

Marcin Hernes, Andrzej Bytniewski

Uniwersytet Ekonomiczny we Wrocławiu

e-mails: marcin.hernes@ue.wroc.pl; andrzej.bytniewski@ue.wroc.pl

JAKOŚĆ DANYCH W ZINTEGROWANYCH SYSTEMACH INFORMATYCZNYCH ZARZĄDZANIA

DATA QUALITY IN INTEGRATED MANAGEMENT INFORMATION SYSTEMS

DOI: 10.15611/pn.2017.471.15

JEL Classification: M41, C88

Streszczenie: Zintegrowane systemy informatyczne zarządzania pozyskują dane z wielu źródeł wewnętrznych i zewnętrznych na świecie, a więc dane te mogą posiadać różną jakość. Skuteczne zarządzanie musi jednakże bazować na danych o wysokiej jakości. Ważnymi cechami danych są pilność i aktualność, niezaprzeczalność, zrozumiałość i wiarygodność. Tylko na podstawie takich danych osoby zarządzające mogą podejmować prawidłowe decyzje ekonomiczne. Celem niniejszego artykułu jest analiza problematyki związanej z jakością danych i zarządzaniem tą jakością w zintegrowanych systemach informatycznych zarządzania. W pierwszej części przedstawiono istotę pojęcia jakości danych. Następnie scharakteryzowano problematykę zarządzania jakością danych w zintegrowanych systemach informatycznych zarządzania. Ostatnią część artykułu stanowi dyskusja nad praktycznymi aspektami zarządzania jakością danych.

Słowa kluczowe: jakość danych, zintegrowane systemy informatyczne zarządzania, zarządzanie jakością danych.

Summary: Integrated management support systems acquire data from multiple internal and external sources in the world, so the data may have different quality. However, effective management has to be based on high quality data. The important features of the data are actuality, non-repudiation, comprehensibility and credibility. Only on the basis of such data managers can take the right economic decisions. The aim of this paper is to analyze issues related to data quality and management of the quality in integrated management information systems. The first part presents the essence of the concept of data quality. Next, the problem of data quality management in integrated management information systems has been characterized. The last part of the paper is a discussion on the practical aspects of data quality management.

Keywords: data quality, integrated management information systems, data quality management.

1. Wstęp

Jakość danych jest istotnym elementem prawidłowego funkcjonowania organizacji gospodarczej. Tylko na podstawie danych o odpowiedniej jakości osoby zarządzające mogą podejmować prawidłowe decyzje. Szczególnego znaczenia nabiera jakość danych zintegrowanych w systemach informatycznych zarządzania (ZSIZ). Systemy te pozyskują dane z wielu źródeł wewnętrznych i zewnętrznych na świecie, a więc dane te mogą posiadać różną jakość. Jakość mierzona jest w ZSIZ poprzez wykorzystanie funkcji użyteczności. Istotnym problemem jest to, że prawdziwa użyteczność jest znana dopiero w momencie, gdy znane są skutki decyzji podjętej z wykorzystaniem tych danych. Ważnymi cechami danych są pilność i aktualność, prawidłowa decyzja może być bowiem podjęta tylko wtedy, gdy system potrafi szybko pozyskać aktualne dane potrzebne do podjęcia tej decyzji. Kolejnymi własnościami danych są niezaprzeczalność, zrozumiałość i wiarygodność. Często dane uznaje się za wiarygodne, jeżeli można je potwierdzić w wielu źródłach.

Celem niniejszego artykułu jest analiza problematyki związanej z jakością danych i zarządzaniem tą jakością w zintegrowanych systemach informatycznych zarządzania. W pierwszej części przedstawiono istotę pojęcia jakości danych. Następnie scharakteryzowano problematykę zarządzania jakością danych w ZSIZ. Ostatnią część artykułu stanowi analiza praktycznych aspektów zarządzania jakością danych.

W celu realizacji badań przedstawionych w artykule zastosowano takie metody badawcze, jak studia literaturowe, obserwacja zjawisk, synteza i analiza.

2. Istota jakości danych

Odpowiednia jakość danych jest warunkiem koniecznym (ale nie wystarczającym) prawidłowego funkcjonowania zintegrowanych systemów informatycznych zarządzania. W literaturze przedmiotu pojęcie jakości danych definiowane jest na wiele sposobów. W pracy [Błaszczuk, Knosala 2006] jakość danych określana jest na przykład jako kwalifikacja poprawności danych i ich przydatności.

