

# การพัฒนาคุณภาพข้อมูลเพื่อเพิ่มประสิทธิภาพองค์กร

## Data Quality Improvement for Improving Organizations Efficiency

สุกัญญา เรืองสุวรรณ<sup>1</sup>

Sukanya Ruangsuwan

### บทคัดย่อ

คุณภาพของข้อมูลเป็นปัญหาที่สำคัญในการดำเนินงานขององค์กร ข้อมูลที่คุณภาพไม่ดีย่อมทำให้องค์กรเกิดต้นทุนมูลค่าสูง ดังนั้นบทความนี้จึงได้วิเคราะห์งานวิจัยและบทความที่เกี่ยวข้องกับปัญหาคุณภาพข้อมูล โดยการวิเคราะห์บทความ 8 เรื่อง เพื่อทำความเข้าใจปัญหาคุณภาพข้อมูล เพื่อพัฒนาคุณภาพของข้อมูล ผลการศึกษาพบว่า คุณภาพข้อมูลระบุได้จากหลายมิติคุณภาพ เช่น ความถูกต้อง ความสอดคล้อง ความครบถ้วน ความทันเวลา และความเกี่ยวข้อง ผู้ใช้ต้องเลือกมิติคุณภาพที่สำคัญสำหรับข้อมูลและประเมินคุณภาพของข้อมูล โดยการสร้างตัวชี้วัดให้เหมาะสมกับข้อมูลเหล่านั้น จากนั้นองค์กรสามารถเริ่มต้นดำเนินการพัฒนาคุณภาพข้อมูล เพื่อปรับปรุงประสิทธิภาพขององค์กรโดยปฏิบัติตามแนวทางที่นำเสนอไว้ในบทความนี้

### Abstract

Data quality is a critical issue in organizational operations. The poor quality data is costing large amounts of money to organizations. Therefore, this article is the reviews of researches and articles relating to data quality issue. By doing document analysis upon eight articles, it is very important to understand data quality issue in order to improve data quality. The study results that data quality has been identified in many dimensions such as accuracy, consistency, completeness, timeliness and relevance. Users have to select significant dimensions for data and assess data quality by construct metrics which is appropriate to them. Then organizations initiate data quality to improve their efficiency by follow the framework that proposed in this article.

คำสำคัญ: คุณภาพข้อมูล คุณภาพสารสนเทศ

Keywords: Data quality, Information quality

<sup>1</sup> ผู้ช่วยศาสตราจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

## บทนำ

ในปัจจุบันทุกประเทศทั่วโลกต่างพัฒนาเข้าสู่ยุคแห่งสังคมความรู้ (Knowledge Society) โดยต้องมีการพัฒนาระบบเศรษฐกิจไปสู่ระบบเศรษฐกิจแบบฐานความรู้ (Knowledge Base Economy) ควบคู่กันไปด้วย ซึ่งองค์กรสมัยใหม่ (Modern Organization) ในยุคนี้จำเป็นต้องเน้นการสร้างความรู้และนวัตกรรมใหม่เป็นหลัก ดังนั้นสิ่งสำคัญที่องค์กรต้องใช้ในการบริหารสมัยใหม่ เพื่อให้การบริหารงานเป็นไปอย่างมีประสิทธิภาพก็คือ สารสนเทศและความรู้ซึ่งถือเป็นทุนทางปัญญาขององค์กรที่ต้องได้รับการจัดการอย่างเป็นระบบ ไม่ว่าจะเป็นการใช้สารสนเทศเพื่อการวางแผน การจัดการ หรือแม้แต่การใช้ในการทำงานประจำวัน (กุลธิดา ท้วมสุข, ม.ป.ป.)

สารสนเทศและความรู้ที่นำไปใช้ในการบริหารงานในองค์กรย่อมมาจากการมีข้อมูลที่มีคุณภาพ ข้อมูลจึงเป็นพื้นฐานที่สำคัญของประสิทธิภาพองค์กร เนื่องจากทุกกิจกรรมที่เกิดขึ้นในองค์กรย่อมจำเป็นต้องใช้ข้อมูลทั้งสิ้น (Haug, Arlbjorn and Pedersen, 2009) บางองค์กรลงทุนไปกับการสร้างระบบสารสนเทศเป็นจำนวนเงินมหาศาล เช่น การนำระบบบริหารทรัพยากรองค์กร (Enterprise Resource Planning, ERP) และเทคโนโลยีเพื่อสนับสนุนการทำงานด้านต่างๆ มาใช้หวังว่าจะช่วยให้ได้เปรียบในการแข่งขัน (Xu, Nord,

