

Tác động của chất lượng dữ liệu xấu đến kết quả hoạt động kinh doanh của doanh nghiệp

Trịnh Hoàng Hồng Huệ¹, Trần Thiện Vũ^{2,*}



Use your smartphone to scan this QR code and download this article

TÓM TẮT

Trong thời đại số, dữ liệu được xem là chủ thể trung tâm của các hoạt động kinh doanh của doanh nghiệp. Bên cạnh yếu tố nhân lực và tài lực, dữ liệu được xem là "tài sản" quý giá của doanh nghiệp quyết định mức độ thành công của hoạt động kinh doanh. Bởi vì chất lượng dữ liệu "xấu" (poor data quality) ở đầu vào sẽ có tác động tiêu cực đến chất lượng đầu ra, đó chính là chất lượng của quyết định ở cấp độ cá nhân, tổ chức, và doanh nghiệp. Chất lượng dữ liệu xấu sẽ là rào cản cho doanh nghiệp hướng đến việc xây dựng chiến lược có hiệu lực và quản lý có hiệu quả xuyên chức năng kinh doanh của doanh nghiệp đó. Vấn đề quản lý chất lượng của dữ liệu (data quality management) đã được các doanh nghiệp và Đại học tại các quốc gia phát triển như Mỹ, Pháp, Thụy Sĩ, Úc, Hàn Quốc quan tâm, phát triển, ứng dụng và cải thiện quy trình quản lý chất lượng dữ liệu từ những năm 1990s. Tuy nhiên vấn đề này các doanh nghiệp cũng như trường học tại Việt Nam vẫn còn ở dạng sơ khởi và mới mẻ. Mục tiêu của bài báo này là nêu lên tổng quan về chất lượng dữ liệu, các vấn đề của chất lượng dữ liệu, các mô hình đã được nghiên cứu, tác động của chất lượng dữ liệu xấu đến kết quả hoạt động kinh doanh của doanh nghiệp, bằng phương pháp định tính – tổng hợp các tài liệu và sử dụng thông tin thứ cấp. Từ đó tác giả đưa ra những thảo luận và hàm ý chính sách quản trị về vấn đề chất lượng dữ liệu trong bối cảnh ở Việt Nam.

Từ khoá: Chất lượng dữ liệu, hoạt động kinh doanh, doanh nghiệp Việt Nam

¹Trường Đại học Kinh tế – Luật, ĐHQG HCM, Việt Nam

²Trường Đại học Công nghệ Thông tin và Truyền thông Việt-Hàn, Đại học Đà Nẵng, Việt Nam

Liên hệ

Trần Thiện Vũ, Trường Đại học Công nghệ Thông tin và Truyền thông Việt-Hàn, Đại học Đà Nẵng, Việt Nam

Email: vutrandn@gmail.com; ttvu@vku.udn.vn

Lịch sử

- Ngày nhận: 24-9-2022
- Ngày chấp nhận: 07-6-2023
- Ngày đăng: 30-6-2023

DOI:

<https://doi.org/10.32508/stdjelm.v7i2.1224>



Bản quyền

© ĐHQG Tp.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



ĐẶT VẤN ĐỀ

Ngày nay, dữ liệu được xem là "tài sản" quan trọng của tổ chức. Bởi vì, các hoạt động trong mỗi tổ chức đều liên quan đến việc sử dụng dữ liệu như là một yếu tố đầu vào cho các hoạt động ra quyết định ở các cấp độ từ tác nghiệp, chiến thuật cho đến cấp độ chiến lược. Chính vì vậy dữ liệu "xấu" (poor data) ở đầu vào sẽ tác động tiêu cực đến chất lượng đầu ra, chính là chất lượng của quyết định ở cấp độ cá nhân, tổ chức, và doanh nghiệp. Mặc dù chất lượng dữ liệu (data quality) là một khái niệm trừu tượng, chưa được nhiều cá nhân và tổ chức biết đến và quan tâm nhiều, nhưng chất lượng dữ liệu lại quyết định đến sự thành công của một doanh nghiệp¹. Chất lượng dữ liệu xấu (poor data quality) có thể tác động tiêu cực đến hiệu quả vận hành và quản lý chung của một tổ chức². Cụ thể, chất lượng dữ liệu xấu sẽ là rào cản cho tổ chức hướng đến xây dựng chiến lược có hiệu lực và quản lý có hiệu quả xuyên chức năng kinh doanh của tổ chức³. Loshin nêu ra tác động tích cực của dữ liệu tốt đến việc khai thác tốt kho dữ liệu, tăng khả năng thu hút khách hàng mới, tăng tỷ suất hoàn vốn (ROI – Return on Investment) trong đầu tư dự án công nghệ thông tin; ngược lại, tác động tiêu cực của dữ liệu xấu đến vận hành kém hiệu quả, ra quyết định sai lầm, đưa ra các kết

luận không chính xác, giảm tỷ lệ giữ chân khách hàng, giảm độ tin cậy của tổ chức³. Tương tự với quan điểm của Loshin, bản báo cáo về chất lượng dữ liệu năm 2012 cho rằng dữ liệu xấu có thể đưa đến những quyết định sai lầm, gây thiệt hại lớn cho doanh nghiệp cũng như dẫn tới dẫn đến sự lãng phí tài nguyên, hiệu quả dịch vụ thấp và gia tăng chi phí cao trong mọi lĩnh vực⁴.

Về cốt lõi, dữ liệu đo lường các hoạt động kinh doanh bên ngoài và bên trong của doanh nghiệp. Do đó, chất lượng dữ liệu xấu sẽ gây ảnh hưởng đáng kể đến các doanh nghiệp. Theo kết quả khảo sát từ công cụ chất lượng dữ liệu nguồn mở Great Expectations với 500 chuyên gia, có 77% doanh nghiệp gặp vấn đề về chất lượng dữ liệu. Trong đó, có 91% báo cáo rằng chất lượng dữ liệu xấu có ảnh hưởng đến kết quả hoạt động doanh nghiệp, và chỉ 11% báo cáo không có vấn đề liên quan đến chất lượng dữ liệu. Nghiên cứu do Gartner thực hiện cho thấy chi phí trung bình hàng năm mà các công ty phải gánh chịu do chất lượng dữ liệu xấu là khoảng gần 13 triệu đô la mỗi năm⁵.

Trong thời đại số, dữ liệu được xem là đối tượng trung tâm của các hoạt động giao dịch kinh doanh và thương mại toàn cầu. Dữ liệu là tài sản quý giá của doanh nghiệp, bên cạnh các yếu tố nhân lực và tài lực. Khối tài sản dữ liệu này quyết định mức độ thành

Trích dẫn bài báo này: Huệ T H H, Vũ T T. Tác động của chất lượng dữ liệu xấu đến kết quả hoạt động kinh doanh của doanh nghiệp. *Sci. Tech. Dev. J. - Eco. Law Manag.*; 7(2):4381-4392.

