiązanie systemów i wymiany danych. Według TOGAF do skutecznego definiowania architektury danych na poziomie przedsiębiorstwa niezbędne jest przestrzeganie następujących zasad:

- przedsiębiorstwo musi ustalić początkowe wspólne słownictwo dla firmy; definicje będą stosowane jednolicie w całym przedsiębiorstwie;
- gdy wymagana jest nowa definicja danych, wysiłek w zakresie definicji zostanie skoordynowany i uzgodniony z firmowym „glosariuszem” opisów danych; proces powinien być skoordynowany przez administratora danych przedsiębiorstwa;
- niejednoznaczności wynikające z wykorzystywania wielu (często potocznych lub hermetycznych) definicji danych muszą ustąpić na rzecz definicji obowiązujących na poziomie przedsiębiorstwa;
- konieczne jest skoordynowanie inicjatyw dotyczących standaryzacji danych;
- należy przypisać funkcyjną odpowiedzialność zarządzania danymi na poziomie przedsiębiorstwa (np. koordynacja standaryzacji).

Dane powinny być chronione przed nieautoryzowanym użyciem i ujawnieniem. Każda organizacja posiada zasoby, które wymagają dodatkowego zabezpieczenia wybranych danych poprzez wdrożenie odpowiednich procedur i wykorzystanie adekwatnych środków technicznych. Zabezpieczenia mogą obejmować m.in. kontrolę dostępu do danych, ograniczanie czynności możliwych do wykonania na danych (np. ograniczenie możliwości wydruku czy edycji dokumentu), kontrolę przepływu danych w organizacji i do podmiotów zewnętrznych. Według TOGAF podstawowe wytyczne dotyczące bezpieczeństwa w architekturze danych są następujące:

- bezpieczeństwo musi być od początku projektowane w elementach danych; systemy, dane i technologie muszą być zabezpieczo-

ne przed nieautoryzowanym dostępem i manipulacją; informacje muszą być chronione przed niezamierzoną lub nieautoryzowaną zmianą, sabotażem, katastrofą lub ujawnieniem;

- można zaimplementować mechanizmy, które będą ograniczać dostęp do danych typu „tylko do zobaczenia” (np. bez możliwości drukowania czy kopiowania); należy wdrożyć możliwość etykietowania danych pod kątem stopnia ich zabezpieczenia, np. etykiety dla danych kluczowych do podejmowanych działań biznesowych, niejawnych, wrażliwych lub zastrzeżonych;
- aby w odpowiedni sposób zapewnić dostęp do informacji publicznie dostępnych przy jednoczesnym zabezpieczeniu danych, które wymagają kontroli dostępu, należy zidentyfikować i opracować kwestie bezpieczeństwa na poziomie danych, a nie na poziomie aplikacji;
- należy rozważyć architekturę danych pod kątem przechowywania danych jawnych i niejawnych; tradycyjne podejście polegające na utrzymywaniu oddzielnych systemów do składowania różnie sklasyfikowanych danych może być optymalnym rozwiązaniem w kontekście rozwoju oprogramowania i sprzętu; nowoczesne systemy bazodanowe udostępniają możliwość przechowywania i oddzielania danych, które potrzebują dodatkowego zabezpieczenia (np. szyfrowania) w systemach służących do przechowywania danych jawnych; przechowywanie danych jawnych w systemach przeznaczonych do pracy z danymi niejawnymi będzie rozwiązaniem droгим i nieefektywnym.

Założenia odnośnie do możliwych do implementacji cech danych, uświadomienie sobie związanych z tym ograniczeń i możliwości oraz potencjalnych problemów będzie wpływało na możliwość opracowania dobrej architektury danych.