Brown and Nord, 2002) แต่ถ้าหากมองข้ามคุณภาพของข้อมูลที่อยู่ในระบบแล้วสิ่งที่ลงทุนไปนั้นก็เท่ากับการการสูญเปล่าโดยสิ้นเชิง โดย Lucas (2010) กล่าวว่า สำหรับธุรกิจในสหรัฐอเมริกา นั้น ต้องเสียค่าใช้จ่ายอันเนื่องมาจากปัญหาในเรื่องของคุณภาพข้อมูลถึงปีละ 600 ล้านเหรียญสหรัฐเลยทีเดียว ดังนั้นในบทความฉบับนี้จะได้กล่าวถึงความสำคัญของคุณภาพข้อมูล มิติคุณภาพข้อมูล การประเมินคุณภาพข้อมูล และกรอบการพัฒนาคุณภาพข้อมูล เพื่อให้เกิดความตระหนักถึงความสำคัญของคุณภาพข้อมูล และเป็นแนวทางสำหรับการพัฒนาข้อมูลในองค์กรให้มีคุณภาพเพื่อการตัดสินใจที่ถูกต้องต่อไป

## แหล่งในการรวบรวมเอกสาร

การศึกษาครั้งนี้เป็นการศึกษาวิเคราะห์เอกสาร โดยการคัดเลือกบทความวิจัย และบทความปริทัศน์ที่เกี่ยวข้องกับเรื่องของคุณภาพข้อมูล จากฐานข้อมูลอิเล็กทรอนิกส์ IEE/IEEE Electronic Library Online, Emerald Management Xtra175&Engineering Database และ Science Direct โดยใช้คำค้นว่า "Data Quality" โดยมีเกณฑ์ว่าเป็นบทความที่ดีพิมพ์อยู่ระหว่างปี ค.ศ. 2000 ถึง 2010 มีเนื้อหาเกี่ยวกับการจัดการคุณภาพข้อมูลในระบบสารสนเทศเป็นหลัก ได้บทความทั้งหมด 8 เรื่องดังรายชื่อในตารางที่ 1

ตารางที่ 1 บทความที่เกี่ยวข้องกับคุณภาพข้อมูลในระบบสารสนเทศ

ผู้วิจัย	วารสาร	วัตถุประสงค์	วิธีการศึกษา
Lucas (2010)	Information Systems and Technologies	ศึกษาการจัดการคุณภาพข้อมูลในบริษัท เหตุใดจึงริเริ่มการจัดการคุณภาพข้อมูล และทำอย่างไร	ศึกษาจากกรณีศึกษาแล้วเปรียบเทียบกับทฤษฎีจากวรรณกรรม
Haug, Arlbjorn and Pedersen (2009)	Industrial Management & Data Systems	ศึกษามิติคุณภาพข้อมูลที่เกี่ยวข้องกับการประเมินคุณภาพข้อมูลใน ERP	สร้างตัวแบบเพื่อแบ่งคุณภาพข้อมูลของระบบ ERP ตรวจสอบความถูกต้องของตัวแบบและหาความสัมพันธ์ของมิติคุณภาพ โดยศึกษาจากกรณีศึกษา 3 แห่ง

ผู้วิจัย	วารสาร	วัตถุประสงค์	วิธีการศึกษา
Marshall and Harpe (2009)	Journal of Information Management	สร้างความตระหนักในปัจจัยที่ทำให้การใช้ประโยชน์จากสารสนเทศเป็นเรื่องยาก ตรวจสอบว่าปัจจัยเหล่านั้นมีผลต่อกระบวนการตัดสินใจ ในสภาพแวดล้อม BI	ใช้การวิจัยเชิงคุณภาพ โดยการสัมภาษณ์แบบไม่มีโครงสร้าง สมาชิกของฝ่าย BI ในองค์กรขนาดใหญ่ในแอฟริกาใต้
Scannapieco, Virgillito, Marchetti, Mecella and Baldoni (2004)	Information Systems	เสนอสถาปัตยกรรมสำหรับการจัดการคุณภาพข้อมูลในระบบสารสนเทศรวม	สร้างกรอบการทำงาน และทดสอบแนวคิดโดยการใช้กับตัวอย่างข้อมูล
Moody and Shanks (2003)	Information Systems	อธิบายการใช้กรอบการทำงานเพื่อพัฒนาตัวแบบคุณภาพข้อมูล และคุณภาพในการวิเคราะห์ข้อมูล	วิจัยเชิงปฏิบัติการโดยการใช้กรอบการทำงานกับองค์กรขนาดใหญ่ 2 แห่ง ในออสเตรเลีย แล้วจึงวิเคราะห์ความแตกต่างของตัวแบบข้อมูลที่พัฒนาขึ้น กับผู้เชี่ยวชาญ แล้วพัฒนาเป็นระบบอัตโนมัติ
Pipino, Lee and Wang (2002)	Communications of the ACM	นำเสนอหลักการเพื่อช่วยองค์กรในการพัฒนาตัวชี้วัดคุณภาพข้อมูล	วิเคราะห์เอกสาร และสรุปเป็นหลักการ
Xu, Nord, Brown and Nord (2002)	Industrial Management & Data Systems	ศึกษาปัญหาเกี่ยวกับคุณภาพข้อมูลในการใช้ระบบ ERP ภายในองค์กร และกำหนดปัจจัยสำคัญที่องค์กรควรมุ่งเน้นเพื่อคุณภาพข้อมูล	ศึกษากรณีศึกษาในองค์กรขนาดใหญ่ 2 แห่งในออสเตรเลีย
Divorski and Scheirer (2001)	Evaluation and Program Planning	สร้างกรอบการทำงานเพื่อประเมินและพัฒนาคุณภาพข้อมูล	ศึกษาหน่วยงานของรัฐ 6 แห่ง เพื่อตรวจสอบวิธีการและการนำเสนอบริบทของการวัดประสิทธิภาพ