Maślankowski [2015] stwierdza, że jakość danych definiowana jest przez takie kryteria, jak:

- dokładność,
- przydatność danych, oznaczająca ich relewantny charakter dla odbiorcy,
- aktualność danych¹, czyli to, że powinny być dostarczane w określonym czasie.

Zgodnie z europejskim systemem statystycznym jakość danych określona jest przez następujące komponenty [Trzęsiok 2014]:

- przydatności,
- dokładności,

¹ Opis problematyki dotyczącej zwiększenia aktualności danych zawarty jest w pracy [Bytniewski 2009].

- terminowości² i punktualności³,
- dostępności i przejrzystości,
- porównywalności,
- spójności.

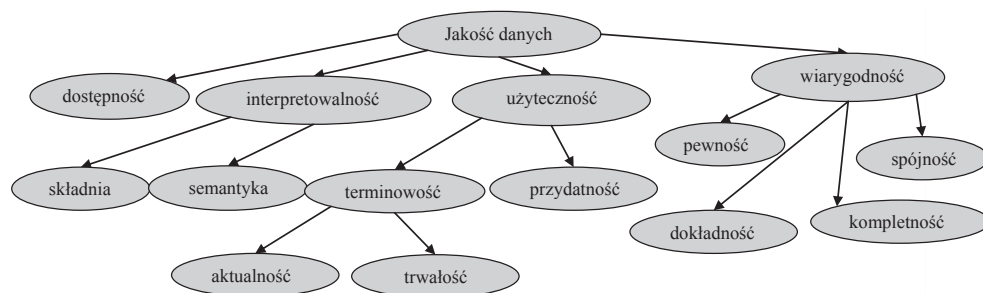
Na cechy jakościowe danych ma wpływ złożoność wielu problemów, do których można zaliczyć (por. [Kes 2013, s. 263]):

- dane przychodzące zbyt późno, a dotyczą zdarzeń, które wcześniej zaistniały, i nie mamy już wpływu na przebieg zaistniałych procesów,
- dane są zbyt szczegółowe, choć w przypadku systemów informatycznych nie należy tego traktować jako wielkiej wady, gdyż można te dane łatwo automatycznie zagregować,
- dane są zbyt rozległe, czyli obejmują dość duży okres czasowy,
- dane są ukierunkowane na przyszłość,
- dane mają charakter tylko kwantyfikatywny,
- dane wzajemnie się wykluczają,
- dane dla celów strategicznych są niejasne.

W analityce biznesowej jakość danych jest utożsamiana z kompletnością i spójnością jako centralnymi wymiarami jakości danych systemów analitycznych [Kwon i in. 2014]. Rozpatrywanie jakości danych biznesowych powinno odbywać się w trzech etapach [Haug i in. 2014; Weichbroth 2009]:

- na wejściu,
- podczas przetwarzania,
- na wyjściu.

Lee i in. [2006] określają jakość danych jako hierarchię atrybutów (rys. 1).



Rys. 1. Hierarchia atrybutów jakości danych

Źródło: opracowanie własne na podstawie [Lee i in. 2006].

² Terminowy (z łac. *terminus*, czyli kres, granica), mający swój termin, dokładnie określoną chwilę zakończenia danego działania [Pszczółowski 1978, s. 253].

³ Punktualny – działania podmiotu są zsynchronizowane przez ten podmiot z czynami innych podmiotów. Punktualność odnosi się do każdej fazy działania [Pszczółowski 1978, s. 201].

Autorzy podkreślają, że w relacji tej atrybut nadrzędny jest prawdziwy (spełniony), jeśli prawdziwe są wszystkie atrybuty podrzędne (iloczyn logiczny). Na przykład, dane są interpretowalne, jeśli mają prawidłową składnię i semantykę.

Można zatem stwierdzić, że jakość danych jest pojęciem o bardzo szerokim znaczeniu. Określenie atrybutów (jak i ich hierarchii), które muszą być spełnione przy analizowaniu jakości danych, odbywa się najczęściej na podstawie charakteru sytuacji decyzyjnej. Przykładowo, dane, na podstawie których podejmowana jest decyzja o działalności marketingowej, nie muszą być w pełni spójne lub kompletne, natomiast w odniesieniu do danych dotyczących sterowania produkcją niespójność, czy też niekompletność jest niedopuszczalna, gdyż mogłaby doprowadzić do uszkodzenia maszyn/urządzeń na linii produkcyjnej, a nawet do wypadku przy pracy.