W wyniku zaprojektowania architektury danych powinniśmy otrzymać opis docelowego (pożądanego) rozwiązania, które z jednej strony odpowiada na potrzeby użytkowników systemu poprzez zagwarantowanie odpowiednich funkcjonalności, ale również przedstawia opis działań pozwalających to rozwiązanie wdrożyć. Zarówno Zachman, jak i TOGAF wymagają zdefiniowania komponentów odpowiedzialnych za dane na trzech poziomach szczegółowości:

- pojęciowym (konceptualnym) – jest to ogólne opisanie głównych obiektów biznesowych i procesów zachodzących pomiędzy nimi dla zrozumienia funkcjonowania organizacji; na tym poziomie następuje opisanie przez interesariuszy modelu danych wysokiego poziomu (opisującego całą organizację, ale bez wyjaśnienia szczegółów) przy pomocy wspólnego i zrozumiałego dla wszystkich zestawu pojęć;

- logicznym – jest to szczegółowy model danych wykonany przy pomocy obiektów modelowania danych (*Data Modeling Objects*) – szerzej opisanych w rozdziale dotyczącym baz danych – opisany niezależnie od konkretnego rozwiązania bazodanowego lub technologii składowania danych;
- fizycznym – jest to szczegółowy plan pozwalający na realizację planu logicznego osadzonego w konkretnych systemach DBMS i rozwiązaniach sprzętowych; na tym poziomie szczegółowości wykorzystuje się konkretne cechy wybranych rozwiązań do osiągnięcia zakładanych funkcjonalności, ale również do opracowania szczegółowej architektury mającej zagwarantować odpowiednią wydajność i niezawodność.

Zaprojektowanie fizycznego modelu oznacza zdefiniowanie konkretnych sposobów składowania danych, ale również opisania wszystkich pozostałych elementów niezbędnych do ciągłej pracy systemu. System informacyjny musi zagwarantować odpowiednie przepływy danych, mechanizmy zarządzania.

## 1.4. Podsumowanie

Celem architektury korporacyjnej jest optymalizacja zachodzących w organizacji procesów (automatycznych lub ręcznych) w zintegrowanym środowisku, które potrafi reagować na zmiany i wspiera osiągnięcie celów strategii biznesowej. Efektywne wykorzystanie i zarządzanie systemami informacyjnymi jest jednym z kluczowych czynników pozwalających osiągnąć przewagę konkurencyjną i zapewnić sukces organizacji. Architektura korporacyjna w zdecydowany sposób wiąże rozwój systemów IT organizacji z osiągnięciem celów biznesowych, a dobra architektura korporacyjna potrafi określić poprawną równowagę pomiędzy efektywnością IT a innowacyjnością biznesową.

Do korzyści, które wynikają z dobrze zaprojektowanej architektury korporacyjnej, można zaliczyć [Open Group, 2011]:

- bardziej wydajne działania (operacje) biznesowe:
  - niższe koszty operacyjne,
  - możliwość adaptacji do sytuacji w otoczeniu organizacji,
  - możliwości biznesowe współdzielone na przestrzeni organizacji,
  - niższe koszty związane z zarządzaniem zmianą,



- bardziej elastyczni pracownicy,
- zwiększona wydajność pracy;
- bardziej wydajna działalność IT:
  - niższe koszty tworzenia i utrzymania oprogramowania,
  - zwiększenie możliwości przeniesienia aplikacji,
  - zwiększenie interoperacyjności i uproszczenie zarządzania siecią i systemami,
  - zwiększona zdolność do radzenia sobie z zagadnieniami krytycznymi dla całej organizacji, np. zagadnienia bezpieczeństwa,
  - ułatwiona aktualizacja i wymiana komponentów systemu obsługującego organizację;
- większy zwrot z poczynionych inwestycji, mniejsze ryzyko przyszłych inwestycji:
  - zredukowana złożoność środowiska biznesowego i IT,
  - maksymalny zwrot z inwestycji w istniejące przedsięwzięcia biznesowe i infrastrukturę IT,
  - elastyczność w podejmowaniu decyzji odnośnie tworzenia, zakupu lub wykorzystania zewnętrznych rozwiązań biznesowych i IT;
- szybsze i łatwiejsze dostarczenie produktu:
  - decyzje kupna są prostsze, ponieważ informacje dotyczące zamówień są łatwo dostępne w spójnym planie,
  - proces zamówień jest szybszy – maksymalizacja szybkości i elastyczności zamówień bez uszczerbku dla spójności architektonicznej,
  - zdolność do pozyskiwania heterogenicznych, otwartych systemów wielu dostawców,
  - zdolność do zabezpieczenia większej możliwości ekonomicznej.