จากบทความทั้ง 8 เรื่องข้างต้นทำการศึกษา โดยการวิเคราะห์เอกสารในส่วนของบริษัทย่อย ผลการศึกษา การอภิปรายผล และสรุปผลการศึกษา ใน 4 ประเด็นหลัก ได้แก่ ความหมายและความสำคัญของคุณภาพข้อมูล มิติคุณภาพข้อมูล การประเมินคุณภาพข้อมูล และแนวทางการพัฒนาคุณภาพข้อมูล โดยการบันทึกความถี่ของมิติคุณภาพที่เกิดจากแนวคิดในบทความนั้นๆ แล้ววิเคราะห์มิติคุณภาพที่มีความถี่มากที่สุด

### ความหมายและความสำคัญของคุณภาพข้อมูล

จากที่ได้กล่าวมาแล้วว่า ข้อมูลเป็นพื้นฐานที่สำคัญขององค์กร ข้อมูลที่มีคุณภาพย่อมทำให้กิจกรรมต่างๆในองค์กร มีประสิทธิภาพตามไปด้วย อย่างไรก็ตาม มีผู้ให้ความหมายของคุณภาพข้อมูลไว้หลายประเด็น Lucas (2010) ได้ให้ความหมายของคำว่า "คุณภาพข้อมูล" คือ ข้อมูลที่มีคุณภาพสูง พอที่จะใช้ในการดำเนินงานตัดสินใจ และวางแผนได้ ซึ่ง Marshall and Harpe (2009) กล่าวในประเด็นที่มองถึงการนำไปใช้งานว่า

คุณภาพข้อมูล คือ ข้อมูลที่เหมาะสมกับการใช้งาน และตรงตามวัตถุประสงค์ โดยผู้ใช้สามารถเข้าถึงข้อมูลได้อย่างถูกต้องด้วย

อย่างไรก็ตาม หากข้อมูลขององค์การไม่มีคุณภาพ ย่อมทำให้เกิดปัญหาในหลายด้าน ดังนี้ ดันลูกค้า Lucas (2010) พบว่าบริษัทที่ศึกษาได้เห็นความสำคัญของการตรวจสอบคุณภาพข้อมูลเมื่อพวกเขาต้องส่งบิลเรียกเก็บเงินไปยังลูกค้า แต่ส่งไปผิดที่เนื่องจากในคลังข้อมูลของบริษัทเก็บที่อยู่ของลูกค้าผิดพลาดทำให้บริษัทต้องเสียค่าใช้จ่ายในการส่งเอกสารกลับไปกลับมาหลายครั้ง ลูกค้าก็ไม่พอใจที่ได้รับเอกสารล่าช้า พวกเขาจึงเริ่มตระหนักถึงความสำคัญของคุณภาพข้อมูลในด้านค่าใช้จ่าย Haug, et al. (2009) ยกตัวอย่างข่าวจาก NBC รายงานว่า ข้อมูลการแจกแอสแตมปีอาหารของสหรัฐอเมริกาขาดความทันสมัยถึงขนาดส่งแอสแตมปีอาหารให้กับคนที่ตายไปแล้ว ทำให้รัฐบาลต้องเสียค่าใช้จ่ายนับล้านดอลลาร์ทีเดียว ในด้านการจัดการ Marshall and Harpe (2009) พบว่าในการสั่งซื้อสินค้าผู้จัดการ

จำเป็นต้องมีข้อมูลที่คำนวณจากระดับสินค้าที่มีอยู่ในมือ หากข้อมูลนี้มีข้อผิดพลาดการคำนวณจำนวนสั่งซื้อก็จะผิดพลาดไปด้วยเช่นกัน แล้วถ้าสั่งซื้อสินค้ามากเกินไปก็ทำให้เสียค่าใช้จ่ายในการเก็บรักษาสินค้า แต่ถ้าสั่งมาน้อยเกินไปก็ทำให้สินค้าไม่พอขายซึ่งทำให้เสียรายได้และโอกาสในการขาย ดังนั้นคุณภาพข้อมูลจึงมีความสำคัญอย่างยิ่งในการดำเนินกิจกรรมต่างๆ ในองค์การ

