

JACEK MAŚLANKOWSKI

Wydział Zarządzania
Uniwersytet Gdański

Zarządzanie zmianami metodologicznymi w hurtowniach danych

1. Wstęp

Współczesna administracja publiczna, podejmując decyzje, powinna coraz częściej sięgać do informacji pozyskiwanych na podstawie danych zapisanych w źródłach administracyjnych. Wskazane jest, aby dane pozyskane z takich źródeł były przechowywane w hurtowniach danych. Przykładem jest chociażby przewidywanie przez urzędy pracy liczby osób poszukujących pracy, jaka pojawi się w ich regionie np. po zakończeniu roku szkolnego czy akademickiego. Podstawą takiej prognozy mogą być dane administracyjne dotyczące liczby studentów na ostatnim roku studiów według grup i podgrup kierunków studiów. Urzędy pracy, znając zapotrzebowanie na poszczególne grupy zawodów, mogłyby z góry przewidzieć, jaka liczba absolwentów będzie miała problem ze znalezieniem pracy po zakończeniu edukacji.

Do takich analiz potrzebne są informacje zarówno o liczbie absolwentów w ubiegłych latach, jak i o podaży wolnych miejsc pracy i liczbie rejestrujących się absolwentów każdego roku według kierunków. Wydawać się może, że tego typu analiza może być bez przeszkód wykonana przy wykorzystaniu typowego narzędzia klasy *Business Intelligence*. W praktyce jednak nie jest to łatwe zadanie, gdyż wraz z upływem czasu zachodzi wiele zmian metodologicznych, takich jak zmiany klasyfikacji (przesunięcia kierunków pomiędzy grupami kierunków studiów¹) czy też zmiany metodologiczne, czego przykładem jest chociażby

¹ Klasyfikacja kierunków studiów bazuje na międzynarodowej klasyfikacji systemu edukacji ISCED97 i jest co roku aktualizowana poprzez dodawanie nowych kierunków,

definicja bezrobotnego². Hurtownia danych powinna zatem być konstruowana w sposób umożliwiający zarządzanie zmianami metodologicznymi, co zapewni porównywalność dłuższych szeregów czasowych.

Niniejszy artykuł podejmuje ważną kwestię, jaką jest zarządzanie zmianami metodologicznymi w danych, co jest kluczowym aspektem funkcjonowania współczesnych hurtowni danych. Główną tezą opracowania jest założenie, że w tradycyjnej hurtowni danych istnieje możliwość zastosowania rozwiązań pozwalających na obsługę zmian metodologicznych, które zachodzą w danych w niej zawartych. Artykuł ma na celu zaprezentowanie koncepcji architektury hurtowni danych pozwalającej na przechowywanie danych z zachowaniem zasady porównywalności w szeregu czasowym, pomimo zachodzących zmian metodologicznych. Zostało to przedstawione dwuaspektowo – pierwszy aspekt dotyczy częstych zmian zachodzących w klasyfikacjach dostępnych w wymiarach hurtowni, drugi aspekt obejmuje zmiany w definicji pojęć, czyli tzw. metadanych definicyjnych. Zaprezentowano również metody realizacji takiego rozwiązania w bazie danych.

2. Współczesna struktura hurtowni danych

Jak powszechnie wiadomo, hurtownia danych jest źródłem danych w firmie pozwalającym na wykonywanie zapytań³ w celu sporządzania raportów i analiz. Warunkiem koniecznym uznania takiego źródła danych za hurtownię danych jest spełnianie czterech reguł, zwanych również atrybutami hurtowni danych: uporządkowania tematycznego, zawierania wymiaru czasowego, zintegrowania oraz nieulotności danych⁴. Pomimo że definicja ta została sformułowana w 1991 r., a sama koncepcja hurtowni danych istniała jeszcze wcześniej⁵, to wdrażane

usuwanie istniejących bądź przesunięcia kierunków pomiędzy grupami kierunków studiów.

² Definicja „bezrobotnego” w latach 2000–2012 była wielokrotnie aktualizowana. Ponadto inna jest metodologia określania bezrobotnego według różnych badań, np. europejskiego LFS czy narodowego spisu powszechnego, co sprawia, że poziom bezrobocia może przyjmować różne wartości w zależności od przyjętej metodologii jego obliczania.

