

CYFRYZACJA

Zarządzanie danymi w organizacji

pod redakcją
Beaty Gontar

Zarządzanie danymi w organizacji



WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

CYFRYZACJA

Zarządzanie danymi w organizacji

pod redakcją
Beaty Gontar

Beata Gontar – Uniwersytet Łódzki, Wydział Zarządzania, Katedra Informatyki
90-237 Łódź, ul. Matejki 22/26

RECENZENT

Gabriela Idzikowska

REDAKTOR INICJUJĄCY

Monika Borowczyk

OPRACOWANIE REDAKCYJNE

Aleksandra Urzędowska

SKŁAD I ŁAMANIE

AGENT PR

PROJEKT OKŁADKI

Katarzyna Turkowska

Zdjęcie wykorzystane na okładce: © Depositphotos.com/everythingposs

Katarzyna Ciach (rozdział IV), Piotr Czerwonka (rozdział I, VI), Zbigniew Gontar (rozdział III)
Ryszard Kurzyjamski (rozdział II), Grzegorz Podgórski (rozdział V)
Jerzy S. Zieliński (rozdział VI)

© Copyright by Authors, Łódź 2019

© Copyright for this edition by Uniwersytet Łódzki, Łódź 2019

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.08380.17.0.K

Ark. wyd. 11,0; ark. druk. 14,375

ISBN 978-83-8142-629-9

e-ISBN 978-83-8142-630-5

Wydawnictwo Uniwersytetu Łódzkiego

90-131 Łódź, ul. Lindleya 8

www.wydawnictwo.uni.lodz.pl

e-mail: ksiegarnia@uni.lodz.pl

tel. (42) 665 58 63

Spis treści

Wstęp	9
Rozdział 1	
Architektura danych	13
1.1. Wprowadzenie	13
1.2. Dane i rozwój organizacji	14
1.3. Architektura korporacyjna i architektura danych	17
1.4. Podsumowanie	28
Pytania kontrolne	29
Studium przypadku	30
Literatura	32
Rozdział 2	
Bazy danych	33
2.1. Wprowadzenie	33
2.2. Podstawowe pojęcia baz danych	34
2.3. Użytkownicy baz danych	38
2.4. Modele danych	41
2.4.1. Model relacyjny	41
2.4.2. Model encyjno-relacyjny	43
2.5. Funkcje Systemu Zarządzania Bazą Danych (SZBD)	44
2.6. Wymogi integralności	50
2.7. Struktura SZBD	52
2.8. Organizacja danych w bazach danych	58
2.9. Kontrola dostępu	60
2.10. Język SQL	62
2.11. Podsumowanie	66
Pytania kontrolne	66
Studium przypadku	67
Literatura	72

Rozdział 3

<i>Business Intelligence</i>	73
3.1. Wprowadzenie	73
3.2. Podstawowe pojęcia	77
3.2.1. Proces zarządzania efektywnością i łańcuch poznawczy dane-informacja-wiedza-mądrość (<i>Data-to-Information-to-Knowledge-to-Wisdom</i> , DIKW)	77
3.2.2. Model referencyjny dla kluczowych wskaźników efektywności	80
3.2.3. Analiza i monitoring stanu przedsiębiorstwa	84
3.2.4. <i>Benchmarking</i>	85
3.2.5. Analiza granulacyjna	86
3.2.6. Wizualizacja procesu	87
3.3. Wielowymiarowy model danych	89
3.4. <i>Business Intelligence</i> na poszczególnych poziomach piramidy informacyjnej	90
3.4.1. Poziom danych	90
3.4.2. Poziom informacji	91
3.4.3. Poziom wiedzy	92
3.4.4. Poziom mądrości strategicznej	94
3.5. Charakterystyka Systemów <i>Business Intelligence</i>	94
3.5.1. Pojęcie Systemu <i>Business Intelligence</i>	94
3.5.2. Cechy systemów BI	95
3.5.3. Struktura systemów BI	96
3.6. Języki programowania systemów BI	98
3.7. Podsumowanie	107
Pytania kontrolne	109
Studium przypadku	109
Literatura	122

Rozdział 4

Wizualizacja informacji w biznesie	123
4.1. Wprowadzenie	123
4.2. Zmysł wzroku i jego zdolność do przyswajania informacji	123
4.3. Pojęcie i geneza wizualizacji informacji	129
4.4. Wybrane funkcje wizualizacji informacji	130
4.4.1. Funkcja redukcji nadmiaru informacji	131
4.4.2. Funkcja analityczna	132