### มิติคุณภาพของข้อมูล

มิติคุณภาพข้อมูลคือ คุณลักษณะที่มีอยู่ในข้อมูลตามธรรมชาติ (Scannapieco et al., 2004) โดยที่คุณภาพข้อมูลนั้นมีความหมายต่อผู้บริโภคข้อมูลแตกต่างกันไปขึ้นอยู่กับการใช้งานไปใช้ (Lucas, 2010) ซึ่งจากการวิเคราะห์เอกสาร พบว่านักวิจัยแต่ละคนให้ความสำคัญกับมิติคุณภาพที่แตกต่างกันไป จาก 8 บทความนี้พบมิติคุณภาพข้อมูลถึง 21 ลักษณะ โดยสามารถสรุปแนวคิดในเรื่องมิติคุณภาพจากแต่ละบทความได้ดังตารางที่ 2

ตารางที่ 2 มิติคุณภาพของข้อมูลที่สำคัญตามแนวคิดของนักวิจัยแต่ละคน

มิติคุณภาพของข้อมูล	Lucas (2010)	Haug et al. (2009)	Marshall & Harpe (2009)	Scannapieco et al. (2004)	Moody & Shanks (2003)	Pipino, et al. (2002)	Xu et al. (2002)	Divorski & Scheirer (2001)	รวมความถี่
Accuracy / Free of Error	✓	✓	✓	✓	✓	✓	✓	✓	8
Consistency / Integration	✓		✓	✓	✓	✓	✓	✓	7
Completeness	✓	✓		✓	✓	✓	✓	✓	7
Timeliness	✓					✓	✓	✓	4
Relevancy	✓	✓			✓	✓			4
Accessibility / Availability		✓	✓			✓			3
Understandability			✓		✓	✓			3
Ease of Use					✓			✓	2
Concise Representation		✓				✓			2
Value - Added		✓				✓			2

มิติคุณภาพของข้อมูล	Lucas (2010)	Haug et al. (2009)	Marshall & Harpe (2009)	Scannapieco et al. (2004)	Moody & Shanks (2003)	Pipino, et al. (2002)	Xu et al. (2002)	Divorski & Scheirer (2001)	รวมความถี่
Currency				✓					1
Validity								✓	1
Appropriate						✓			1
Amount of Data									
Believability						✓			1
Ease of Manipulation						✓			1
Interpretability						✓			1
Objectivity						✓			1
Reputation						✓			1
Security						✓			1
Unambiguous		✓							1
Meaningful		✓							1

มิติคุณภาพที่มีความถี่สูงสุด 5 ลำดับแรก ได้แก่

**1) ความถูกต้องของข้อมูล (Accuracy)** เป็นมิติคุณภาพที่บทความทั้ง 8 เรื่องมีความเห็นตรงกันว่า เป็นเรื่องที่สำคัญต่อการแสดงคุณภาพของข้อมูล ความถูกต้องของข้อมูลหรือข้อมูลที่ปราศจากข้อผิดพลาดคลาดเคลื่อน หมายถึงขอบเขตที่ข้อมูลถูกต้องเชื่อถือได้ (Pipino et al., 2002) เช่น ลูกค้าชื่อ John แต่พิมพ์เป็น Jhn ถือว่ามีความถูกต้องต่ำ (Scannapieco, et al., 2004) ข้อมูลที่ไม่ถูกต้องนี้เป็นสาเหตุหลักของปัญหาอื่นๆ ที่จะตามมา (Marshall and Harpe, 2009)

**2) ความสอดคล้อง (Consistency)** เป็นเรื่องรองลงมาที่คนให้ความสำคัญ ความสอดคล้องในความหมายของ Pipino, et al. (2002) คือ ข้อมูลถูกนำเสนอในรูปแบบเดียวกัน โดย Marshall and Harpe (2009) ยกตัวอย่างของความไม่สอดคล้องกันเช่น ชื่อที่มาจากแหล่งหนึ่งกำหนดความยาวไว้ 20 ตัวอักษร แต่อีกแหล่งหนึ่งกำหนดไว้ 40 ตัวอักษร หรือรูปแบบการนำเสนอวันที่เป็นแบบ dd/mm/yyyy ที่หนึ่งส่วนอีกที่หนึ่งกำหนดเป็น mm/dd/yyyy เป็นต้น ส่วน Scannapieco, et al.(2004) มองว่าข้อมูลที่มีหลายๆ ค่าไม่ควรถัดแย้ง