³ R. Kimball, L. Reeves, M. Ross, W. Thornthwaite, *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing, Inc., Indianapolis 1998, s. 19.

⁴ W.H. Inmon, *Building the Data Warehouse*, Wiley Publishing Inc., Indianapolis 2005, s. 29.

⁵ Na przykład hurtownie informacji biznesowej zaproponowane w 1988 r. przez B. Devlina oraz P. Murphy’ego z firmy IBM Ireland, por. A. Januszewski, *Funkcjonalność informatycznych systemów zarządzania*, t. 2, *Systemy Business Intelligence*, Wydawnictwo Naukowe PWN, Warszawa 2008, s. 21.

obecnie systemy tego typu wciąż bazują na tych czterech atrybutach. Można jedynie stwierdzić, że niewielkie odstępstwa dotyczą kwestii związanych z nie-ulotnością, gdyż dopuszcza się usuwanie danych archiwalnych. Jednak należy zauważyć, że również Bill Inmon, prezentując wczesną koncepcję hurtowni danych, wyodrębnił warstwę danych archiwalnych, do której nie trzeba mieć szybkiego dostępu⁶. Obecnie dane wprowadzane do hurtowni danych są pozyskiwane z systemów ERP (*Enterprise Resource Planning*), CRM (*Customer Relationship Management*), SCM (*Supply Chain Management*) lub innych⁷.

Typowa hurtownia danych zwykle działa w architekturze scentralizowanej lub rozproszonej w postaci wielu hurtowni tematycznych (ang. *data mart*), które mogą być zależne lub niezależne, co wynika ze stopnia ich zintegrowania oraz rodzaju zastosowanego procesu ekstrakcji, transformacji i ładowania – współdzielonego lub niezależnego⁸. Najczęściej wybieranym schematem architektury danych jest gwiazda, rzadziej płatek śniegu, konstelacja gwiazd czy schemat hybrydowy, co ma związek z wysoką wydajnością hurtowni danych w schemacie gwiazdy w stosunku do innych schematów⁹. To, w jakiej architekturze i z jakim schematem nastąpi wdrożenie, nie ma odzwierciedlenia w sposobie traktowania zmian metodologicznych. Koncepcja prezentowana w niniejszym artykule pozwala zatem na zastosowanie jej w dowolnym schemacie i dowolnej architekturze hurtowni danych.

Minimalne założenia dotyczące schematu hurtowni danych nakładają na nią ograniczenia w postaci konieczności wyodrębnienia wymiaru czasu (zgodnie z jednym z czterech podstawowych atrybutów hurtowni). Jednak ze względu na często wykonywane analizy przestrzenne należy również wyodrębnić wymiar terytorialny, w literaturze zwykle znany jako region¹⁰. Dodatkowo, jeżeli hurtownia danych dotyczy wielu zjawisk, należy dodać wymiar, który pozwoli na ich rozróżnienie. Wymiar ten zarazem reprezentuje metadane, czyli informacje o danych przechowywanych w hurtowni. To wszystko prowadzi do powstania

⁶ Porównaj rysunek 2.5 w: W.H. Inmon, op.cit., s. 34.

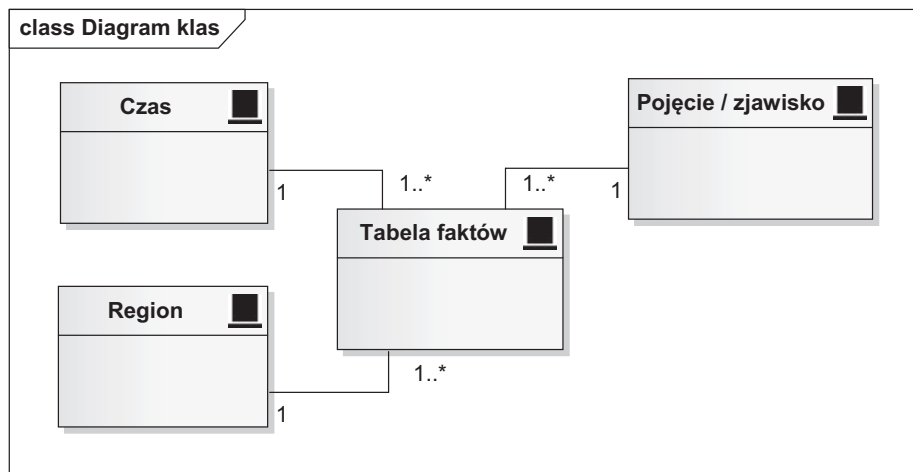
⁷ E. Skrzypek, M. Hofman, *Zarządzanie procesami w przedsiębiorstwie*, Wolters Kluwer Polska, Warszawa 2010, s. 103.