3. Zarządzanie jakością danych

Proces zarządzania jakością danych składa się z następujących faz [Geiger 2004; Kwon i in. 2014]:

1. Profilowanie danych – jest to proces analizy i zrozumienia istniejących danych. Dane, które już istnieją w systemie, muszą być prawidłowo interpretowane i wykorzystywane. Faza profilowania danych stanowi zatem punkt wyjścia do realizacji kolejnych faz. W wyniku realizacji tej fazy uzyskuje się odpowiedź, czy dane są kompletne i dokładne, czyli:

- posiadają odpowiedni zakres (np. wiek osoby 200 lat jest spoza odpowiedniego zakresu),
- są ustandaryzowane (przykładowo, jeśli wzrost jednej osoby jest zapisany w centymetrach, a wzrost innej osoby w metrach, to dane nie są ustandaryzowane),
- mają poprawne wartości (np. jeśli płeć może przybierać tylko wartości „K” lub „M”, a dane zawierają wartość „L”, to ich wartość jest niepoprawna),
- są zgodne z ustawieniami regionalnymi i przyjętymi formatami (przykładowo, w różnych regionach świata występują różne formaty daty – może więc występować ich niezgodność, mogą występować również różne waluty),
- występuje spójność powiązań pomiędzy danymi (np. jeżeli dane zawierają niepoprawny kod pocztowy miejscowości, to powiązania są niespójne).

2. Jakościowanie pozyskanych danych. Na podstawie wyników uzyskanych w poprzedniej fazie zbudowane są reguły poprawności w odniesieniu do pozyskanych danych. Następnie sprawdza się, czy pozyskane dane są zgodne z tymi regułami. W wyniku realizacji tej fazy podejmowane są następujące działania:

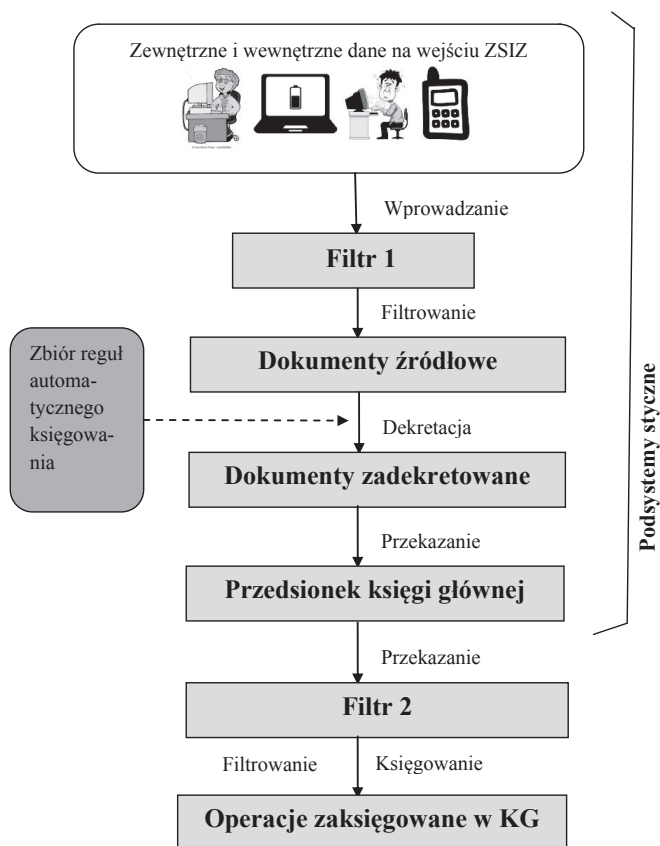
- wykluczenie danych; jeśli błąd w danych jest zbyt duży, najlepszym podejściem może być usunięcie tych danych,
- akceptacja danych; nawet jeśli wiadomo, że istnieją błędy w danych, ale są one w granicach tolerancji; czasem najlepszym rozwiązaniem jest, aby zaakceptować dane z błędem,
- skorygowanie danych; jeżeli wiadomo, w jaki sposób można to wykonać,

- wstawianie wartości domyślnej; czasami ważne jest, aby pole posiadało jakąś wartość (np. nieznaną).