công của doanh nghiệp trong kinh doanh. Kể từ đầu thế kỷ 21, những thay đổi vượt bậc trong ngành công nghệ thông tin với sự xuất hiện của điện toán đám mây, Internet vạn vật và mạng xã hội đã khiến lượng dữ liệu tăng liên tục và tích lũy với tốc độ nhanh chóng mặt, từ đó hình thành nên Big Data – Dữ liệu lớn⁶. Các vấn đề về chất lượng dữ liệu mà nhiều doanh nghiệp gặp phải hiện nay có thể liên quan đến sự phát triển công nghệ vì khi khối lượng dữ liệu tăng lên, sự phức tạp của việc quản lý dữ liệu cũng tăng theo⁷. Mặc dù chất lượng dữ liệu cao là điều kiện tiên quyết để đạt được hiệu quả của doanh nghiệp nhưng cho đến nay vẫn chưa nhận được quan tâm đúng mức⁸. Một số rào cản khiến chất lượng dữ liệu xấu có thể bao gồm thiếu sự phân cấp trách nhiệm bảo trì dữ liệu; thiếu chế độ khen thưởng để đảm bảo dữ liệu hợp lệ; thiếu các quy trình kiểm soát dữ liệu; thiếu năng lực của nhân viên; và sự thiếu thân thiện với người dùng của phần mềm được sử dụng để quản lý dữ liệu⁹. Nhìn chung, vấn đề quản lý chất lượng của dữ liệu đã được các doanh nghiệp và đại học tại các quốc gia phát triển như Mỹ, Pháp, Thụy Sĩ, Úc, Hàn Quốc quan tâm, phát triển, ứng dụng và cải thiện qui trình quản lý chất lượng dữ liệu từ những năm 1990. Trong đó, có rất nhiều những nghiên cứu liên quan đến mô hình đánh giá chất lượng dữ liệu, khảo sát về chất lượng dữ liệu, thách thức mà chất lượng dữ liệu của các doanh nghiệp đang phải đối mặt¹⁰⁻¹⁴. Tuy nhiên, ở Việt Nam, đây là một vấn đề còn sơ khởi và khá mới mẻ đối với các doanh nghiệp, các trường đại học cũng như các học giả nghiên cứu. Do đó, mục tiêu của bài báo này là khái quát hóa tổng quan về chất lượng dữ liệu, các vấn đề của chất lượng dữ liệu, các mô hình đã được nghiên cứu, tác động của chất lượng dữ liệu xấu đến hoạt động doanh nghiệp bằng phương pháp nghiên cứu định tính – tổng hợp tài liệu và phân tích thông tin thứ cấp. Từ đó, nhóm tác giả đưa ra những thảo luận và hàm ý quản trị nhằm nâng cao chất lượng dữ liệu trong bối cảnh kinh doanh và thương mại ở Việt Nam.

CƠ SỞ LÝ LUẬN VỀ CHẤT LƯỢNG DỮ LIỆU

Khái niệm về chất lượng dữ liệu (Data quality)

Hiện nay, các nhà khoa học cũng như các chuyên gia trong các lĩnh vực vẫn chưa đưa ra thống nhất chung về định nghĩa chất lượng dữ liệu. Do đó, thuật ngữ này được định nghĩa dựa trên nhiều quan điểm cũng như trên các phương diện khác nhau. Theo cách tiếp cận cổ điển, chất lượng dữ liệu là “phù hợp để sử dụng”, tức là mức độ mà dữ liệu phục vụ thành công các mục

đích của người dùng^{7,15}. Định nghĩa này chủ yếu dựa vào bối cảnh và mang tính tương đối. Tuy nhiên, theo quan điểm nhận thức về đánh giá chất lượng dữ liệu trong bối cảnh⁷, các mô hình đánh giá chất lượng thông tin thường có xu hướng bỏ qua tác động của chất lượng theo bối cảnh đối với việc sử dụng thông tin và kết quả ra quyết định. Theo định nghĩa của Juran và Godfrey (1999)¹⁶, chất lượng dữ liệu để cập đến vấn đề “*dữ liệu có chất lượng cao nếu dữ liệu đáp ứng được nhu cầu sử dụng cho vận hành, ra quyết định và lập kế hoạch*”. Ngoài ra, dữ liệu phải liên quan đến mục đích sử dụng, đủ chi tiết, đủ số lượng, với độ chính xác cao và đầy đủ, thống nhất với các nguồn dữ liệu khác và được trình bày một cách hợp lý. Dữ liệu có giá trị là trung tâm của các hoạt động cấp chiến lược, tác nghiệp và vận hành của mỗi tổ chức. Dữ liệu có chất lượng cao được cung cấp đúng thời điểm sẽ hỗ trợ doanh nghiệp ra quyết định đúng và giúp cho doanh nghiệp thành công trong kinh doanh.

Các nghiên cứu chỉ ra rằng có nhiều cách phân loại chất lượng dữ liệu khác nhau^{17,18}. Cụ thể, Ballou và Pazer (1985) đã phân chia chất lượng dữ liệu thành bốn khía cạnh bao gồm: độ chính xác, tính kịp thời, tính đầy đủ và tính nhất quán¹⁷. Theo nhóm tác giả này, độ chính xác là dễ đánh giá nhất, vì đây chỉ là sự khác biệt giữa giá trị chính xác và giá trị thực tế được ghi lại. Tương tự, việc đánh giá tính kịp thời cũng có thể được thực hiện theo cách trên trong khi đó, việc đánh giá tính đầy đủ cũng tương đối đơn giản, miễn là dữ liệu có đầy đủ hay không hơn là mức độ hoàn chỉnh của dữ liệu. Mặt khác, việc đánh giá tính nhất quán khá phức tạp do đòi hỏi phải có hai hoặc nhiều sơ đồ biểu diễn để so sánh¹⁷. Trong khi đó, Wand và Wang đã giới hạn chất lượng dữ liệu nội tại và xác định bốn yếu tố nội tại, bao gồm tính đầy đủ, tính rõ ràng, tính có ý nghĩa và tính đúng đắn¹⁸. Theo nghiên cứu của Redman về tác động của chất lượng xấu đến doanh nghiệp, việc phân loại dữ liệu của một doanh nghiệp có thể theo các vấn đề như chế độ xem dữ liệu (mức độ liên quan, mức độ chi tiết); giá trị dữ liệu (độ chính xác, tính nhất quán, đơn vị và tính đầy đủ); trình bày dữ liệu (tính thích hợp của định dạng, tính dễ giải thích); và các vấn đề như quyền riêng tư, bảo mật và quyền sở hữu¹⁹. Trong khi đó, Wang và Strong chia chất lượng dữ liệu thành bốn loại bao gồm tính nội tại với các khía cạnh như độ tin cậy, độ chính xác, tính khách quan và danh tiếng; theo ngữ cảnh; tính đại diện; và khả năng truy cập (Bảng 1)¹⁸. Ngoài ra, Haug và cộng sự đã xác định ba loại chất lượng dữ liệu bao gồm tính nội tại, khả năng truy cập và tính hữu ích⁹. Về cơ bản, Haug và cộng sự⁹ ủng hộ định nghĩa do Wang và Strong²⁰ và Wang²¹ đề xuất nhưng họ cho rằng “chất lượng dữ liệu mang tính đại diện” có

Bảng 1: Các khía cạnh của chất lượng dữ liệu²⁰

Thể loại	Khía cạnh	Mô tả
Tính nội tại	Chính xác (Accuracy)	Dữ liệu là chính xác
	Đáng tin cậy (Believability)	Mức độ dữ liệu phản ánh đáng tin
	Khách quan (Objectivity)	Dữ liệu khách quan
	Uy tín (Reputation)	Nội dung dữ liệu và nguồn được xem xét kỹ lưỡng
Theo ngữ cảnh	Số lượng phù hợp (Appropriate Amount)	Số lượng dữ liệu phù hợp
	Hoàn chỉnh (Completeness)	Mức độ đầy đủ của dữ liệu
	Liên quan (Relevancy)	Dữ liệu khả dụng và hữu ích
	Giá trị gia tăng (Value-added)	Dữ liệu cung cấp lợi thế cạnh tranh
	Kịp thời (Timeless)	Thời gian tồn tại của dữ liệu
Tính đại diện	Thể hiện súc tích (Concise representation)	Dữ liệu thể hiện chặt chẽ, ngắn gọn
	Dễ hiểu (Ease of understanding)	Dữ liệu rõ ràng, dễ đọc và dễ hiểu
	Diễn giải rõ ràng (Interpretability)	Ý nghĩa dữ liệu được giải thích
	Nhất quán (Consistency)	Dữ liệu thể hiện tính liên tục theo cấu trúc giống nhau
Khả năng truy cập (Accessibility)	Bảo mật truy cập (Access security)	Việc truy cập được đảm bảo hoặc bị hạn chế
	Truy xuất (Accessibility)	Mức độ dữ liệu có thể truy xuất

Nguồn: Wang và Strong (1996)

thể được coi là một dạng của “chất lượng dữ liệu có khả năng truy cập”.