⁸ W.H. Inmon, op.cit., s. 370–371.

⁹ L. Hobbs, S. Hillson, S. Lawande, P. Smith, *Oracle Database 10g Data Warehousing*, Elsevier Ltd, Oxford 2004, s. 29.

¹⁰ Takie typowe wymiary są stosowane przez liczne systemy automatyzujące projektowanie hurtowni danych, por. R. Oliveira, F. Rodrigues, P. Martins, J.P. Moura, *Dimensional Templates in Data Warehouses: Automating the Multidimensional Design of Data Warehouse Prototypes*, w: *Enterprise Information Systems*, red. J. Filipe, J. Cordeiro, Springer-Verlag, Berlin–Heidelberg 2009, s. 190.

przedstawionego na rysunku 1 schematu, który należy uznać za zawierający minimalną liczbę tabel wymiarów.



Rysunek 1. Schemat gwiazdy z minimalną zalecaną liczbą tabel wymiarów

Źródło: opracowanie własne.

Jeżeli dane zawarte w hurtowni dotyczą wielu różnych aspektów związanych z określonymi zjawiskami, np. przechowywania danych o sprzedaży danego produktu w sztukach lub w postaci udziału procentowego w sprzedaży wszystkich produktów ogółem, to należy dodatkowo w schemacie hurtowni danych uwzględnić tabelę wymiaru o nazwie „Jednostka miary”. Nie jest to jednak warunek konieczny, gdyż wiele hurtowni danych wymiar dotyczący jednostki miary zawiera w tabeli wymiarów o nazwie „Pojęcie”.

3. Zarządzanie nieciągłością szeregów czasowych

Wśród istotnych zmian dokonywanych przez organy administracji publicznej należy wyróżnić zmiany terytorialne z 1999 r. (wprowadzenie 16 województw zamiast 49)¹¹, zmiany dokonywane w Polskiej Klasyfikacji Produktów (PKD) w latach 2004 i 2007 czy w Polskiej Klasyfikacji Wyrobów i Usług (PKWiU).

¹¹ Z analizy informacji ze strony <http://www.stat.gov.pl> wynika, że Krajowy Rejestr Urzędowy Podziału Terytorialnego Kraju TERYT jest aktualizowany co najmniej dwa razy w roku. Zmiany te dotyczą głównie podziału gmin, zmian ich rodzajów itp.

W celu zobrazowania zjawiska problemu ciągłości szeregów czasowych w niniejszym artykule jako przykład wykorzystano problem z podziałem terytorialnym kraju na poziomie województw. Jeżeli dane dotyczące sprzedaży do 1999 r. były prezentowane zgodnie z podziałem terytorialnym kraju według 49 województw, a od 1999 r. według 16 województw (zgodnie z nowym podziałem terytorialnym kraju na poziomie NTS2¹²), to istnieją trzy potencjalne rozwiązania problemu związanego z utrzymaniem ciągłości szeregu czasowego:

- 1) przekształcenie danych sprzed 1999 r. na obecnie obowiązujący podział terytorialny,
- 2) przekształcenie danych dotyczących 1999 r. i lat późniejszych na 49 województw,
- 3) pozostawienie bez zmian i analizowanie dwóch szeregów czasowych – pierwszego do 1999 r., drugiego dotyczącego lat późniejszych.