กัน เช่น ชื่อ John ควรเป็น เพศหญิง จึงจะสอดคล้องกัน สำหรับ Divorski and Scheirer (2001) กลับมองว่าข้อมูลที่รวบรวมมานั้นต้องใช้วิธีการและนิยามเดียวกันทุกแหล่งข้อมูลและทุกช่วงเวลาจึงจะเกิดความสอดคล้องกัน

**3) ความครบถ้วนสมบูรณ์ (Completeness)** คือ ข้อมูลไม่ขาดหาย กว้างพอและลึกพอสำหรับการใช้งาน (Pipino, et al., 2002) ข้อมูลครบทั้งหมดตามที่ผู้ใช้ต้องการ (Moody and Shanks, 2003) เช่น ข้อมูล E-mail ของลูกค้าบางคนขาดหายไป อาจเป็นไปได้สองความหมายคือ ลูกค้าคนนั้นไม่มี E-mail นั้นไม่มีผลกระทบกับความครบถ้วนสมบูรณ์ แต่ถ้าลูกค้ามี E-mail แต่ไม่ได้กรอกลงไปถือว่าความครบถ้วนสมบูรณ์ต่ำ (Scannapieco, et al., 2004) นอกจากนี้ Divorski and Scheirer (2001) ยังมองถึงความเพียงพอของข้อมูลที่เก็บรวบรวมมาจากประชากรหรือตัวอย่างว่ามีสัดส่วนที่เพียงพอหรือไม่อีกด้วย

**4) ความทันต่อการใช้งานหรือทันเวลา (Timeliness)** หมายถึงข้อมูลเป็นปัจจุบันทันสมัยเพียงพอต่อการใช้งาน (Pipino, et al., 2002) หรือมีข้อมูลทันต่อการใช้งานทุกครั้งตามที่ผู้ใช้ต้องการ (Divorski and

Scheirer, 2001) เช่น ต้องการทราบยอดขายของบริษัท ประจำไตรมาสที่ 1 แต่กลับมีข้อมูลถึงเดือนมกราคม เท่านั้นแสดงว่าข้อมูลขาดความทันเวลา

**5) ความเกี่ยวข้องกัน (Relevancy)** หมายถึง ข้อมูลนั้นสามารถนำไปใช้ได้กับงานที่ทำอยู่ (Pipino et al., 2002) ภายใต้ข้อจำกัดด้านเวลา งบประมาณและเทคโนโลยีของกิจกรรม (Moody and Shanks, 2003) เช่น ข้อมูล E-mail ของลูกค้าที่มีอยู่ในฐานข้อมูลสามารถนำไปใช้เพื่อส่งโปรโมชั่นประจำเดือนให้กับลูกค้าได้

สำหรับมิติคุณภาพอื่นๆ อีก 16 มิติ (Pipino et al., 2002) (Divorski and Scheirer, 2001) (Haug et al., 2009) ซึ่งผู้ใช้อาจเลือกใช้มิติคุณภาพเหล่านี้ เพื่อบอกคุณภาพข้อมูลตามความจำเป็นของงานที่ทำ สามารถอธิบายความหมายพอสังเขปได้ดังนี้

1. **Accessibility / Availability** คือ มีข้อมูลนั้นอยู่ หรือเข้าถึงได้อย่างรวดเร็ว
2. **Understandability** คือ ข้อมูลนั้นเข้าใจได้ง่าย
3. **Ease of Use** คือ ข้อมูลสามารถนำไปใช้งานได้ง่าย
4. **Concise Representation** คือ ข้อมูลนั้นเป็นตัวแทนที่กระชับ
5. **Value-Added** คือ ข้อมูลนั้นมีประโยชน์และทำให้มีข้อได้เปรียบในการใช้
6. **Currency** คือ มีข้อมูลนั้นเกิดขึ้นเป็นประจำ
7. **Validity** คือ ข้อมูลที่เป็นตัวแทนของการทำงานจริงอย่างเพียงพอ
8. **Appropriate Amount of Data** คือ ข้อมูลนั้นมีจำนวนเหมาะสมเพียงพอที่จะใช้งานได้
9. **Believability** คือ ข้อมูลนั้นเป็นความจริงและน่าเชื่อถือ
10. **Ease of Manipulation** คือ ข้อมูลนั้นสามารถจัดการได้ง่าย และนำไปประยุกต์กับงานต่างๆ ได้
11. **Interpretability** คือ ข้อมูลอยู่ในรูปของภาษา สัญลักษณ์ และหน่วยที่เหมาะสม มีความหมายชัดเจน
12. **Objectivity** คือ ข้อมูลนั้นไม่เอนเอียง ไม่มีอคติ และเป็นกลาง

13. **Reputation** คือ ข้อมูลเป็นที่ยอมรับในแง่ของแหล่งที่มาและเนื้อหา

14. **Security** คือ การเข้าถึงข้อมูลมีความเข้มงวดเหมาะสมกับการคงไว้ซึ่งความปลอดภัย