3. Integracja danych. Dane dotyczące tego samego obiektu lub zjawiska zachodzącego w świecie rzeczywistym przechowywane w różnych bazach danych mogą się różnić. W celu integracji danych mogą być stosowane różne metody, na przykład [Hernes, Sobieska-Karpińska 2016]:

- wyboru,
- negocjacji,
- consensusu.

4. Augmentacja danych (zwiększenie możliwości analitycznych danych). Jest to ostatnia faza zarządzania jakością danych. Dane konwertowane są do postaci umożliwiających dokonywanie głębszej analizy biznesowej (na przykład tworzone są kostki OLAP umożliwiające wielowymiarową analizę danych).



Rys. 2. Procedury zarządzania jakością danych w procesie przetwarzania danych (szczególnie księgowych) przez ZSIZ

Źródło: opracowanie własne.

Zintegrowane systemy informatyczne zarządzania posiadają wbudowane procedury zarządzania jakością danych, które są realizowane w procesie przetwarzania tych danych (rys. 2). Dane do podsystemów stycznych mogą być wprowadzane przez użytkowników lub z urządzeń elektronicznych (internet rzeczy), mogą być to dane wewnętrzne lub zewnętrzne. Wprowadzone dane są filtrowane⁴ (Filtr 1). Przykładowo, sprawdzane są wartości graniczne wprowadzonych danych. Po prze-filtrowaniu dane te stają się dokumentami źródłowymi. Następnie wykonywana jest automatyczna dekretycja tych dokumentów na podstawie reguł księgowania⁵ i dokumenty te przekazywane są do tzw. przedsiionka księgi głównej. W dalszej kolejności dokonywane jest filtrowanie tych dokumentów (Filtr 2) pod względem poprawności zapisów syntetycznych w księdze głównej (KG) i ich zaksięgowanie.

4. Aspekty praktyczne zarządzania jakością danych

W rozwiązaniach praktycznych zarządzanie jakością danych odbywa się przede wszystkim za pomocą procedur/warunków wykonywanych automatycznie przez ZSIZ⁶. Systemy funkcjonujące w dużych i średnich przedsiębiorstwach mogą zawierać nawet około 200 takich procedur/warunków (oprócz warunków implementowanych w module sterowania produkcją, których mogą być tysiące lub nawet setki tysięcy w zależności od poziomu skomplikowania linii technologicznej).

Większość błędów danych jest usuwana w procesie pozyskiwania danych. Należy wyraźnie podkreślić, że dane pozyskiwane z otoczenia wewnętrznego przedsiębiorstwa (na przykład wprowadzane przez pracowników) są sprawdzane i jakościowane już podczas procesu wprowadzania danych. System umożliwia wprowadzenie tylko ustandaryzowanych danych (przykładowo, jeśli jednostką miary jest sztuka, to system umożliwia wprowadzenie tej jednostki jako „szt.” i pracownik nie ma możliwości wpisania „sztuka” itp.).

Wykluczenie danych w ZSIZ może występować jedynie w odniesieniu do danych zewnętrznych, pozyskiwanych np. z cyberprzestrzeni. Przykładowo, można wziąć pod uwagę opinie klientów, pozyskiwane z portali społecznościowych, o produkcie oferowanym przez przedsiębiorstwo. Jeżeli w wyniku automatycznej analizy przez

⁴ J. Kisielnicki wskazuje, że między nadawcą i odbiorcą istnieją filtry (fizyczny, semantyczny, pragmatyczny), które nie przepuszczają całej informacji. Narzędzia informatyczne najczęściej minimalizują straty powstałe w tym procesie przesyłania informacji. Wskazuje on, że drogą prowadzącą do zmniejszenia strat są m.in. rozwiązania wykorzystujące słowniki (tezauryusy) i reguły [Kisielnicki 2013, s. 62–63]. W naszym przypadku mają zastosowanie oba rozwiązania.

⁵ Szczegółowe rozwiązania w zakresie reguł księgowania można znaleźć w pracach [Bytniewski 1996, s. 86–89] i [Bytniewski 2012, s. 87–90].