Najbardziej racjonalnym rozwiązaniem wydaje się pierwszy wariant, jednak mogą zaistnieć obawy, że wraz z nowym podziałem terytorialnym kraju zmieniły się również warunki mające wpływ na tę sprzedaż. Przynajmniej w teorii można przypuszczać, że w miastach, które w wyniku nowego podziału terytorialnego kraju utraciły miano stolic województw, sprzedaż będzie niższa, czego przyczyny należy upatrywać w odpływie ludności.

Problemy tego typu są powszechnie znane i zostały zdefiniowane w literaturze na dwa sposoby:

- 1) zgodnie z zasadami retrospekcji,
- 2) z wykorzystaniem reguł dotyczących wolno zmieniających się wymiarów (ang. *slowly changing dimensions* – SCD).

Retrospekcja może być:

- prawdziwa,
- fałszywa lub
- trwała.

Stosując retrospekcję prawdziwą, zakłada się, że obiekt będzie wiernie odwzorowywał przeszłość. Retrospekcja fałszywa jest stosowana, gdy wraz ze zmianami wartości obiektu będzie zmieniany również obraz jego historii. Ostatni typ retrospekcji – retrospekcja trwała – oznacza, że wartość obiektu nie podlega zmianom w czasie¹³.

¹² NTS2 oznacza poziom wojewódzki, zgodnie z nowym podziałem terytorialnym Polski obowiązującym od 1999 r.

¹³ C. Todman, *Designing a Data Warehouse. Supporting Customer Relationship Management*, Hewlett Packard, Upper Saddle River (NJ) 2001, s. 153–157.

W przypadku reguł dotyczących wolno zmieniających się wymiarów najczęściej wyróżnia się trzy podstawowe typy:

- typ 1 – nadpisanie wartości atrybutu – oznacza, że poprzednia wartość atrybutu zostanie nadpisana przez jej nową wartość; tym samym historyczna wartość jest bezpowrotnie tracona; jest to jedno z niezalecanych rozwiązań, gdyż może prowadzić do zaburzeń szeregów czasowych, jeżeli metodologia dotycząca pojęcia ulega zmianie;
- typ 2 – dodanie nowego wiersza wymiaru – jest najpowszechniej wykorzystywaną metodą zarządzania regułami dotyczącymi wolno zmieniających się wymiarów; w ramach tej metody powstaje nowy atrybut z własnym kluczem głównym, a wszelkie poprzednie wartości zbliżone metodologicznie do nowego atrybutu zostają zachowane z niezmienionym kluczem głównym w bazie danych;
- typ 3 – dodanie nowego atrybutu wymiaru – jest jednym z rozwiązań pozwalających na przechowywanie poprzedniej i bieżącej definicji atrybutu (np. zmienionego obszaru województwa) w jednym miejscu; tym samym użytkownik będzie śledził szereg czasowy, posługując się dwiema definicjami dla pojedynczego wiersza.

Wyróżnia się również tzw. hybrydowe wolno zmieniające się wymiary, które łączą cechy typu 1 z cechami typu 2¹⁴.

4. Koncepcja zarządzania zmianami metodologicznymi

Zaprezentowana w niniejszym punkcie koncepcja stanowi jedno z rozwiązań problemu dotyczącego zmian metodologicznych w prezentowanych pojęciach. Zakłada się, że metadane definicyjne będą przechowywane hierarchicznie na zasadzie pojęcia nadrzędnego i podrzędnego.

W dalszej części do zobrazowania problemu nieciągłości szeregów czasowych posłużono się pojęciem szkoły średniej. Posługując się przykładem, pojęciem nadrzędnym nazwano pojęcie pierwotne (np. szkoła średnia), a pojęciem podrzędnym pojęcie wtórne powstające na jego podstawie (np. szkoła ponadgimnazjalna). Przykłady pojęć nadrzędnych i podrzędnych zostały zaprezentowane w tabeli 1.

¹⁴ R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, B. Becker, *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing, Inc., Indianapolis 2008, s. 257–261.