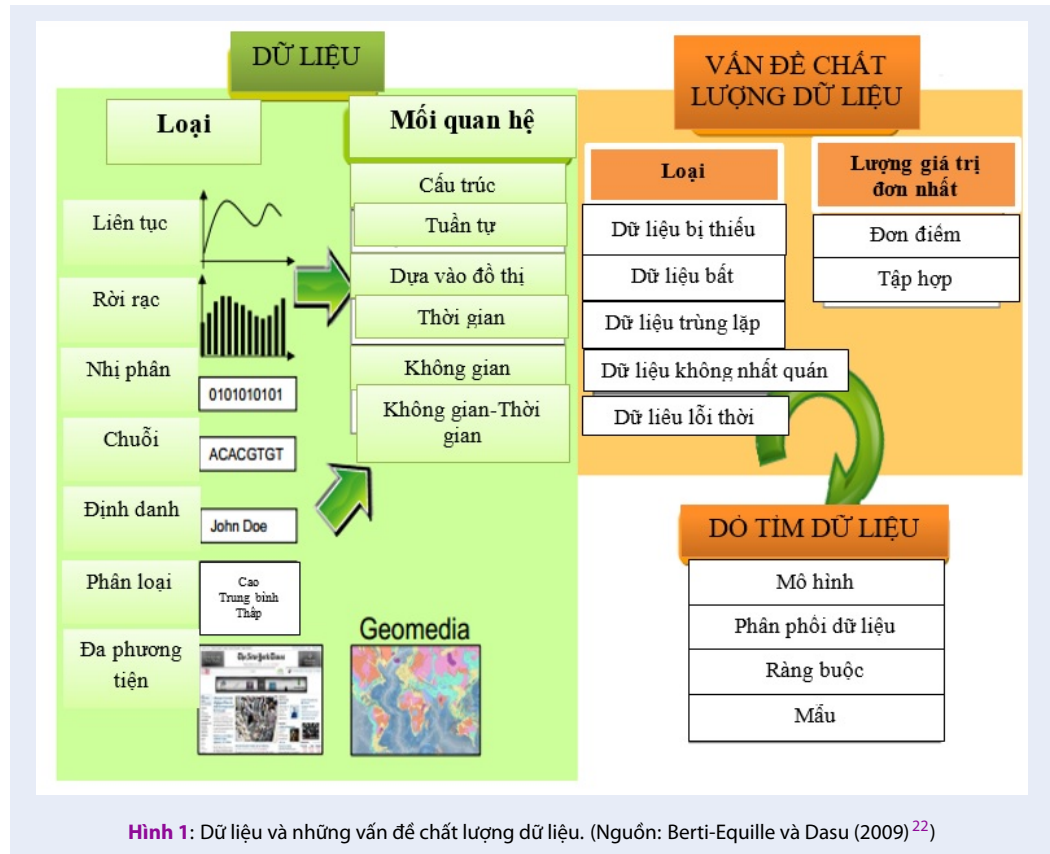
Nhìn chung, dữ liệu có chất lượng cao khi mà dữ liệu phù hợp cho mục đích sử dụng khi doanh nghiệp cần. Điều này có nghĩa rằng dữ liệu phụ thuộc vào hoàn cảnh mà dữ liệu được sử dụng, nên không thể so sánh chất lượng tuyệt đối. Do đó, để định nghĩa chất lượng dữ liệu cần tuân những tiêu chí cơ bản sau: i) Tính hoàn chỉnh (dữ liệu không thiếu và phải đầy đủ); ii) Tính giá trị (dữ liệu phù hợp với qui định); iii) Tính đơn nhất (dữ liệu không bị trùng lặp); iv) Tính thống nhất (dữ liệu thống nhất xuyên suốt các kho dữ liệu); v) Tính đúng thời gian (dữ liệu phải có thực khi cần); vi) Tính chính xác (dữ liệu phải thể hiện được thực tế); vii) Tính kịp thời (dữ liệu phải cập nhật và kịp thời); viii) Tính bảo mật (dữ liệu phải được kiểm soát và giám sát); ix) Tính nhất quán (dữ liệu giống nhau cho tất cả các trường hợp).

Các vấn đề về chất lượng dữ liệu và dữ liệu xấu

Dữ liệu và thông tin chính là “nguồn sống” của mọi cá nhân, tổ chức, xã hội nhất là trong nền kinh tế số hiện nay. Đặc biệt đối với doanh nghiệp, trong thời đại

thông tin, dữ liệu được xem là tài sản vô giá. Nếu như doanh nghiệp không hiểu, nắm bắt rõ và đầy đủ tầm quan trọng của dữ liệu và thông tin trong nền kinh tế số thì không thể tồn tại và phát triển. Do được xem là tài sản của doanh nghiệp nên dữ liệu không chỉ giúp doanh nghiệp đạt được mục tiêu chiến lược mà còn giúp quản lý một cách toàn diện. Không có doanh nghiệp nào quản lý có hiệu quả và hiệu lực nếu như không có dữ liệu có chất lượng cao. Theo Hiệp hội quản lý dữ liệu (DAMA – DATA Management Association), dữ liệu có chất lượng cao là dữ liệu đáp ứng nhu cầu và mong đợi của người sử dụng, hay nói cách khác là phù hợp với mục đích sử dụng của người dùng. Ngược lại nếu dữ liệu không phù hợp với mục đích sử dụng được xem là dữ liệu có chất lượng kém hay dữ liệu xấu²². Doanh nghiệp dựa vào dữ liệu để ra quyết định, các nhà lãnh đạo dựa vào dữ liệu để tạo lợi thế cạnh tranh thông qua việc am hiểu khách hàng, sử dụng thông tin cải tiến, và vận hành hiệu quả. Do đó việc cải thiện dữ liệu xấu là hết sức quan trọng và cấp thiết²².

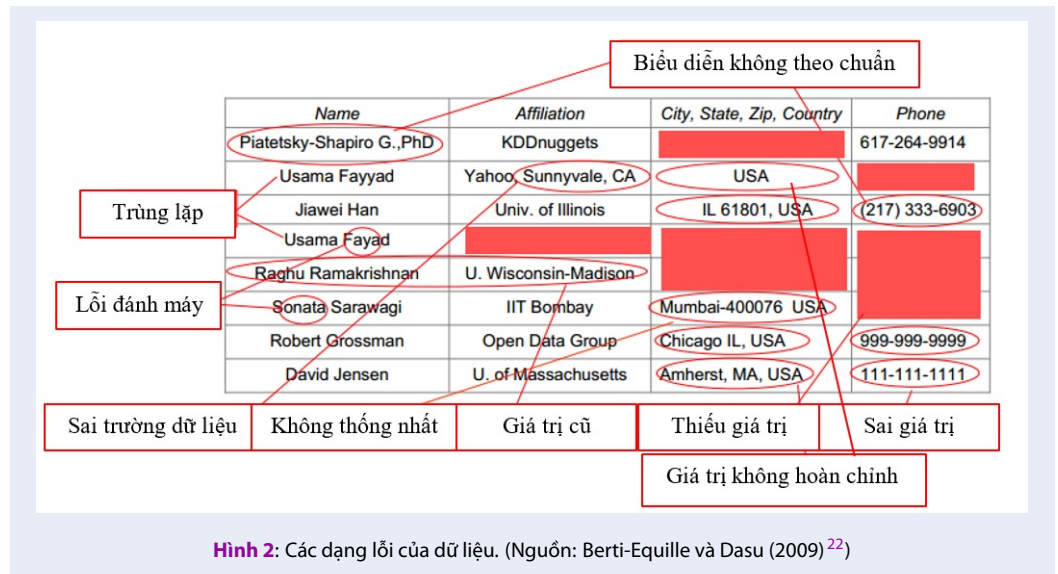
Theo Berti-Equille và Dasu²³, các vấn đề liên quan đến chất lượng dữ liệu xấu bao gồm thiếu dữ liệu (missing data), dữ liệu bị sai (erroneous data), dữ