15. **Unambiguous** คือ ข้อมูลไม่กำกวมสงสัยไม่มี 2 สถานะเกิดขึ้นพร้อมกัน

16. **Meaningful** คือ ข้อมูลต้องมีความหมาย

### การประเมินคุณภาพข้อมูล

ก่อนที่จะทำการประเมินคุณภาพข้อมูลต้องกำหนดวัตถุประสงค์ที่ต้องการประเมินให้ชัดเจนก่อน เลือกข้อมูลและมิติคุณภาพที่ข้อมูลนั้นจำเป็นต้องมีแล้ว จึงสร้างตัวชี้วัดการประเมิน โดย Pipino, et al. (2002) ได้เสนอรูปแบบตัวชี้วัดไว้ 3 ประเภทได้แก่ ค่าสัดส่วนอย่างง่าย, ค่าต่ำสุด-สูงสุดในการปฏิบัติการ และค่าเฉลี่ยถ่วงน้ำหนัก

**ค่าสัดส่วนอย่างง่าย (Simple Ratio)** เป็นการหาสัดส่วนของผลลัพธ์ที่ต้องการกับผลลัพธ์ทั้งหมดซึ่งมักจะใช้กับ มิติความถูกต้อง ความครบถ้วนสมบูรณ์ และความสอดคล้อง เป็นต้น (Pipino, et al., 2002) ในเรื่องความถูกต้องสามารถสร้างตัวชี้วัดที่ใช้ประเมินได้จากจำนวนความผิดพลาดที่เกิดขึ้นในการใช้ข้อมูลหารด้วยจำนวนข้อมูลทั้งหมดแล้วลบออกจาก 1 เช่น มีลูกค้าจำนวน 1000 คน มีการพิมพ์ที่อยู่ผิดไป 200 คน จะได้ ค่าตัวชี้วัดความถูกต้องเป็น 0.80 (คำนวณจาก  $1 - (200/1000)$ ) ส่วนความครบถ้วนสมบูรณ์และความสอดคล้องก็เช่นเดียวกัน ประเมินในรูปของจำนวนของข้อมูลที่ขาดหายไปจากความต้องการหรือจำนวนข้อมูลที่สอดคล้องกับความต้องการ โดยสามารถแสดงในรูปของร้อยละของข้อมูลทั้งหมด (Moody and Shanks, 2003)

**ค่าต่ำสุดหรือสูงสุดในการปฏิบัติการ (Min or Max Operation)** โดยให้ผู้ใช้ให้คะแนนคุณภาพข้อมูลในมิติที่ต้องการอยู่ในช่วง 0 ถึง 1 แล้วกำหนดค่าต่ำสุด (หรือสูงสุด) ที่ยอมรับได้ หากค่าเฉลี่ยของคะแนนคุณภาพที่ได้ต่ำกว่า (หรือสูงกว่า) เกณฑ์ที่กำหนดไว้ ก็ถือว่าข้อมูลมีคุณภาพ ซึ่งสามารถนำไปใช้วัดมิติคุณภาพด้านความ

ทันเวลา ความน่าเชื่อถือ และการเข้าถึงได้ เป็นต้น (Pipino, et al., 2002) ตัวอย่างเช่น กำหนดคะแนนความทันต่อเวลาไว้ที่อย่างน้อย 0.81 ถ้าผู้ใช้จำนวน 100 คน ประเมินให้คะแนนเฉลี่ยต่ำกว่านี้ถือว่าข้อมูลไม่มีคุณภาพ หรืออาจใช้การให้คะแนนเป็น 7 ระดับตามแบบลิเคิร์ต สเกลตามคำแนะนำของ Moody and Shanks (2003) ก็ได้

### ค่าเฉลี่ยถ่วงน้ำหนัก (Weighted Average)

ใช้ในกรณีมีข้อมูลหลายๆ ตัวแปร โดยที่องค์กรจะต้องกำหนดน้ำหนักให้กับตัวแปรแต่ละตัวแปรตามความสำคัญแล้วนำไปหาค่าเฉลี่ยถ่วงน้ำหนัก ค่าถ่วงน้ำหนักแต่ละตัวให้กำหนดอยู่ในช่วง 0 ถึง 1 และทุกตัวมีค่ารวมกันเท่ากับ 1 ซึ่งวิธีการสร้างตัวชี้วัดแบบนี้้องค์กรจะต้องสามารถบอกได้ว่าตัวแปรใดมีความสำคัญเพียงใด จึงจะสามารถกำหนดค่าน้ำหนักได้อย่างเหมาะสม (Pipino, et al., 2002)

เมื่อสามารถสร้างตัวชี้วัดได้แล้วองค์กรสามารถนำตัวชี้วัดเหล่านั้นไปใช้เพื่อบอกสถานะของคุณภาพข้อมูลได้ว่ามีคุณภาพมากน้อยเพียงใด