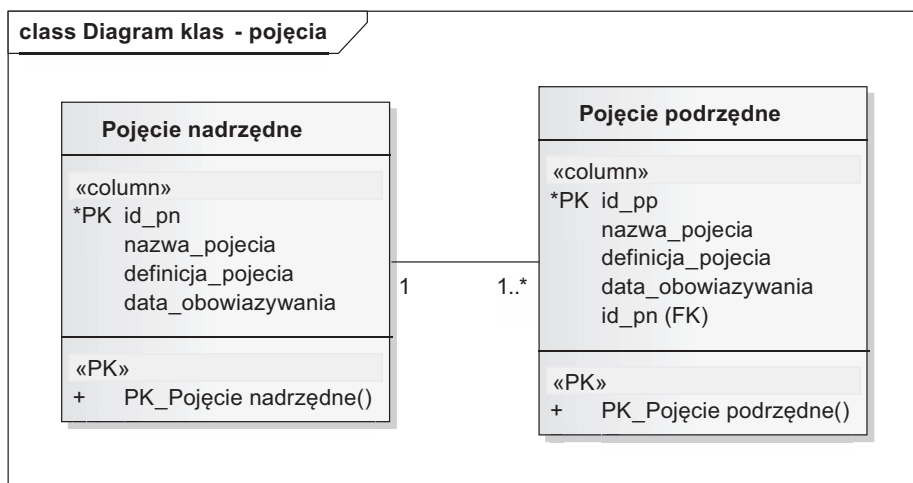
Tabela 1. Przykłady pojęć nadrzędnych i podrzędnych

Pojęcia			
Nadrzędne	Podrzędne	Data obowiązywania	Wersja
Średnia szkoła	Szkoła średnia	1.09.1948–31.08.1999	1
Średnia szkoła	Szkoła ponadgimnazjalna	1.09.1999–	2

Źródło: opracowanie własne.

Przykład ten pokazuje zatem, w jaki sposób mogą zachodzić zmiany metodologiczne. Istotnym punktem jest stwierdzenie, czy można analizować szereg czasowy poprzez prezentowanie danych dotyczących szkół średnich przed 1999 r. i po tym roku.

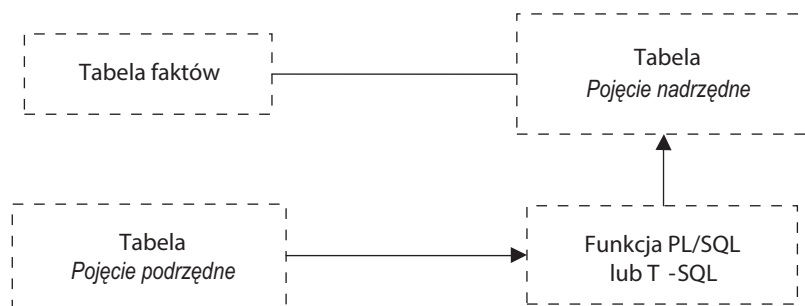
Z technicznego punktu widzenia powyższa propozycja dotycząca zmian metodologicznych może być zrealizowana na wiele sposobów. W niniejszym artykule zaproponowano dwa rozwiązania alternatywne do proponowanych przez przywoływanych w niniejszym artykule teoretyków hurtowni danych i wpasowujące się w funkcjonalności oferowane przez współczesne systemy bazodanowe, będące rdzeniem hurtowni danych. Jednym ze sposobów jest połączenie dwóch tabel na zasadzie tabeli nadrzędnej i podrzędnej. Wówczas zaprezentowany na rysunku 1 schemat gwiazdy zostałby rozszerzony o nową tabelę o nazwie „Pojęcie podrzędne”, połączoną z tabelą „Pojęcie/zjawisko”, którą na rysunku 2 nazwano „Pojęcie nadrzędne”.

**Rysunek 2. Schemat gwiazdy uwzględniający rozszerzony wymiar pojęcia nadrzędnego**

Źródło: opracowanie własne.

Zasadniczą wadą takiego rozwiązania jest zniekształcenie schematu gwiazdy poprzez zastosowanie odwołania do tabeli „Pojęcie nadrzędne” z dwóch miejsc: „Tabeli faktów” oraz „Pojęcia podrzędnego”, w którym występują klucze obce do tabeli „Pojęcie nadrzędne”. Powoduje to pojawienie się wspomnianego na początku artykułu niepotrzebnego narzut związanego z przetwarzaniem tak zapisanego zbioru danych. Należy mieć na uwadze fakt, że baza danych działa wolniej proporcjonalnie do liczby tabel, które są wykorzystywane w zapytaniu. Zatem należałoby poszukać rozwiązania, które nie zmieni istoty schematu gwiazdy, a pozwoli na przyjęcie zasady związanej z pojęciami nadrzędnymi i podrzędnymi.