liệu bất thường (data anomalies), trùng lặp (duplicates), không thống nhất (inconsistent data), dữ liệu lỗi thời (out-of-date data), chưa được chuẩn hóa (undocumented data). Các dạng dữ liệu được biểu diễn dưới dạng liên tục (continuous), rời rạc (discrete), nhị phân (binary), chuỗi (sequence), định danh (nominal), định khoảng (categorical), đa phương tiện (multimedia) hay địa lý (geomedia). Các dữ liệu đặt trong mối quan hệ cấu trúc (structural), chuỗi (sequential), đồ thị (graph-based), theo thời gian (temporal), theo không gian (spatial), theo không gian-thời gian (spatio-tempo). Việc dò tìm dữ liệu xấu dựa trên mô hình (model-based), phân phối dữ liệu (data distribution-based), ràng buộc (constraint-based), mẫu (pattern-based). Hình 1 minh họa các dạng dữ liệu và vấn đề của chất lượng dữ liệu. Bên cạnh đó, theo Berti-Equille và Dasu²³, các dạng lỗi thường gặp của dữ liệu là biểu diễn không theo chuẩn (non-standard representation) lỗi đánh máy (typos), trùng lặp (duplicates), sai trường dữ liệu (misfiled value), không thống nhất (inconsistency), giá trị cũ (obsolete value), giá trị bị thiếu (missing value), giá trị bị sai (incorrect value), giá trị không hoàn chỉnh (incomplete value). Hình 2 minh họa các dạng lỗi của dữ liệu.

Các mô hình và công cụ quản lý chất lượng dữ liệu.

Ngày nay, khảo cứu lý thuyết chỉ ra rằng có nhiều mô hình và bộ khung khác nhau về quản lý chất lượng dữ liệu²⁴. Mô hình sơ khởi quản lý chất lượng dữ liệu toàn diện (TDQM – Total Data Quality Management) được phát triển bởi Wang dựa trên triết lý và mô hình quản lý chất lượng toàn diện (TQM) – Total Quality Management, xem dữ liệu như là sản phẩm vật lý trong dòng sản xuất từ đầu vào đến đầu ra²¹. Chất lượng của dữ liệu được cải thiện thông qua vòng tròn Deming: Lập kế hoạch (Plan), thực hiện (Do), kiểm tra (Check), và hành động (Act)²¹. English Lary phát triển mô hình quản lý chất lượng thông tin toàn diện (TIQM – Total Information Quality Management) thiết lập môi trường chất lượng, đánh giá định nghĩa dữ liệu và chất lượng kiến trúc, đo lường chi phí chất lượng do dữ liệu xấu, tái cấu trúc và làm sạch dữ liệu, và cải thiện quy trình²⁵. Ryu và cộng sự²⁶ chỉ ra rằng chất lượng giá trị dữ liệu và chất lượng dịch vụ dữ liệu thấp là do cấu trúc dữ liệu thấp; vì vậy nhóm tác giả này đề xuất quản trị siêu dữ liệu (metadata) và mô hình trưởng thành quản lý chất lượng dữ liệu (Data Quality Management Maturity Model) với 4 cấp độ. Cấp độ 1 là bước quản lý



dữ liệu ban đầu và quản lý chất lượng cấu trúc dữ liệu thông qua các quy tắc. Cấp độ 2 mô hình hóa dữ liệu mức logic và vật lý. Cấp độ 3 tiêu chuẩn hóa dữ liệu. Cấp độ 4 là tối ưu hóa dữ liệu thông qua quản lý kiến trúc dữ liệu.

Viện thông tin sức khỏe của Canada (CIHI – Canadian Institute for Health Information) đề xuất mô hình chất lượng dữ liệu CIHI, tiếp cận quản lý đối tượng dữ liệu thông qua quy trình dòng dữ liệu (data flow process) và sử dụng phương pháp cải tiến liên tục (continuous improvement). Ba thành phần chính của mô hình bao gồm vòng tròn công việc chất lượng dữ liệu, công cụ đánh giá chất lượng dữ liệu, và tài liệu hóa chất lượng dữ liệu²⁷. Cùng tư duy về quản lý theo tiến trình thay thì quản lý chất lượng dựa trên kết quả, Kim và Lee đề xuất mô hình quản lý chất lượng dữ liệu tổng thể dựa theo triển trình (process-centric)²⁸. Ba tiến trình chức năng là biên tập dữ liệu (data manipulation), giám sát chất lượng (quality monitoring), and cải thiện chất lượng (quality improvement), kết hợp với ba vai trò của nhà quản lý, người kiểm soát, và người sử dụng; tạo nên ma trận 3x3 của tiến trình chất lượng dữ liệu. Dựa trên mô hình này, Trần và cộng sự đề xuất mô hình kết hợp giữa đội ngũ quản lý dữ liệu (data stewards) và quản lý dòng dữ liệu (data flow management)²⁹. Tóm lại, ngày nay có rất nhiều mô hình và bộ khung trong quản lý chất lượng dữ liệu và chưa có sự thống nhất chung ngoại trừ bộ tiêu chuẩn chất lượng dữ liệu ISO 8000 mà tổ chức quốc tế vì tiêu chuẩn hóa đang biên soạn và hoàn thiện.

Đi kèm với các triết lý, mô hình, bộ khung về chất lượng dữ liệu, để mô hình hóa các dữ liệu, các học giả và chuyên gia sử dụng đa dạng các công cụ và phần

mềm để quản lý. Trong đó chúng ta cần phải kể đến công cụ IDEF0 (Definition for Function Modelling) của Hoa Kỳ; IPMAP (Information Quality MAP) của Shankaranarayanan và Wang³⁰, kết hợp giữa IPMAP và UML (IP-UML)³¹. Các phần mềm thông dụng ngày nay có chức năng quản lý chất lượng dữ liệu phải kể đến Informatica Data Quality, Oracle Enterprise Data Quality, SAP Data Services, SAP Information Steward, IBM Infosphere Information, Microsoft Data Quality Service.

PHƯƠNG PHÁP NGHIÊN CỨU

Trong bài báo này, chúng tôi sử dụng phương pháp nghiên cứu bàn giấy (desk research) hay còn gọi là nghiên cứu thứ cấp (secondary research) nhằm thu thập, tổng hợp, phân tích, và đánh giá những nguồn dữ liệu và thông tin các nghiên cứu trước đây trong lĩnh vực chất lượng dữ liệu, quản lý chất lượng dữ liệu, các mô hình chất lượng dữ liệu đã và đang được triển khai tại một số nước phát triển như Mỹ, Pháp, Úc, Hàn Quốc, và các nước khác. Bên cạnh phương pháp nghiên cứu bàn giấy, nhóm tác giả cũng có thời gian nghiên cứu thực địa (field research) tại Văn phòng Tin dụng Hàn Quốc (nơi mà chất lượng dữ liệu được quan tâm hàng đầu) trong thời gian 6 tháng tại Hàn Quốc (đến tháng 9 năm 2019). Tại đây, nhóm tác giả có thực hiện phỏng vấn 03 đối tượng là Giám đốc văn phòng, Trưởng phòng Quản lý chất lượng, và nhân viên của phòng Quản lý chất lượng. Kết hợp giữa lý thuyết và thực tiễn ở nước phát triển, chúng tôi hệ thống hóa lý thuyết và đề xuất kiến nghị những vấn đề liên quan quản lý chất lượng dữ liệu cho bối cảnh Việt Nam.

KẾT QUẢ NGHIÊN CỨU

Có thể nhận thấy rằng bất kỳ tổ chức nào cũng bị tác động bởi chất lượng của dữ liệu. Trong đó, dữ liệu xấu gây ra những hậu quả xấu cho hoạt động kinh doanh của các doanh nghiệp bao gồm quyết định sai lầm, thất bại trong tiến trình kinh doanh và dự án; khiến chi phí tăng hoặc doanh thu giảm, giảm chất lượng cung cấp sản phẩm, dịch vụ và gây ra sự không hài lòng của các bên liên quan khác³².