### แนวทางการพัฒนาคุณภาพข้อมูล

เมื่อองค์กรเห็นความสำคัญของการมีข้อมูลที่มีคุณภาพ องค์กรควรจะมีการกำหนดกรอบในการพัฒนาคุณภาพของข้อมูล ทั้งนี้ Divorski and Scheirer (2001) ได้เสนอกรอบแนวทางในการพัฒนาคุณภาพข้อมูลขององค์กรไว้ 4 ขั้นตอนดังนี้

1) **ผลักดันให้องค์กรมุ่งมั่นและเพิ่มความสามารถเพื่อพัฒนาคุณภาพข้อมูล** โดยการสื่อสารและทบทวนขั้นตอนในการเก็บรวบรวมข้อมูล ซึ่ง Xu, et al. (2002) มองว่าการสื่อสารเป็นเรื่องที่สำคัญเพื่อให้คนในองค์กรมีความเข้าใจในเรื่องคุณภาพข้อมูลได้ถูกต้องตรงกัน โดยเฉพาะการสื่อสารจากผู้บริหารระดับสูงเพื่อสนับสนุนให้เกิดคุณภาพข้อมูล โดยการกำหนดปรัชญาข้อมูลองค์กร (Corporate Data Philosophy) เช่นเดียวกับการกำหนดวิสัยทัศน์ และนโยบายต่างๆ ขององค์กร การกำหนดปรัชญาข้อมูลองค์กร หมายถึง

การที่องค์กรมองข้อมูลเป็นสินทรัพย์อย่างหนึ่งที่มีมูลค่าทางธุรกิจ จากนั้นกำหนดนโยบายและกลยุทธ์ข้อมูลองค์กร (Corporate Data Policy and Strategy) แล้วกำหนดเจ้าหน้าที่ผู้รับผิดชอบข้อมูล (Data Steward) ทำหน้าที่ดูแลข้อมูลทั้งด้านเนื้อหาและคุณภาพข้อมูล (Lucas, 2010) ซึ่งเจ้าหน้าที่เหล่านี้จำเป็นต้องได้รับการฝึกอบรมและแนะนำ (Xu, et al., 2002) ในเรื่องการดูแลคุณภาพข้อมูล การเลือกตัวชี้วัดที่เหมาะสม

2) **ประเมินคุณภาพข้อมูลที่มีอยู่** โดยการทำให้การประเมินคุณภาพข้อมูลเป็นกระบวนการหนึ่งในกระบวนการทำงานปกติ อาจใช้ซอฟต์แวร์เพื่อตรวจสอบและแก้ไขข้อมูลในระบบคอมพิวเตอร์ ที่ Lucas (2010) เรียกว่าเครื่องมือคุณภาพข้อมูล (Data Quality Tools) ควบคู่ไปกับการใช้การป้อนกลับจากผู้ใช้ข้อมูลและผู้มีส่วนได้ส่วนเสียในการประเมินคุณภาพ เช่น ให้ลูกค้าเสนอแนะผ่านทางเว็บไซต์ การโทรศัพท์เพื่อสอบถามความคิดเห็นของผู้ใช้ข้อมูล เป็นต้น แล้วเปรียบเทียบข้อมูลจากแหล่งอื่นหรือให้มีการตรวจสอบจากกลุ่มอิสระหรือเจ้าหน้าที่ (Divorski and Scheirer, 2001)

3) **การตอบสนองต่อข้อจำกัดของข้อมูล** เป็นการเตือนให้ผู้ใช้ระมัดระวังการใช้ข้อมูลโดยการรายงานเกี่ยวกับข้อจำกัดเหล่านั้น หรือให้ข้อมูลเพิ่มเติมเกี่ยวกับปัญหาของข้อมูลนั้น เช่น การรายงานสถิติการวัดปริมาณแอลกอฮอล์ในเลือดของผู้ขับขี่ยานพาหนะแล้วเกิดอุบัติเหตุ มีการแจ้งเพิ่มเติมว่าข้อมูลที่แสดงนั้นอาจมีค่าสูงไป หรือเป็นเพียงการประมาณค่าเท่านั้น เป็นต้น หรือแสดงค่าของตัวชี้วัดคุณภาพประกอบ (Divorski and Scheirer, 2001)

4) **สร้างคุณภาพในการพัฒนาประสิทธิภาพข้อมูล** การพัฒนาคุณภาพข้อมูลโดยการตรวจสอบและแก้ไขข้อมูลที่ผิดพลาดนั้นไม่เพียงพอที่จะป้องกันการผิดพลาดที่จะเกิดขึ้นในอนาคต หากข้อมูลมีการเปลี่ยนแปลง ตัวชี้วัดคุณภาพก็จำเป็นต้องเปลี่ยนแปลงไปด้วย วิธีการที่จะช่วยพัฒนาตัวชี้วัดคุณภาพข้อมูลได้อย่างมีประสิทธิภาพได้แก่ ศึกษาจากงานวิจัยอื่น การขอความคิดเห็นจากผู้เกี่ยวข้องทั้งภายในและภายนอก