Takim rozwiązaniem, które warto zastosować w niewielkich hurtowniach danych, jest schemat zgodny z rysunkiem 3. W tym przypadku w hurtowni danych występuje tylko jedno połączenie w schemacie gwiazdy pomiędzy „Tabelą faktów” a tabelą „Pojęcie nadrzędne”. Szczegółowe informacje dotyczące pojęcia w odniesieniu do okresu, jakiego zapytanie dotyczy, znajdują się w tabeli „Pojęcie podrzędne”, która związana jest z tabelą „Pojęcie nadrzędne” poprzez funkcję PL/SQL (Oracle) lub T-SQL (Microsoft SQL Server). Wówczas zapytania na tabeli faktów wykorzystują typowy schemat gwiazdy, natomiast wszelkie szczegóły metodologiczne są pobierane z tabeli „Pojęcie podrzędne” tylko w przypadku takiej konieczności, zdefiniowanej w założeniach do systemu.



Rysunek 3. Schemat gwiazdy uwzględniający rozszerzony wymiar pojęcia

Źródło: opracowanie własne.

Takie technicznie rozwiązanie jest możliwe poprzez zastosowanie typu danych z wbudowaną funkcją PL/SQL lub T-SQL, a następnie zdefiniowanie kolumny tego typu w tabeli „Pojęcie nadrzędne”. Proponowane rozwiązanie może mieć zastosowanie w przypadku, gdy zachodzą zmiany metodologiczne, które nie wykluczają porównywalności poszczególnych pojęć podrzędnych.