⁶ Zdaniem Z. Kesa problemu z porządkowaniem danych, czytaj zarządzaniem jakością danych, w zintegrowanych systemach informatycznych zarządzania (ZSIZ) nie ma [Kes 2013, s. 265]. Jednak wynika to z tego, że w ZSIZ przeprowadzane są automatyczne kontrole danych wykorzystujące mechanizmy zarządzania jakością danych (filtry), a tym samym podnoszące ich wiarygodność.

system 100 opinii okaże się, że 99 jest pozytywnych, a 1 negatywna, to system może automatycznie wykluczyć tę negatywną opinię.

W podsystemie zarządzania produkcją błędy w danych są niedopuszczalne. Przykładowo, w procesie sterowania produkcją błąd chociażby jednego bitu danych może spowodować bardzo negatywne skutki nie tylko w aspekcie biznesowym i technologicznym, ale również w aspekcie zdrowia i życia ludzkiego (na przykład nieprawidłowe sterowanie maszyną może doprowadzić do wypadku przy pracy). Bardzo często zmiana stanu wyjścia binarnego sterownika wymaga sprawdzenia kilkudziesięciu warunków (stanu czujników, aktualnego stanu linii produkcyjnej itp.).

Błędy danych w ZSIZ mogą wystąpić jedynie przez pewien krótki czas po ich pozyskaniu, dopóki nie zostaną one sprawdzone w sposób automatyczny lub przez człowieka. Jako przykład można podać podsystem finansowo-księgowy, w którym kwoty wszystkich księgowanych dokumentów muszą się zgadzać ze stanem faktycznym. Jeśli nawet wystąpi taki błąd przy wprowadzaniu danych, który nie może być wykryty na „wejściu” przez system, to zostanie on jednak zidentyfikowany w dalszym etapie przetwarzania. Na przykład wprowadzenie błędnej wartości faktury VAT zakupu materiału zostanie ujawnione w przypadku konfrontacji jej z dokumentem przyjęcia materiałów (PZ).

Szczególnego znaczenia nabiera kontrola danych pozyskiwanych z podsystemów stycznych (środków trwałych, logistyki, zarządzania zasobami ludzkimi) z podsystemem finansowo-księgowym, a konkretnie księgi głównej. Dane uzyskiwane z tych podsystemów podlegają automatycznej różnorodnej kontroli według wcześniej zdefiniowanych reguł. Przykładami takich reguł mogą być w zakresie danych dotyczących amortyzacji środków trwałych symbole kont syntetycznych i analitycznych oraz miejsc ich użytkowania. Powyższe symbole muszą być zgodne z symboliką kont księgi głównej, w przeciwnym razie proces automatycznego księgowania jest wstrzymywany i wymaga interwencji człowieka w celu sprostowania błędnie zadeklarowanego konta w podsystemie środków trwałych, bądź może to być przypadek, że błąd wystąpił w księdze głównej, gdyż nie było wcześniej takie konto utworzone.

Innym przykładem kontroli pozyskiwania (księgowania) danych z podsystemu logistyki do księgi głównej jest sprawdzenie prawidłowości przypisania symbolu konta syntetycznego i analitycznego do pozycji kartoteki towarowej, gdyż dane definiujące wzorzec księgowania zawarte w tej kartotece są podstawą automatycznego utworzenia dekretu dotyczącego przyjęcia towaru. Na przykład, gdy pewna grupa towarowa (np. soki) jest księgowana na konkretnym koncie analitycznym towarów, a we wzorcu kartoteki w trakcie jej zakładania wpisano błędnie symbol piwa, wówczas wcześniej zdefiniowane reguły w fazie profilowania analizujące ten dekret (na wejściu do księgi głównej) wykażą, że wystąpił błąd w automatycznie sporządzonym dekrecie w podsystemie logistyki. W takiej sytuacji księgowanie dokumentu przyjęcia towaru w księdze głównej jest wstrzymywane i wymaga korekty. Korekty tej może dokonać program, ale tylko wtedy, gdy wystąpił już taki błąd i jest on już

ujęty w regułach poprawności, w przeciwnym wypadku wymagana jest interwencja człowieka prostującego ten błąd.

Kolejnym przykładem badania poprawności danych z podsystemów stycznych księgowanych w księdze głównej jest kontrolowanie dokumentów pod względem wartości granicznych. Gdy są one o wartościach zerowych bądź przekraczają wartości minimalne lub maksymalne, nie są księgowane i wymagają poprawy lub eliminacji.