Ngày nay, các doanh nghiệp đang nỗ lực đầu tư vào kỹ thuật và công nghệ thu thập, lưu trữ, và xử lý khối lượng lớn dữ liệu. Tuy nhiên, các doanh nghiệp nhận thấy có những trở ngại trong nỗ lực biến chuyển dữ liệu thành thông tin có ý nghĩa, mà các doanh nghiệp sử dụng trong việc cải thiện quy trình kinh doanh, ra quyết định tốt hơn, và tạo lợi thế cạnh tranh. Vấn đề nằm ở chỗ chất lượng của dữ liệu và thông tin có thể tạo ra những trở ngại này đứng về phía kỹ thuật như tích hợp dữ liệu từ nhiều nguồn khác nhau, hay phi kỹ thuật như thiếu chiến lược đồng bộ toàn doanh nghiệp để đảm bảo các bên liên quan có quyền thông tin và có thông tin đúng thời điểm, đúng nơi, và đúng thời gian³³.

Phương pháp đánh giá những tác động

Phạm vi nghiên cứu tác động của chất lượng dữ liệu đến doanh nghiệp là phát triển phương pháp đánh giá những tác động, kiểm soát tác động là tích cực hay tiêu cực. Theo đó, có năm lĩnh vực cần phân nhỏ để nghiên cứu tác động, bao gồm: các hệ thống ứng dụng Công nghệ thông tin (CNTT) trong doanh nghiệp và bị tác động bởi chất lượng dữ liệu như Quản trị quan hệ khách hàng (CRM – Customer Relationship Management), Quản trị tri thức (KM – Knowledge Management), Quản trị chuỗi cung ứng (SCM – Supply Chain Management), Lập kế hoạch nguồn lực doanh nghiệp (ERP – Enterprise Resource Planning); kết quả hoạt động kinh doanh của tổ chức và cá nhân, phân tích lợi ích chi phí, và ảnh hưởng đến hoạt động vận hành; lĩnh vực quản lý CNTT, đầu tư dự án CNTT, quản trị CNTT; thay đổi tiến trình kinh doanh và tổ chức; và cuối cùng là lĩnh vực chiến lược và chính sách.

Chất lượng dữ liệu xấu

Theo nghiên cứu của Gartner³⁴, chất lượng dữ liệu xấu là nguyên nhân chính của 40% tất các hoạt động và dự án kinh doanh thất bại trong việc đạt được mục tiêu lợi nhuận. Chất lượng dữ liệu xấu làm giảm năng suất lao động đến 20%. Khi mà các tiến trình kinh doanh được tự động hóa, chất lượng dữ liệu trở thành nhân tố quyết định đến chất lượng toàn bộ quy trình

kinh doanh. Cụ thể, chất lượng dữ liệu tác động lên ba yếu tố của doanh nghiệp. Yếu tố thứ nhất là quản lý các tiến trình kinh doanh (BPM - Business Processes Management). Ngày nay, các tiến trình kinh doanh được số hóa và công nghệ thông tin giúp tự động hóa trong các giao dịch, như tự động bán hàng, thiết kế, gia công sản phẩm mới, tuyển dụng, vấn đề dữ liệu trở nên quan trọng. Chi phí ước tính cho việc quản lý dữ liệu là 20% so với tổng chi phí các tiến trình kinh doanh. Nếu doanh nghiệp chi tiêu cho 100 triệu đô-la cho toàn bộ hoạt động bán hàng tự động thì vấn đề quản lý chất lượng dữ liệu có chi phí 20 triệu đô-la. Chi phí này không được tính một lần mà là con số tích lũy theo hàng năm. Do đó, chi phí về vấn đề chất lượng dữ liệu trong giao dịch là rất lớn. Yếu tố thứ hai là năng suất (Productivity). Tại Mỹ, nhờ ứng dụng công nghệ thông tin vào trong công việc nên năng suất lao động tăng hàng năm khoảng 2%, với năng suất trung bình của lao động nước Mỹ là 70%. Sự tăng trưởng năng suất này tiết kiệm được trung bình 20 triệu đô-la cho doanh nghiệp, nhưng vấn đề chất lượng dữ liệu xấu lại gây thiệt hại 10% trong khoảng tiết kiệm này. Yếu tố thứ ba là ra quyết định (Decision making). Theo tính toán, có 60% mục tiêu lợi nhuận đạt được, điều này đồng nghĩa với 40% là không đạt được mục tiêu trong quá trình ra quyết định. Ước tính trung bình các doanh nghiệp Mỹ tổn thất khoảng 2 triệu đô cho việc chất lượng dữ liệu xấu³⁴.

Quản lý dữ liệu

Theo Redman (1998)¹⁹, những vấn đề thường gặp trong quản lý dữ liệu là dữ liệu không chính xác (inaccurate data) – nếu dữ liệu bị sai từ 1% đến 5% trong trường dữ liệu thì được xem là lỗi; sự không thống nhất trong cơ sở dữ liệu; dữ liệu cần cho hoạt động vận hành và ra quyết định nhưng không có. Từ đó, có những tác động đến doanh nghiệp ở các cấp độ khác nhau. Cụ thể, thứ nhất ở cấp hoạt động (Operational level), chất lượng dữ liệu xấu tác động trực tiếp đến sự không hài lòng của khách hàng, gia tăng chi phí (theo đánh giá là từ 8% đến 12%), và làm giảm sự hài lòng của nhân viên. Ví dụ đối với khách hàng, nếu dữ liệu bị sai tên hay sai địa chỉ được ghi trên sản phẩm hoặc dịch vụ phân phối đến khách hàng sẽ dẫn đến mất thời gian và chi phí cho việc giao hàng lại. Lỗi nhỏ nhưng gây ra thiệt hại lớn cho cả khách hàng và doanh nghiệp. Thứ hai, ở cấp chiến lược (Strategic level), chất lượng dữ liệu xấu tác động đến chiến lược không rõ ràng hơn so với cấp tác nghiệp vì chiến lược mang tính chất lâu dài và yêu cầu dữ liệu không những bên trong mà còn bên ngoài. Do đó, tác động đến cấp chiến lược là khó đánh giá và mang tính tiềm

ấn. Tuy nhiên, chất lượng dữ liệu xấu dẫn đến khó khăn cho ban giám đốc hoạch định chiến lược, thực thi chiến lược, xác định vấn đề liên quan đến sở hữu dữ liệu, đồng bộ hóa các hoạt động của tổ chức, và làm sai lệch trong quản trị. Thứ ba, sẽ gây ra một số tác động điển hình khác (typical impacts) như chất lượng dữ liệu xấu sẽ ảnh hưởng đến tiêu cực trong việc ra quyết định (garbage in, garbage out), làm mất thời gian trong việc quản lý kho dữ liệu (data warehouses), làm khó cho việc tái kiến trúc doanh nghiệp, và giảm lòng tin của tổ chức.