4.4.3. Funkcja zarządzająca	133
4.4.4. Funkcja intelektualna	133
4.4.5. Funkcja komunikacyjna	134
4.4.6. Funkcja manipulacyjna	135
4.4.7. Funkcja społeczna	137
4.4.8. Funkcja edukacyjna	138
4.5. Metody wizualizacji	138
4.5.1. Proces wizualizacji	138
4.5.2. Przegląd metod wizualizacji	139
4.5.2.1. Dane liczbowe	139
4.5.2.2. Dane tekstowe	148
4.5.2.3. Daty	150
4.5.2.4. Dane przestrzenne	151
4.6. Komunikacja wizualna	153
4.7. Podsumowanie	156
Pytania kontrolne	157
Studium przypadku	157
Literatura	164

Rozdział 5

Bezpieczeństwo danych	167
5.1. Wprowadzenie	167
5.2. Wartość informacji i danych oraz ich znaczenie dla organizacji	168
5.3. Podstawowe definicje	172
5.4. Zagrożenia	175
5.5. Sprawy przestępstw komputerowych	178
5.6. Wybrane zagrożenia bezpieczeństwa informacji	180
5.7. Ryzyko, zarządzanie ryzykiem, analiza ryzyka	184
5.8. Polityka bezpieczeństwa informacji (PBI)	188
5.9. Audyt	191
5.10. Mobilność użytkowników w środowisku IT	192
5.11. Podsumowanie	199
Pytania kontrolne	201
Studium przypadku	201
Literatura	204

Rozdział 6	
Nowe trendy i technologie	207
6.1. Wprowadzenie	207
6.2. Wpływ technologii mobilnych na ewolucję źródeł i sposobów dostępu do danych	208
6.3. Internet Rzeczy (IoT – <i>Internet of Things</i>)	215
6.3.1. Definicje	215
6.3.2. Obszary zastosowań IoT	217
6.3.3. Wybrane przypadki koncepcji zastosowań IoT	218
6.3.3.1. Współdziałanie z chmurą obliczeniową	218
6.3.3.2. Zastosowania w elektroenergetyce	218
6.3.3.3. Inne zastosowania	220
6.4. Podsumowanie	221
Pytania kontrolne	222
Studium przypadku	222
Literatura	226
Zakończenie	229

Wstęp

Rozwój technologii nie jest kwestią wyboru – trzeba się do niego dostosować¹. Współczesna technologia, oparta na automatyce, robotyce i sztucznej inteligencji, jest modularna i doskonale mobilna², posiada łatwość generowania danych i pracy z nimi. Dane dostępne w organizacjach liczone są już nie w petabajtach, a w eksabajtach i wyżej. Rozwiązywanie problemów w takich organizacjach, wzmacniające zdolność organizacji do dostosowywania się do nowych warunków, zdolność do podnoszenia się z kryzysów, konfliktów i zapaści oraz zdolność do zrównoważonego rozwoju³, wymaga sporych inwestycji w zarządzanie oparte na danych i zainteresowanie aplikacjami i narzędziami *data mining* (DM) i *data warehousing* (DW). Przyczyną tego stanu rzeczy należy szukać we wzroście dynamiki zdarzeń biznesowych wymuszającym zwiększenie elastyczności działania organizacji oraz jednoczesne wykorzystanie danych online i danych archiwalnych⁴, gwałtownym postępie w transformacji cyfrowej organizacji wymuszającym wykorzystanie nowych modeli biznesowych bazujących na danych⁵ oraz szybkim postępie cywilizacyjnym w zakresie nauki i techniki