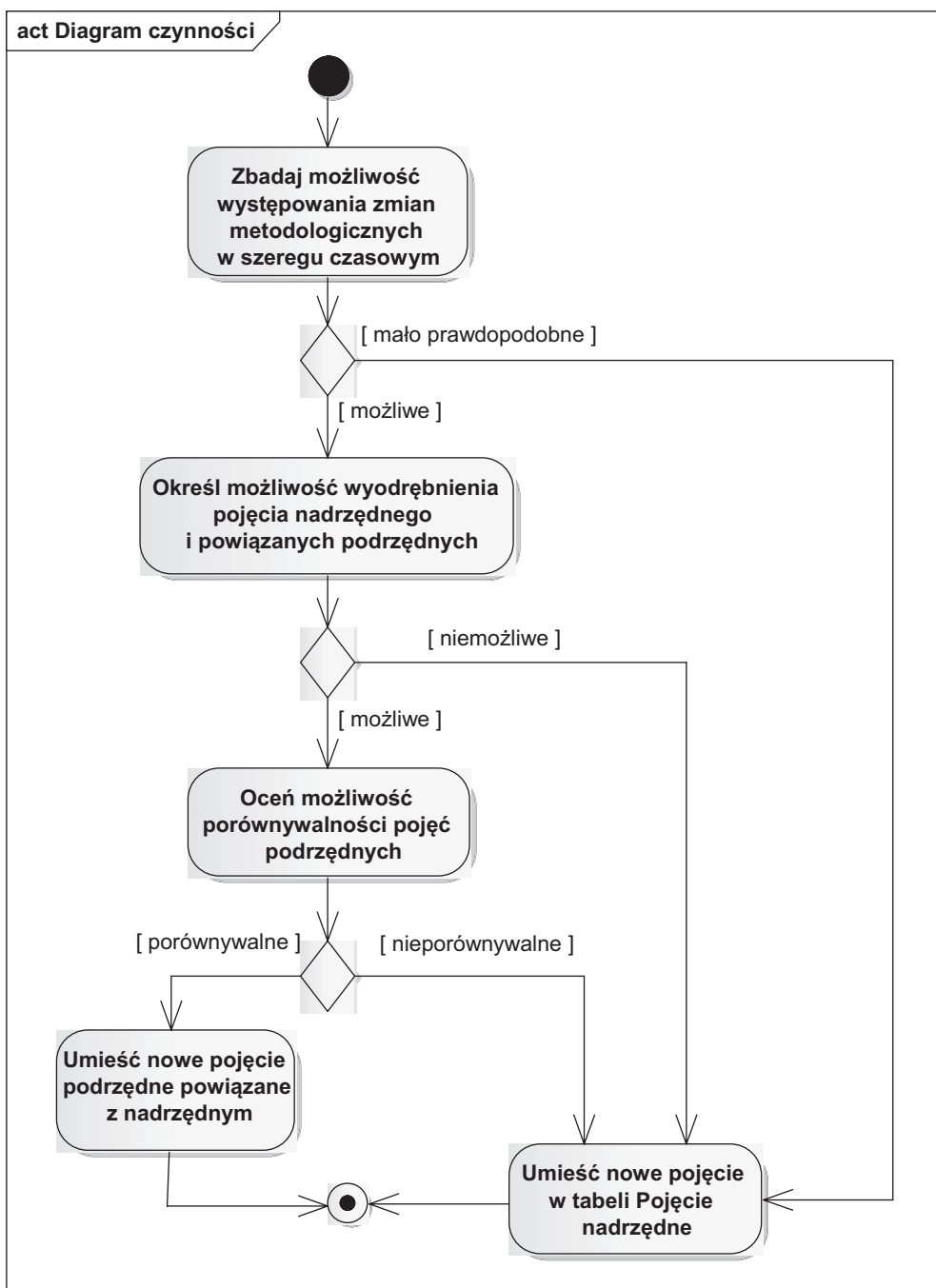
5. Proponowany schemat działania

Proponowany schemat działania dotyczący określania możliwości zastosowania koncepcji związanej z pojęciami nadrzędnymi i podrzędnymi jest stosunkowo prosty. Został on zaprezentowany na rysunku 4. Na początku należy określić, czy w szeregu czasowym, jaki ma się znaleźć w hurtowni danych, mogą zachodzić zmiany metodologiczne. Jeżeli nie przewiduje się dla danego pojęcia występowania zmian metodologicznych, to należy zdefiniować i umieścić w bazie jedynie pojęcie nadrzędne, bez przypisywania do niego pojęć podrzędnych. Przykładem pojęcia będącego zarazem nazwą klasyfikacji, które nie ulegnie zmianie w czasie, jest płeć. Przykładem pojęcia, które będzie się zmieniało w czasie, jest np. szkoła – istnieje prawdopodobieństwo, że w długim okresie powstaną nowe typy szkół, które będą zawierały się w tej definicji. Zatem zmienność będzie zachodziła w klasyfikacji uzupełniającej definicję tego pojęcia.

Jeżeli zmiany metodologiczne mogą pojawić się w pojęciach występujących w szeregu czasowym, wówczas należy określić i wyodrębnić pojęcie nadrzędne oraz powiązane z nim pojęcia podrzędne. Jeżeli takie rozróżnienie nie jest możliwe, wówczas należy utworzyć tylko pojęcie nadrzędne. W innym przypadku należy utworzyć jedno pojęcie nadrzędne i powiązane z nim podrzędne. Przykładem jest chociażby wspomniana wcześniej szkoła średnia i szkoła ponadgimnazjalna. Oba pojęcia mogą dotyczyć tych samych szkół i zostać sklasyfikowane jako jedno pojęcie nadrzędne, z zaznaczeniem w uwagach metodologicznych, że do 30 sierpnia 1999 r. bazuje się na pojęciu podrzędnym o nazwie „szkoła średnia”, a później „szkoła ponadgimnazjalna”.

Kolejną czynnością jest zbadanie porównywalności szeregów czasowych – jeżeli wyodrębnione pojęcia podrzędne są zbliżone metodologicznie, wówczas można zakładać możliwość ich porównywania z zachowaniem ostrożności wynikającej z zachodzących w nich zmian. Następuje zatem na tym etapie zdefiniowanie pojęć podrzędnych i powiązanie ich z nadrzędnymi. W innym przypadku należy utworzyć niezależne pojęcia nadrzędne, które z założenia nie będą porównywalne.