## Pytania kontrolne

1. Opisz sześćoetapowy model rozwoju technologii informacyjnych w organizacji.
2. Wyjaśnij pojęcie architektury korporacyjnej i architektury danych.
3. Podaj pięć przykładów źródeł danych i pięć przykładów danych, które mogą być wykorzystywane przez organizację.
4. Wymień główne składowe architektury korporacyjnej według TOGAF.
5. Wymień i opisz główne cechy danych według TOGAF.

# Studium przypadku

## Przypadek 1

Fikcyjna organizacja HykerExcelence podejrzewa, że w ostatnim czasie dochodziło do wycieku wewnętrznych danych finansowych do konkurencji i chce w bardziej restrykcyjny sposób kontrolować możliwość pracy na nich. Wdrożone mechanizmy powinny umożliwiać klasyfikowanie i nadawanie materiałom elektronicznym (dokumentom, wiadomościom poczty elektronicznej) łatwo rozpoznawalnych sygnatur, które stanowiłyby dla użytkownika informację o specjalnym statusie dokumentu (np. dokument tylko do użytku działu, dokument zawierający dane wrażliwe, materiały finansowe). Dodatkowo powinna być dostępna możliwość zablokowania pewnych funkcjonalności, które mogłyby ułatwić udostępnienie danych niepowołanym użytkownikom (np. drukowanie i kopiowanie dokumentu, przesyłanie dalej wrażliwych wiadomości poczty elektronicznej).

Dostępną funkcjonalność można uzyskać poprzez usługę AD RMS (*Active Directory Rights Management Services*). Jest ona jednak dostępna w dwóch modelach wdrożenia – w siedzibie organizacji (*on-premise*) i poprzez usługę w chmurze obliczeniowej – Azure RMS. W pierwszym przypadku do wdrożenia potrzebne są następujące komponenty (lista nie uwzględnia liczby serwerów niezbędnych do osiągnięcia zadowalającego poziomu dostępności usługi):

- kontroler domeny Active Directory,
- serwer z zainstalowaną usługą AD RMS,
- serwer poczty elektronicznej Exchange,
- komputery użytkowników z zainstalowanym klientem RMS i oprogramowaniem pozwalającym z niego korzystać (np. system Windows lub pakiet biurowy Office).

W przypadku wdrożenia usługi w chmurze wszystkie niezbędne komponenty mogą być uruchomione na serwerach usługodawcy. Dodatkowo użytkownicy otrzymują możliwość pracy z dokumentami Office poprzez przeglądarkę komputerową.

Udostępniane przez usługę funkcjonalności wydają się proste do uzyskania – „schowanie” przed użytkownikiem kilku opcji (np. drukowanie, kopiowanie) czy dodanie do dokumentu odgórnie znaku wodnego. W sytuacji jednak, kiedy mechanizm ma być automatyczny i w pełni kontrolowany, tj. kiedy dokumenty mają być automatycznie klasyfikowane do odpowiedniej polisy ochronnej według założonych kryteriów

(np. wszystkie wiadomości poczty elektronicznej zawierające numery PESEL nie mogą zostać wysłane poza organizację; dokumenty utworzone przez użytkowników z działu księgowości mogą być drukowane i kopiowane tylko przez ten dział i menedżerów), okazuje się, że wykorzystywane do tego mechanizmy są bardzo skomplikowane.

Organizacja może w związku z tym zdecydować się ze względu na barierę technologiczną (wykorzystanie skomplikowanych narzędzi i konieczność przeszkolenia kadry IT) i ekonomiczną (wysoki koszt licencji i sprzętu niezbędnego do uruchomienia infrastruktury na miejscu) na wykorzystanie rozwiązań opartych na chmurze obliczeniowej.