องค์กรการ ศึกษาจากแผนงาน เอกสาร และรายละเอียด การเก็บข้อมูลและรายงานผล การฝึกอบรมผู้กำกับดูแล คุณภาพข้อมูล (Divorski and Scheirer, 2001)

## สรุป

คุณภาพข้อมูลเป็นประเด็นสำคัญที่มีผลกระทบต่อการทำงานทั้งหมดขององค์กรการ ดังนั้นเพื่อพัฒนา ประสิทธิภาพองค์กรการให้มีขีดความสามารถในการแข่งขัน ในตลาดปัจจุบันได้ องค์กรการควรตระหนักถึงความสำคัญ ของคุณภาพข้อมูล โดยมีการกำหนดมิติคุณภาพข้อมูลที่ จำเป็นสำหรับข้อมูลซึ่งย่อมแตกต่างกันไปตามบริบท ขององค์กรการแต่ละองค์กรการ แต่สิ่งสำคัญที่ควรคำนึงถึง คือ บางครั้งการมุ่งเน้นที่มีมิติคุณภาพหนึ่งอาจ ทำให้มิติ คุณภาพอีกด้านหนึ่งลดลงไป เช่น องค์กรการต้องการให้ ข้อมูลมีความถูกต้องมากที่สุด ทำให้เสียเวลาในการ ตรวจสอบข้อมูลจนทำให้คุณภาพของมิติด้านความทัน ต่อเวลาลดลง ดังนั้นองค์กรการจำเป็นต้องหาสมดุลระหว่าง มิติคุณภาพด้านต่าง ๆ ที่สอดคล้องกับความต้องการและ ธรรมชาติขององค์กรการ อีกประการหนึ่งคือ การกำหนด ตัวชี้วัดคุณภาพซึ่งจำเป็นต้องอาศัยประสบการณ์ของ ผู้วิเคราะห์เพื่อกำหนดวิธีการสร้างตัวชี้วัดให้กับมิติ คุณภาพแต่ละด้าน และตัวชี้วัดต้องชี้บอกระดับคุณภาพ ของข้อมูลให้ได้ จึงต้องมีการกำหนดระดับคะแนน มาตรฐาน หากข้อมูลมีคุณภาพในระดับต่ำกว่ามาตรฐาน จำเป็นต้องได้รับการแก้ไข ทั้งสองเรื่องนี้นับเป็นเรื่องยาก ที่ต้องอาศัยความร่วมมือจากหลายฝ่ายในองค์กรการ ร่วมกันตัดสินใจ โดยอาศัยกรอบการพัฒนาคุณภาพ ข้อมูลดังที่เสนอไว้แล้วในบทความนี้

## เอกสารอ้างอิง

กุลชิตา ท้วมสุข. (ม.ป.ป). เอกสารประกอบการสอน วิชา 412 806 เรื่องประเด็นสาระหลักที่ 1 ประเด็นการวิจัยด้านสารสนเทศและการจัดการองค์กรการสมัยใหม่. ขอนแก่น : กลุ่มวิชาการจัดการสารสนเทศและการสื่อสาร คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น.

- Divorski, S. and Scheirer, M.A. (2001). Improving Data Quality for Performance Measures: Results from a GAO Study of Verification and Validation. **Evaluation and Program Planning**, 24, 83-94.
- Haug, A., Arlbjørn, S.A. and Pedersen, A. (2009). A Classification Model of ERP System Data Quality. **Industrial Management & Data Systems**, 109(8), 1053-1068.
- Lucas, A. (2010). Corporate Data Quality Management From Theory to Practice. **Information Systems and Technologies(CISTI)**. 5<sup>th</sup> Iberian Conference: 16-19 June 2010, 1-7.
- Marshall, L. and Harpe, R. de la. (2009). Decision Making in the Context of Business Intelligence and Data Quality. **Journal of Information Management**, 11(2), 1-15.
- Moody, D. L. and Shanks, G. G. (2003). Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework. **Information Systems**, 28, 619-650.
- Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002). Data Quality Assessment. **Communication of the ACM**, 45(April), 211-218.
- Scannapieco, M., Virgillito, A., Marchetti, C., Mecella, M. and Baldoni, R. (2004). The DaQuinCIS Architecture: a Platform for Exchanging and Improving Data Quality in Cooperative Information Systems. **Information Systems**, 29, 551-582.
- Xu, H., Nord, J. H., Brown, N. and Nord, G. D. (2002). Data Quality Issues in Implementing an ERP. **Industrial Management & Data Systems**, 102(1), 47-58.