Proponowana koncepcja jest doskonałym rozwiązaniem dla osób wykonujących złożone analizy na danych w bardzo długich szeregach czasowych. Umożliwia wykonywanie analiz bez konieczności ręcznego łączenia poszczególnych pojęć. Dzięki temu rozwiązaniu osoba korzystająca z bazy danych nie musi posiadać pełnej wiedzy na temat zmian metodologicznych. Możliwość porównywania szeregów czasowych jest z góry określana przez architekta danych biorącego udział w projektowaniu hurtowni danych.



Rysunek 4. Czynności prowadzące do określenia zasadności przyjęcia pojęć podrzędnych i nadrzędnego

Źródło: opracowanie własne.

6. Podsumowanie i kierunki dalszych badań

Podsumowując treści zawarte w niniejszym artykule, można wysunąć tezę, że hurtownie danych wdrażane w administracji publicznej mogą cechować się dużo większą złożonością niż ich odpowiedniki w biznesie. Na złożoność tę wpływa mnogość źródeł administracyjnych, jaka może zostać wprowadzona do tego typu systemu. W biznesie natomiast wciąż najczęściej będą występować hurtownie danych opisujące zjawiska związane z działalnością danego podmiotu gospodarczego. Klasycznym tego przykładem i zarazem potwierdzeniem jest wciąż bardzo często występujące w literaturze hurtownie danych, których głównym celem jest analiza sprzedaży.

Zmienność metadanych jest jednym z podstawowych aspektów, jakie należy rozważyć, budując systemy hurtowni danych bazujące na złożonych źródłach administracyjnych. Pomimo że istnieje wiele potencjalnych rozwiązań pozwalających zarządzać tymi zmianami, jak chociażby wspomniane kwestie retrospekcji czy też reguły dotyczące wolno zmieniających się wymiarów, to jednak wiele hurtowni danych nie jest dostosowanych do zbyt częstych zmian metodologicznych, co prowadzi do konieczności wykonywania analiz danych na osobnych szeregach czasowych. Dlatego bardzo istotne jest zastosowanie rozwiązania zarządzania zmianami metodologicznym, którego istota została opisana w niniejszym artykule wraz z propozycją dotyczącą jego implementacji.

Bibliografia

1. Hobbs L., Hillson S., Lawande S., Smith P., *Oracle Database 10g Data Warehousing*, Elsevier Ltd, Oxford 2004.
2. http://www.stat.gov.pl/bip/36_PLK_HTML.htm [dostęp 02.09.2012].
3. Inmon W.H., *Bulding the Data Warehouse*, Wiley Publishing, Inc., Indianapolis 2005.
4. Januszewski A., *Funkcjonalność informatycznych systemów zarządzania*, t. 2, *Systemy Business Intelligence*, Wydawnictwo Naukowe PWN, Warszawa 2008.
5. Kimball R., Reeves L., Ross M., Thornthwaite W., *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing, Inc., Indianapolis 1998.
6. Kimball R., Ross M., Thornthwaite W., Mundy J., Becker B., *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing, Inc., Indianapolis 2008.

7. Oliveira R., Rodrigues F., Martins P., Moura J.P., *Dimensional Templates in Data Warehouses: Automating the Multidimensional Design of Data Warehouse Prototypes*, w: *Enterprise Information Systems*, red. J. Filipe, J. Cordeiro, Springer-Verlag, Berlin–Heidelberg 2009.
8. Skrzypek E., Hofman M., *Zarządzanie procesami w przedsiębiorstwie*, Wolters Kluwer Polska, Warszawa 2010.
9. Todman C., *Designing a Data Warehouse. Supporting Customer Relationship Management*, Prentice Hall, Upper Saddle River (NJ) 2001.

* * *

Methodological changes management in data warehouses

Summary

Nowadays the data warehouse is a complex system that is built to gather data from several subjects of the enterprise orientation. As there is lots of information, several different variables must exist in such environment. The article shows how to manage data warehouse that is exposed to several methodological issues. It is presented how typical methodological changes should be treated, especially in the context of retrospection issues and different types of slowly changing dimensions.

Keywords: data warehouse, business intelligence, metadata management