Phân loại các loại chi phí

Có nhiều lập luận khác nhau về việc phân loại các loại chi phí mà doanh nghiệp phải chịu khi gặp vấn đề về chất lượng dữ liệu xấu^{34,35}. Theo nghiên cứu của Epler và Helfert³⁵, có hai loại chi phí về chất lượng dữ liệu chính doanh nghiệp sẽ bị tổn thất trong đó chia ra làm năm loại cụ thể bao gồm chi phí trực tiếp, chi phí gián tiếp, chi phí phòng ngừa, chi phí tìm lỗi, chi phí sửa chữa (Bảng 2). Gartner³⁴ cho rằng vấn đề chất lượng dữ liệu ảnh hưởng không chỉ đến bộ phận công nghệ thông tin mà còn ảnh hưởng liên đới đến những bộ phận khác trong doanh nghiệp (Bảng 3). Đối với bộ phận marketing và bán hàng, các dữ liệu về sản phẩm và khách hàng bị sai hoặc không chính xác sẽ gây mất lòng tin, giảm sự hài lòng hay mất thời gian cho khách hàng. Khách hàng sẽ bỏ đi tìm nhà cung cấp khác để nhận cung cấp dịch vụ. Đối với bộ phận tài chính và kế toán, dữ liệu sai sót về doanh thu, lợi nhuận, chi phí sẽ ảnh hưởng đến việc lập sai ngân sách, các mục tiêu về tài chính; khó kiểm soát mục tiêu chiến lược và mức độ hoàn thành công việc. Các chỉ số tài chính không phản ánh đúng thực trạng doanh nghiệp. Đối với bộ phận mua hàng trong chuỗi cung ứng, việc sai sót dữ liệu sẽ sai số lượng và thời gian giao nhận hàng, có những lúc thiếu hàng và thừa hàng. Hoặc sản phẩm được tồn kho lâu ngày do dữ liệu thông báo bị sai dẫn đến sản phẩm có khả năng bị hư hỏng, không hợp thời, giảm giá trị. Đối với bộ phận CNTT, dữ liệu sai sót ảnh hưởng đến quyết định đầu tư hay không đầu tư dự án CNTT cho các giai đoạn tiếp theo của doanh nghiệp. Theo nghiên cứu của Spruit và Linden³⁶, dữ liệu không có chất lượng cao có thể gây ra 13 tác động đến hoạt động doanh nghiệp, bao gồm (1) mất cơ hội bán hàng, (2) chi phí dịch vụ khách hàng tăng, (3) giảm hài lòng của khách hàng, (4) giảm doanh thu, (5) giảm hiệu quả vận hành, (6) trì trệ hệ thống, (7), không tuân thủ quy định, quy chế doanh nghiệp, (8) ra quyết định kém, (9) mất cơ hội kinh doanh, (10) giảm lòng tin/ đạo đức của nhân viên, (11) giảm độ tin cậy của hệ thống.

Tóm lại, có rất nhiều nghiên cứu chỉ ra những tác động của dữ liệu chất lượng xấu đến hoạt động của doanh nghiệp, bao gồm ảnh hưởng xấu đến quá trình ra quyết định, tiến trình kinh doanh, các hoạt động cấp chiến lược, chiến thuật và tác nghiệp của doanh nghiệp (Hình 3). Trong khi đây là những hoạt động có tác động đến giá trị lâu dài của doanh nghiệp.

THẢO LUẬN VÀ KẾT LUẬN

Chủ đề về quản lý chất lượng dữ liệu đã được nghiên cứu ở Mỹ từ những năm 1990s, và các nước phát triển khác như Pháp, Úc, Thụy Sĩ, Hàn Quốc từ những năm 2000s. Ứng dụng quản lý dữ liệu đã giúp cho các ngành công nghiệp ở các nước này tiết kiệm chi phí hàng tỷ đô-la. Về mặt thực tiễn, khái niệm về quản lý chất lượng dữ liệu còn khá mới mẻ và các doanh nghiệp ở Việt Nam dường như chưa biết đến nhiều. Các hệ thống thông tin quản trị còn chưa đồng bộ và toàn diện (ví dụ như ứng dụng và triển khai hệ thống ERP, CRM, SCM trong doanh nghiệp) và gắn kèm với chức năng quản lý chất lượng dữ liệu trong tiến trình kinh doanh còn nhiều hạn chế. Các doanh nghiệp lớn ở Việt Nam đã ứng dụng hệ thống lập nguồn lực doanh nghiệp của Oracle, SAP; nhưng các doanh nghiệp vừa và nhỏ ứng dụng hệ thống thông tin quản lý còn ít và chưa đồng bộ. Đi kèm theo đó là vấn đề quản lý chất lượng dữ liệu chưa được quan tâm đúng mức. Các doanh nghiệp hiện tại dường như chưa có các công cụ công nghệ thông tin để áp dụng trong quản lý chất lượng với đối tượng là dữ liệu.

Về mặt học thuật, quản trị chất lượng dữ liệu ngày nay vẫn còn thiếu nền tảng và sự thống nhất. Phần lớn các chuyên gia và học giả trên thế giới dựa trên nền tảng của TMQ dành cho sản phẩm và dịch vụ, cải tiến cho quản lý chất lượng dữ liệu theo định hướng tiến trình và tư duy cải tiến liên tục. Hiện tại vẫn chưa có sự thống nhất về quan điểm và mô hình chất lượng cho toàn thế giới, ngoại trừ bộ tiêu chuẩn ISO-8000 (bộ tiêu chuẩn toàn cầu về Chất lượng dữ liệu và dữ liệu tổng thể doanh nghiệp, tương tự bộ tiêu chuẩn ISO-9000 liên quan đến sản phẩm và dịch vụ) nhưng vẫn chưa hoàn thiện.

Bài báo này xem xét chất lượng dữ liệu dưới góc cạnh quản lý trong bối cảnh ở Việt Nam về mặt ứng dụng thực tiễn và kiểm soát quản trị. Bài báo này hướng đến hàm ý quản trị cho bốn đối tượng như sau:

Thứ nhất là doanh nghiệp có sử dụng dữ liệu. Về con người, doanh nghiệp cần tổ chức xây dựng bộ phận, đội nhóm về quản lý chất lượng dữ liệu (data quality teams), tương tự như bộ phận kiểm tra chất lượng sản phẩm (KCS) trong nhà máy sản xuất. Đội nhóm này bao gồm nhà quản lý, nhân viên chuyên trách về sử dụng và quản lý dữ liệu (data stewardship), chuyên

Bảng 2: Phân loại chi phí chất lượng dữ liệu. ³⁴

Chi phí do chất lượng dữ liệu xấu	Chi phí trực tiếp	Chi phí xác nhận
		Chi phí nhập lại dữ liệu
		Chi phí bù lỗ
	Chi phí gián tiếp	Chi phí vì giảm danh tiếng
		Chi phí vì ra quyết định sai hoặc hành động sai
		Các chi phí lặn
Chi phí cải thiện và ngăn ngừa chất lượng dữ liệu	Chi phí phòng ngừa	Chi phí đào tạo
		Chi phí giám sát
		Chi phí phát triển và khai thác lại dữ liệu chuẩn
	Chi phí tìm lỗi	Chi phí phân tích
		Chi phí báo cáo
	Chi phí sửa chữa	Chi phí lập kế hoạch sửa chữa
		Chi phí thực hiện sửa chữa

Nguồn: Eppler và Helfert (2004)

Bảng 3: Chi phí của dữ liệu có chất lượng thấp. ³⁴

	Vấn đề	Tác động
Marketing và bán hàng	Giảm sự hài lòng của khách hàng Nhiều địa chỉ giao hàng bị thay đổi. Giảm lòng tin	Khách hàng sẽ bỏ đi Tổn chi phí giao hàng lại Đánh mất chi phí cơ hội do mất doanh thu bán hàng.
Tài chính	Lập ngân sách không chính xác Sai số giữa thực tế và ngân sách Giảm lòng tin trong báo cáo	Lập ngân sách không có hiệu lực Không kiểm soát được chi phí Bị phạt do báo cáo bị sai.
Chuỗi cung ứng	Cạn dự trữ hàng tồn kho Sản phẩm có chất lượng kém Giảm lòng tin	Giảm doanh thu bán hàng Tăng chi phí vận chuyển
Công nghệ thông tin	Nhiều dự án công nghệ thông tin thất bại. Ít sử dụng các ứng dụng CNTT Giảm lòng tin	Lãng phí trong đầu tư dự án CNTT Năng suất lao động thấp