5. Zakończenie

Jakość danych jest kluczowym elementem prawidłowego funkcjonowania zintegrowanych systemów informatycznych zarządzania. Istnieją różnorodne definicje pojęcia jakości danych oraz różne podejścia do procesu zarządzania tą jakością. Ich wspólnym mianownikiem jest końcowy efekt przetwarzania danych, czyli dostarczenie przez system danych umożliwiających podejmowanie szybkich, skutecznych i efektywnych decyzji ekonomicznych. Należy wyraźnie podkreślić, że duża liczba procedur kontroli danych wbudowanych w ZSIZ pozwala zapewnić odpowiednią jakość danych.

Dalsze prace badawcze mogą dotyczyć na przykład analizy jakości danych w systemach przetwarzających dane niekompletne, niepewne i niespójne, takich jak systemy wieloagentowe, jak również analizy problematyki wartości informacji i wiedzy przechowywanej w ZSIZ.

Literatura

- Błaszczuk K., Knosala R., 2006, *Problem jakości danych w hurtowniach*, Prace Naukowe Akademii Ekonomicznej w Katowicach, *Systemy wspomaganie organizacji SWO 2006*, s. 45–54.
- Bytniewski A., 1996, *Założenia teoretyczne robotyzacji system rachunkowości*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Bytniewski A., 2009, *Problemy opóźnień przepływu informacji z podsystemów stycznych do podsystemu finansowo-księgowego w zintegrowanych systemach informatycznych zarządzania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 55.
- Bytniewski A., 2012, *Robotyzacja systemu rachunkowości jako proces wspomaganie rachunkowości zarządczej i controllingu*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 251, *Rachunkowość a controlling*, red. E. Nowak, M. Nieplowicz.
- Das T.K., Kumar P.M., 2013, *BIG Data Analytics: A Framework for Unstructured Data Analysis*, International Journal of Engineering Science & Technology, vol. 5, iss. 2, February, s. 153–156.
- Geiger J., 2004, *Data quality management: The most critical initiative you can implement*, The Twenty-Ninth Annual SAS User Group International Conference.
- Haug A., Arlbjorn J.S., Zachariassen F., Schlichter J., 2014, *Master data quality barriers: an empirical investigation*, Industrial Management & Data Systems, vol. 113, no. 2, s. 234–249.
- Hazen B., Boone C., Ezell J., Jones-Farmer L.A., 2014, *Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications*, International Journal of Production Economics, vol. 154, August, s. 72–80.

- Hernes M., Sobieska-Karpińska J., 2016, *Application of the consensus method in a multiagent financial decision support system*, Information Systems and e-Business Management, vol. 14(1), s. 167–185.
- Kes Z., 2013, *Informatyczne wspomaganie controllingu*, [w:] *Controlling dla menedżerów*, red. E. Nowak, wyd. 1, CeDeWu.pl, Warszawa.
- Kisielnicki J., 2013, *Systemy informatyczne zarządzania*, wyd. 1, Placet, Warszawa.
- Kwon O., Lee N., Shin B., 2014, *Data quality management, data usage experience and acquisition intention of big data analytics*, International Journal of Information Management, vol. 34, iss. 3, June, s. 387–394.
- Lee Y.W., Pipino L., Funk J.D., Wang R.Y., 2006, *Journey to Data Quality*, MIT Press.
- Maślankowski J., 2015, *Analiza jakości danych pozyskiwanych ze stron internetowych z wykorzystaniem rozwiązań Big Data*, Roczniki Kolegium Analiz Ekonomicznych, z. 38, s. 167–177.
- Pszczółowski T., 1978, *Mata encyklopedia prakseologii i teorii organizacji*, Ossolineum, Wrocław.
- Shankaranarayanan G., Cai Y., 2006, *Supporting data quality management in decision-making*, Decision Support Systems, vol. 42, iss. 1, s. 302–317.
- Trzęsiok M., 2014, *O jakości danych w kontekście obserwacji oddalonych w wielowymiarowej analizie regresji*, Studia Ekonomiczne. Zastosowania metod matematycznych w ekonomii i zarządzaniu, nr 191, red. J. Mika, K. Zeug-Żebro, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, s. 75–88.
- Weichbroth P., 2009, *Odkrywanie reguł asocjacyjnych z transakcyjnych baz danych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Informatyka Ekonomiczna, nr 82, *Rynek usług informatycznych*, s. 301–309.