#### Przypadek 2

Fikcyjna organizacja Adatum wykorzystuje infrastrukturę bazodanową do zbierania, składowania i analizy danych ze swojego sklepu internetowego. Dotychczasowa baza danych sklepu i dzienniki wejść na witryny WWW firmy znajdowały się na tym samym serwerze, ale wraz ze wzrostem popularności serwisu obciążenie związane z monitorowaniem sklepu w znacznym stopniu zaczęło obciążać serwer i zmniejszać komfort pracy klientów. W związku z wykorzystywanym – i wystarczającym na początku aktywności sklepu – mechanizmem przekazywania danych do bazy danych (zdarzenia na witrynach sklepu były odkładane w plikach tekstowych o maksymalnym rozmiarze 100 MB i po wypełnieniu aktualnego zbioru odczytywane i zapisywane do bazy danych) analiza bieżącej aktywności była obciążona pewną zwłoką. W trakcie przesyłania danych do bazy na serwerze można było odnotować znaczny wzrost wykorzystania procesorów i zmniejszenie wydajności sklepu. Dodatkowo problem nasilał się wraz ze zwiększoną aktywnością w sklepie.

Dział IT musiał zmierzyć się z dwoma wyzwaniami – rozwiązaniem problemu spadającej wydajności i ulepszeniu istniejącej lub zaproponowaniu nowej platformy pozwalającej na proste i elastyczne tworzenie raportów przez pracowników na podstawie opisywanych danych.

W wyniku analizy procesu zidentyfikowano główne źródło problemu – błąd w projekcie bazy danych i mało efektywny sposób przekazywania danych do bazy. Postanowiono jednak radykalnie zmodyfikować cały mechanizm, aby zagwarantować oczekiwane funkcjonalności raportowania i zagwarantować skalowalność rozwiązania, aby w przyszłości uniknąć podobnych problemów.

Podstawą nowego rozwiązania stał się klaster dwóch serwerów baz danych: Elasticsearch i serwer raportowania Kibana. Zrezygnowano z relacyjnego serwera baz danych na rzecz silnika, który jest zaprojektowany jako system do pracy w układzie rozproszonym i pozwala na łatwe rozkładanie obciążenia pomiędzy węzłami klastra. Oferuje on również

prosty język zorientowany na wyszukiwanie informacji i jest natywnym rozwiązaniem dla obsługi dzienników zdarzeń. Serwer raportowania Kibana pozwala na efektywne i proste tworzenie raportów opartych na wielu silnikach baz danych, m.in. Elasticsearch. Użytkownicy systemu po krótkim przeszkoleniu powinni móc projektować raporty i wykorzystywać platformę do przeprowadzania analiz w oparciu o dane przestrzenne, czas i inne atrybuty przechowywane w bazie.

## Literatura

- Goikoetxea A. (2004), *A Mathematical Framework for Enterprise Architecture Representation and Design*, „International Journal of Information Technology & Decision Making”, Vol. 3, No. 1.
- Graves T. (2016), *Enterprise Architecture*, IT Governance Publishing, Cambridgeshire.
- Harrison R. (2013), *TOGAF 9 Foundation Study Guide*, Van Haren Publishing, Her-togenbosch.
- Nolan R.L. (1979), *Managing the Crises in Data Processing*, „Harvard Business Review”, March–April.
- Open Group (2005), *The Open Group Architecture Framework*, Book Edition, September 2005, Amsterdam.
- Open Group (2011), *TOGAF Version 9.1 Revised Edition*, Van Haren Publishing, Her-togenbosch.
- Sobczak A. (2013), *Architektura korporacyjna. Aspekty teoretyczne i wybrane zastosowania praktyczne*, Ośrodek Studiów nad Cyfrowym Państwem, Poznań.
- Zachman J.A. (1987), *A Framework for Information Systems Architecture*, „IBM Systems Journal”, Vol. 26, Issue 3.