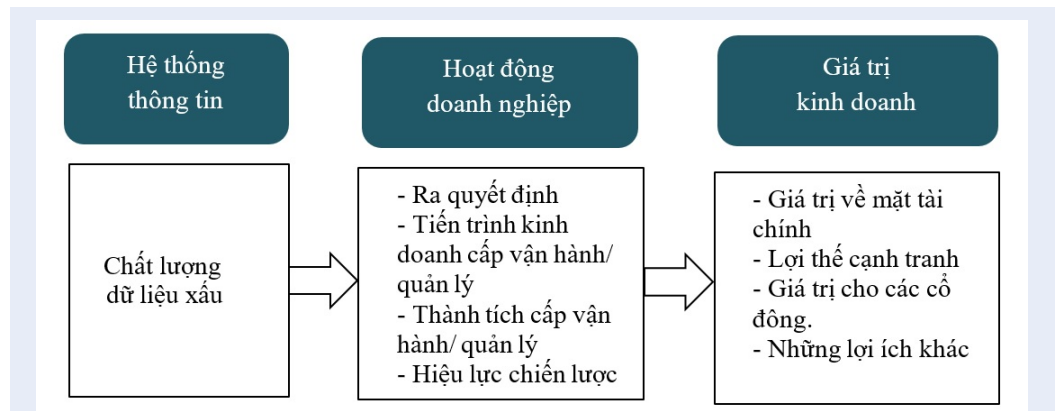
Nguồn: Gartner (2011)

gia phân tích chất lượng dữ liệu (data quality analyst), chuyên gia tích hợp (integration specialist). Về mặt kỹ thuật, doanh nghiệp cũng nên đầu tư phần mềm để kiểm soát chất lượng dữ liệu; về mặt kinh doanh, doanh nghiệp có thể xây dựng lại tiến trình kinh doanh (business process) và dòng chảy công việc (workflow) để truy vết dữ liệu, tránh sai sót trong quá trình dịch chuyển dữ liệu. Tran và cộng sự (2010) đề xuất mô hình kết hợp giữa đội ngũ quản lý dữ liệu (data stewards) và quản trị dòng dữ liệu (data workflow) ²⁸, tuy nhiên mô hình này cần được mở rộng

trong sự phối hợp với các chức năng khác trong bối cảnh toàn doanh nghiệp.

Thứ hai là doanh nghiệp đầu tư, phát triển phần mềm quản lý dữ liệu và doanh nghiệp cung cấp dịch vụ hỗ trợ quản lý dữ liệu. Các doanh nghiệp này có tiềm năng xây dựng phần mềm phục vụ doanh nghiệp trong nước và thương mại hóa phục vụ cho hoạt động quản trị doanh nghiệp chú trọng đến cải thiện chất lượng dữ liệu.

Ngày nay, với sự tiến bộ của công nghệ mới, công nghệ blockchain ngày càng phổ biến và các hợp đồng



Hình 3: Tác động của chất lượng dữ liệu đến hoạt động kinh doanh của doanh nghiệp. Nguồn: Nhóm tác giả, năm 2022

thông minh blockchain sẽ tăng chất lượng dữ liệu lên đến 50% vào năm 2023, theo báo cáo của Gartner³⁷. Tran và Trinh đã phân tích việc ứng dụng công nghệ blockchain vào chuỗi cung ứng thực phẩm nông nghiệp ở Việt Nam³⁸, và một số nền tảng như Agri-ndential giúp nâng cao việc quản lý dữ liệu. Chính vì vậy, phát triển các hệ thống thông tin quản lý và ứng dụng công nghệ blockchain vào trong quản lý dữ liệu tạo nên một nền tảng CNTT là một định hướng tiềm năng và cần được xem xét và phát triển trong bối cảnh cần phải nâng cao chất lượng dữ liệu trong doanh nghiệp.

Thứ ba là nhà khoa học và các nhà hoạch định chính sách. Hiện nay lĩnh vực quản lý chất lượng dữ liệu ở Việt Nam còn ở mức sơ khởi, trong khi đó các nước phát triển đã trong giai đoạn tăng trưởng trong việc xây dựng hệ thống quản lý chất lượng dữ liệu. Các nhà khoa học cần đẩy mạnh nghiên cứu vấn đề chất lượng dữ liệu, xây dựng một nền tảng và bộ khung (framework) cho Việt Nam về vấn đề chất lượng dữ liệu, các tiến trình kinh doanh tương ứng với dữ liệu được sử dụng. Các nhà hoạch định chính sách đã công bố TCVN 10249 năm 2013 dựa trên nền tảng bộ ISO 8000 năm 2011. Tuy nhiên, phiên bản ISO 8000 năm 2022 có nhiều điểm cập nhật bổ sung, chính vì vậy các nhà hoạch định chính sách cũng nên chỉnh sửa, cập nhật, và tiến đến hoàn thiện. Trên cơ sở ISO 8000, các nhà hoạch định chính sách cũng nên bắt đầu dự án xây dựng bộ khung về chất lượng dữ liệu TCVN phù hợp với đặc thù bối cảnh của Việt Nam. Đồng thời các nhà hoạch định chính sách và thực thi nên phổ biến cho các đối tượng liên quan biết, hiểu, nắm rõ và vận dụng TCVN 10249 hoặc ISO 8000 vào trong các tổ chức.

Thứ tư là nhà trường. Trong mối quan hệ nhà trường – doanh nghiệp – sinh viên, khi doanh nghiệp có

những yêu cầu nhân lực về chất lượng dữ liệu, nhà trường nên có những học phần hoặc modules các học phần về quản lý dữ liệu tổng thể (master data management) và chất lượng dữ liệu để trang bị kiến thức, kỹ năng cho sinh viên sau khi ra trường đáp ứng những việc làm mà doanh nghiệp cần trong thời đại số.

Bài báo này trình bày sơ khởi và đặt nền móng cho vấn đề quản lý chất lượng dữ liệu và đề xuất một số gợi ý trong bối cảnh của Việt Nam, nên cần những nghiên cứu lý thuyết, như đề xuất mô hình về quản lý chất lượng dữ liệu theo bộ tiêu chuẩn của Việt Nam và các nghiên cứu định lượng, như khám phá về mức độ ứng dụng và chấp nhận sử dụng các công nghệ và các công cụ thông tin trong các doanh nghiệp tại Việt Nam. Chủ đề về quản lý chất lượng dữ liệu tại Việt Nam còn mới nên hy vọng các học giả, nhà nghiên cứu, chuyên gia về công nghệ thông tin và quản lý tại doanh nghiệp sẽ cụ thể hóa chủ đề này nữa trong tương lai.

DANH MỤC CÁC TỪ VIẾT TẮT

BPM: Quản lý tiến trình kinh doanh (Business Process Management)

CNTT: Công nghệ thông tin

CRM: Quản trị quan hệ khách hàng (Customer Relationship Management)

DAMA: Hiệp hội quản lý dữ liệu (DATA Management Association)

ERP: Lập kế hoạch nguồn lực doanh nghiệp (Enterprise Resource Planning)

KM: Quản trị tri thức (Knowledge Management)

SCM: Quản trị chuỗi cung ứng (Supply Chain Management)

ROI: Tỷ số hoàn vốn (Return on Investment)

TCVN: Tiêu Chuẩn Việt Nam

UML: Ngôn ngữ mô hình thống nhất (Unified Modeling Language)

XUNG ĐỘT LỢI ÍCH

Các tác giả không có bất kỳ xung đột lợi ích nào trong bài báo này.

ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Hai tác giả đồng khởi tạo ý tưởng, nghiên cứu, và viết bản thảo cho bài báo này.