-
- 1 A.K. Koźmiński, *Czwarta fala globalizacji*, „Rzeczpospolita”, 8.05.2017.
 - 2 *Ibidem*. „Procesy tworzenia wartości mogą być realizowane w najróżniejszych konfiguracjach organizacyjnych i terytorialnych. Zarządzanie tymi procesami polega na wyodrębnieniu modułów realizowanych za pomocą konkretnych konfiguracji zasobów (ludzkich, materialnych, technologicznych, wiedzy), na posadowieniu każdego z modułów w optymalnym dla niego środowisku i na koordynacji procesu, czyli na zarządzaniu łańcuchem dostaw”.
 - 3 D. Helbing, Projekt *FuturICT*, <https://futurict.inn.ac> [dostęp: 3.04.2018].
 - 4 Skrócenie cyklu życia organizacji, ale też technologii i wytwarzanych przy ich użyciu produktów i usług, czasu realizacji zamówienia etc.
 - 5 Tworzenie i sprzedaż danych (np. Thomson Reuters sprzedaje dane finansowe, dane o badaniach naukowych, dane dotyczące medycyny oraz newsy, Twitter sprzedaje *tweets* poprzez np. DataSift), zbieranie danych i sprzedaż reklam (np. Facebook, Google, Twitter), zbieranie danych i sprzedaż usług w zakresie analityki biznesowej, dodawanie danych do produktu (Babolat specjalizujący się w produkcji rakiet tenisowych oferuje model *Babolat Play Pure Drive* w szeregu czujników, które pozwalają na pozyskanie i analizę danych zebranych podczas gry tą rakieta), dodawanie danych do usług (np. holenderskie linie lotnicze KLM w usłudze *Meet & Sit*, korzystając z danych pochodzących z Facebooka i LinkedIn, proponują miejsca w samolocie obok osoby, której profil wynikający z analizy danych pochodzących ze wspomnianych źródeł odpowiada naszemu profilowi, który powstał również na podstawie analizy tych danych), tworzenie nowych usług będących rezultatem analiz typu eksploatacja danych (np. platformy typu MOOC w edukacji czy Uber w przewozach osobowych);

wynikającym z globalizacji dającej szybką amortyzację kosztów, rozwijającego się środowiska smart, umożliwiającego śledzenie, zapisywanie, analizowanie i kontrolowanie każdego zdarzenia w procesie biznesowym, zmian demograficznych wymuszających rozwój nowych rozwiązań pozwalających na zaadaptowanie się społeczeństw do nowych warunków środowiskowych wynikających z wyczerpywania się zasobów paliw kopalnych.

Książka jest wynikiem wieloletnich doświadczeń autorów w pracy naukowej i dydaktycznej na Uniwersytecie Łódzkim. Stanowi ona zbiór materiałów, które mogą być wykorzystywane podczas nauki zagadnień związanych z pracą z danymi w różnego typu systemach informatycznych, ich strukturą i bezpieczeństwem, zarówno na zajęciach dydaktycznych, jak i w pracy indywidualnej. Publikacja może być zatem traktowana jako podręcznik do przedmiotów związanych z technologiami informacyjnymi i systemami informatycznymi, które od kilku lat są w minimach programowych większości kierunków studiów I stopnia.

Celem niniejszego opracowania jest przybliżenie czytelnikom, w jaki sposób technologie informatyczne i telekomunikacyjne wspierają zarządzanie. W poszczególnych rozdziałach omówiono zagadnienia związane z takimi zagadnieniami, jak:

- architektura danych, gdzie znajdziemy definicję danych i ich miejsce w rozwijającej się organizacji; przedstawiony zostanie model Zachmana oraz architektura korporacyjna TOGAF;
- teoria baz danych, a w szczególności zagadnienia związane z projektowaniem bazy danych oraz wykorzystaniem języka SQL do ich tworzenia i formułowania zapytań;
- systemy *Business Intelligence* i hurtownie danych;
- wizualizacja i sposoby przedstawiania danych oraz informacji;
- bezpieczeństwo danych;

oraz nowe technologie, jak *Internet of Things*, które coraz częściej są wykorzystywane w organizacjach.

Układ publikacji został tak pomyślany, aby mogła ona stanowić kompendium wiedzy do właściwego zrozumienia omawianych tematów, stanowiąc opracowanie wprowadzające w zagadnienia związane z systemami informacyjnymi szeroko rozumianego zarządzania.

por. C. Levallois, *Six Business Models Based on Data*, <https://www.slideshare.net/seinecle/six-business-models-based-on-data> [dostęp: 18.08.2017]; inaczej klasyfikuje to J. Lokitz, podając trzy modele: *Data as a Service (DaaS)*, *Information as a Service (IaaS)*, *Answers as a Service (AaaS)*, *Exploring Big Data Business Models & The Winning Value Propositions Behind Them*, <https://www.linkedin.com/pulse/exploring-big-data-business-models-winning-value-behind-justin-lokitz> [dostęp: 18.08.2017].

Struktura poszczególnych rozdziałów jest jednakowa. Każda część rozpoczyna się od *Wprowadzenia*, gdzie skrótowo zaprezentowano tematy, które zostaną szerzej omówione, oraz cel, jaki dana część książki realizuje. Rozdziały zakończone są krótkim *Podsumowaniem*, *Pytania-
mi kontrolnymi* i przypadkami do rozważenia zaprezentowanymi w formie *Studium przypadku*.

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 nieudziałające 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 obsługiwanych 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).