TÀI LIỆU THAM KHẢO

1. Madnick S, Wang R & Xian X. The design and implementation of a corporate householding knowledge processor to improve data quality. *Journal of Management Information Systems* 2004; 20(1):41-9; Available from: <https://doi.org/10.1080/07421222.2003.11045772>.
2. Reid A & Catterall M. Invisible data quality issues in a CRM implementation. *Journal of Database Marketing & Customer Strategy Management* 2005; 12:305-314; Available from: <https://doi.org/10.1057/palgrave.dbm.3240267>.
3. Loshin D. *Enterprise Knowledge Management: The Data Quality Approach*. The 1st edition. The Morgan Kaufmann 2001; Available from: <https://doi.org/10.1016/B978-012455840-3/50001-7>.
4. Chen M, Song M, Han J & E.Haihong. Survey on data quality. *2012 World Congress on Information and Communication Technologies* 2012; 1009-1013; Available from: <https://doi.org/10.1109/WICT.2012.6409222>.
5. Gartner. *How to Improve Your Data Quality*. 2021; Available from: <https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality>.
6. Meng XF & Ci X. Big Data Management: Concepts, Techniques and Challenges. *Journal of Computer Research and Development* 2013; 50:146-169;.
7. Watts S, Shankaranarayanan G & Even A. Data quality assessment in context: A cognitive perspective. *Decision support systems* 2009; 48(1):202-211; Available from: <https://doi.org/10.1016/j.dss.2009.07.012>.
8. Marsh R. Drowning in dirty data? It's time to sink or swim: a four-stage methodology for total data quality management. *Database Marketing & Customer Strategy Management* 2005; 12(2):105-12; Available from: <https://doi.org/10.1057/palgrave.dbm.3240247>.
9. Haug A & Stentoft Arlbjorn J. Barriers to master data quality. *Journal of Enterprise Information Management* 2011; 24(3):288-303; Available from: <https://doi.org/10.1108/1741039111122862>.
10. Piprani B, Ernst D. A Model for Data Quality Assessment. In: Meersman, R., Tari, Z., Herrero, P. (eds) *On the Move to Meaningful Internet Systems: OTM 2008*. Lecture Notes in Computer Science 2008; 5333. Springer, Berlin, Heidelberg; Available from: https://doi.org/10.1007/978-3-540-88875-8_99.
11. Xiang JY, Lee S & Kim JK. Data quality and firm performance: empirical evidence from the Korean financial industry. *Information Technology Management* 2013; 14:59-65; Available from: <https://doi.org/10.1007/s10799-012-0145-6>.
12. Cai L & Zhu Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal* 2015; 14:2; Available from: <http://doi.org/10.5334/dsj-2015-002>.
13. O'Brien T. Accounting for Data Quality in Enterprise Systems. *Procedia Computer Science* 2015; 64:442-449; Available from: <https://doi.org/10.1016/j.procs.2015.08.539>.
14. Laranjeiro N, Soydemir SN and Bernardino J. A Survey on Data Quality: Classifying Poor Data. *2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC)* 2015; 179-188; Available from: <https://doi.org/10.1109/PRDC.2015.41>.
15. Tayi GK and Ballou D. Examining Data Quality. *Communication of the ACM* 1998; 41(2):54-57; Available from: <https://doi.org/10.1145/269012.269021>.
16. Juran JM & Godfrey AB. *Juran's Quality Handbook*. Fifth edition, p. 22, McGraw-Hill 1999;.
17. Ballou DP & Pazer HL. Modeling Data and Process Quality in Multiple Input and Output Information Systems. *Management Science* 1985; 31(2):150-162; Available from: <https://doi.org/10.1287/mnsc.31.2.150>.
18. Wand Y & Wang RY. Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM* 1996; 90(11):86-95; Available from: <https://doi.org/10.1145/240455.240479>.
19. Redman TC. The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM* 1998; 41(2):79-82; Available from: <https://doi.org/10.1145/269012.269025>.
20. Wang RY & Strong DM. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 1996; 12(4):5-33; Available from: <https://doi.org/10.1080/07421222.1996.11518099>.
21. Wang RY. A Product Perspective on Total Data Quality Management. *Communication of the ACM* 1998; 41(2):58-66; Available from: <https://doi.org/10.1145/269012.269022>.
22. DAMA. *The Data Management Body of Knowledge*. The 2nd edition. DAMA International 2017;.
23. Berti-Equille L & Dasu T. *Data Quality Mining: New Research Directions*. KDD 2009; Paris, France;.
24. Cichy C & Rass S. An Overview of Data Quality Frameworks. *IEEE Access* 2019; Available from: <https://doi.org/10.1109/ACCESS.2019.2899751>.
25. English LP. *Improving Data Warehouse and Business Information Quality*. John Wiley & Sons, New York 1999;.
26. Ryu KS, Park JS & Park HL. A Data Quality Management Maturity Model. *ETRI Journal* 2006, 28(2):191-200; Available from: <https://doi.org/10.4218/etrij.06.0105.0026>.
27. Canadian Institute for Health Information. *The CIHI Data Quality Framework*, Ottawa 2005;.
28. Kim SH & Lee CS. A Master Data Quality Management Framework. *Entrée Journal of Information Technology* 2010; 9(2);.
29. Tran TV, Kim SH & Hsiao M. Data Stewardship and Flow Management for Data Quality Improvement. *The International Conference on Advanced Information Technology*. 2010, 1(1):1-13; Available from: http://163.17.20.49/AIT2013/ft_061.pdf.
30. Shankaranarayanan G & Wang R. IPMAP: Representing the Manufacture of an Information Product. *Proceedings of the 2000 Conference on Information Quality*, 2000;.
31. Scannapieco M, Pernici B & Pierce E. IP-UML: Towards a methodology for Quality Improvement based on the IP-MAP Framework. *Proceedings of the 7th International Conference on Information Quality* 2005;.
32. Fürber, C. *Data Quality Management with Semantic Technologies*, 1st ed.; Springer Gabler: Wiesbaden, Germany, 2015; Available from: https://doi.org/10.1007/978-3-658-12225-6_1.
33. Madnick SE, Wang RY, Lee YW & Zhu H. Overview and Framework for Data and Information Quality Research. *ACMJ Data Information* 2009; (1)1; Available from: <https://doi.org/10.1145/1515693.1516680>.
34. Gartner. *Measuring the Business Value of Data Quality*. Gartner 2011;.
35. Eppler MJ & Helfert M. A Classification and Analysis of Data Quality Costs. *Proceedings of the Ninth International Conference on Information Quality (ICIQ - 04)* 2004;.
36. Spruit M & Linden VVD. BIDQI: The Business Impacts of Data Quality Interdependencies model. *Technical Report UU-CS-2019-001*, 2019;.
37. Ledger Insights. Gartner by 2023 blockchain can improve data quality by 50. 2022; Available from: <https://www.ledgerinsights.com/gartner-by-2023-blockchain->

- [can-improve-data-quality-by-50/](#).
38. Tran TV & Trinh HHH. Blockchain technology for sustainable supply chains of agri-food in Vietnam: a SWOT analy-

sis. Science & Technology Development Journal-Economics-Law and Management 2021; 5(1):1278-1289; Available from: <https://doi.org/10.32508/stdjelm.v5i1.675>.

The impacts of poor data quality on business performance

Trinh Hoang Hong Hue¹, Thien Vu Tran^{2,*}



Use your smartphone to scan this QR code and download this article

ABSTRACT

In the digital era, data plays a pivotal role in the operational transactions of businesses across the globe. Along with human and capital resources, data is considered an invaluable asset for enterprises driving the success of business performance. Nevertheless, poor-quality data as inputs will negatively impact on the quality of the output and outcomes are the quality of decisions at individual, organizational, and corporate levels. Poor data quality will be a barrier for an enterprise to build an effective strategy and efficient management across its business function. Since the 1990s, businesses, and academic institutions in advanced economies such as the USA, France, Switzerland, Australia, and South Korea have focused on developing, implementing, and using data quality management systems to improve business operations. However, Vietnam has given less attention to data quality issues. This paper aims to summarize the introduction of data quality, the problems with data quality, and the impacts of poor data quality on operations at functional and strategic levels using the qualitative approach – literature synthesis and secondary data analysis. Thus, we propose some policy prescriptions for Vietnam's industries and institutions alike.

Key words: Data quality, functional operations, enterprises, Vietnam

¹University of Economics and Law,
VNU-HCM, Vietnam

²Vietnam-Korea University of
Information and Communication
Technology, The University of Danang,
Vietnam

Correspondence

Thien Vu Tran, Vietnam-Korea University
of Information and Communication
Technology, The University of Danang,
Vietnam

Email: vutrandn@gmail.com;
ttvu@vku.udn.vn

History

- Received: 24-9-2022
- Accepted: 07-6-2023
- Published: 30-6-2023

DOI : <https://doi.org/10.32508/stdjelm.v7i2.1134>



Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



Cite this article : Hue T H H, T V T. **The impacts of poor data quality on business performance** . *Sci. Tech. Dev. J. - Eco. Law Manag.*; 2023, 7(2):4381-4392.