

**คาบ 6 สถาปัตยกรรมข้อมูล**

ดร.ยรรยง เต็งอำนาจ



Slide 1 หลังจากเรียนรู้เกี่ยวกับหลักการของระบบสารสนเทศและเทคโนโลยีสารสนเทศในคาบที่ผ่านมา รวมถึงโครงสร้างพื้นฐานสำหรับระบบเหล่านี้ คาบนี้จะได้นำองค์ประกอบที่สำคัญยิ่งยวดอีกส่วนหนึ่งของระบบสารสนเทศซึ่งมักถูกมองข้ามไป นั่นคือ "ข้อมูล" โดยเน้นไปที่สถาปัตยกรรมของข้อมูลก่อน แล้วจึงค่อยแนะนำประเด็นอื่นๆ เกี่ยวกับข้อมูลในคาบถัดไป

ก่อนอื่นต้องย้ำให้ผู้เรียนเข้าใจถึงความสำคัญของข้อมูล สิ่งนี้เป็นเรื่องใหญ่มาก เพราะมักถูกมองข้ามด้วยว่าเทคโนโลยีสารสนเทศนั้นมีพัฒนาการที่รวดเร็ว มีผลิตภัณฑ์ใหม่ๆ

ออกมาอย่างต่อเนื่อง ทำให้บดบังความสำคัญของการมีข้อมูลที่ตีไปเสียหมด ผู้บริโภคจะเห็นว่าเป็นเรื่องธรรมดาที่เปิดเครื่องแล้วต้องได้ข้อมูล

หากพิจารณาพัฒนาการของระบบสารสนเทศตามที่ได้กล่าวไว้ในคาบต้นๆ จะเห็นว่าในยุคแรกเป็นยุคของคอมพิวเตอร์ ซึ่งคือกล่องที่มีเทคโนโลยีขั้นสูง เป็นของเล่นไฮเทคที่ราคาแพง น่าสนใจ แม้เมื่อพัฒนาขึ้นมาเป็นยุคของเทคโนโลยีสารสนเทศแล้ว ก็ยังมีลักษณะเป็นกล่องไฮเทคกองหนึ่ง มีความสลับซับซ้อนและราคาสูง น่าสนใจอยู่ดี แต่ในปัจจุบันเป็นยุคของระบบสารสนเทศซึ่งมีแนวโน้มในการให้บริการที่สอดคล้องกับงานขององค์กรอย่างเป็นระบบ มีการสร้างระบบงานสารสนเทศเพื่อประมวลข้อมูลออกมาเป็นสารสนเทศอย่างต่อเนื่อง ให้เทคโนโลยีขั้นสูงทางวิศวกรรมซอฟต์แวร์ ทำให้เกิดเป็นยุคที่ซอฟต์แวร์เริ่มมีราคาแพงกว่าฮาร์ดแวร์



ปัญหาคือระบบสารสนเทศที่ซับซ้อนราคาแพงเหล่านี้ ไม่เป็นประโยชน์ได้หากไม่สามารถผลิตสารสนเทศที่เป็นประโยชน์ต่อองค์กรได้ แต่สารสนเทศเหล่านั้นต้องได้มาจากการประมวลผลข้อมูล ดังนั้นคำถามสำคัญใน Slide 2 คือข้อมูลนั้นอยู่ที่ใดในระบบสารสนเทศ ได้มาจากไหน ข้อมูลนั้นอยู่ๆ จะเกิดขึ้นเองไม่ได้ ข้อมูลที่เป็นจำเพาะของธุรกิจขององค์กรนั้นไม่สามารถหาซื้อได้ จำต้องเก็บเอง สร้างเอง สะสมขึ้นเองโดยองค์กร

ข้อมูลจำเพาะขององค์กรนั้นหากไม่มีกระบวนการวิธีในการหา การสร้าง การเก็บที่เหมาะสมจะทำให้ข้อมูลเหล่านั้น "ผิด" คือไม่ตรงหรือไม่ถูกต้องกับความเป็นจริง อาจ "ซาด" คือไม่ครบถ้วน ซาดข้อมูลที่จำเป็นบางส่วนไป หรือ "ซ้า" คือข้อมูลนั้นไม่ตรงกับสภาพความเป็นจริงเพราะล่าช้า เช่นข้อมูลจรรยา หากล่าช้าไปครึ่งชั่วโมงย่อมไม่เกิดประโยชน์ต่อผู้ใช้รถใช้ถนนโดยเฉพาะในช่วงชั่วโมงเร่งด่วน ลักษณะของข้อมูลที่ผิดพลาดเหล่านี้เกิดเป็นคติพจน์ที่ว่า Garbage in คือ Garbage out

หมายถึงหากใส่ "ขยะ" เป็นข้อมูลเข้าไปในระบบสารสนเทศจะได้ "ขยะ" กลับออกมานั่นเอง

ปัจจุบันเป็นที่ยอมรับกันในวงกว้างถึงระบบสารสนเทศเพื่อการดำเนินกิจกรรมต่างๆ ทั้งด้านพาณิชย์และการใช้ชีวิตในสังคม โดยเฉพาะเมื่อมีเทคโนโลยีเว็บ 2.0 ดังได้กล่าวไว้ในคาบก่อนเกี่ยวกับเทคโนโลยีเว็บ ซึ่งช่วยให้ผู้ประกอบการสารสนเทศสามารถป้อนหรือสร้างข้อมูลของตนเองเข้าไปในระบบได้ สภาพการณ์เช่นนี้ทำให้เกิดข้อมูลจำนวนมากมหาศาล จากหลากหลายแหล่ง ทำให้เกิดปัญหาอย่างมากในการรวบรวมข้อมูลจากแหล่งเหล่านี้ การจัดเก็บข้อมูลจำนวนมากซึ่งต้องเก็บไว้เป็นระยะเวลานาน การแยกแยะหมวดหมู่ของข้อมูลที่หลากหลายซับซ้อนเหล่านี้ให้เป็นระบบเป็นระเบียบ และการวางแผนทางเพื่อให้สามารถนำข้อมูลเหล่านี้มาใช้ประโยชน์โดยผ่านการค้นคืนข้อมูลจำนวนมากนี้ตามความต้องการได้

ดังนั้นนอกจากเทคโนโลยีจะสำคัญสำหรับระบบสารสนเทศแล้ว การที่มีข้อมูลที่เหมาะสมจึงเป็นรากฐานสำคัญของการนำไปสู่สารสนเทศที่เป็นประโยชน์ขององค์กร เป็นผลลัพธ์สำคัญที่ต้องมีระบบสารสนเทศ

องค์ประกอบสำคัญของระบบสารสนเทศที่เกี่ยวกับข้อมูลนี้ มีประเด็นหลักอยู่จำนวนหนึ่งซึ่งจะได้อธิบายถึงไว้ 4 ส่วนด้วยกันตาม **Slide 3** ในคาบนี้เป็นประเด็นหลักแรกคือสถาปัตยกรรมข้อมูล โดยต้องเข้าใจธรรมชาติของข้อมูล แหล่งข้อมูลและ



การได้มาซึ่งข้อมูลที่ต้องการ ความถูกต้องของข้อมูล การเก็บข้อมูลทั้งในระยะสั้นและระยะยาว และการนำข้อมูลมาใช้งาน ประเด็นที่สองคือมาตรฐานของข้อมูลซึ่งจำต้องระบุให้ชัดเจนเพราะเมื่อระบบซับซ้อนมากขึ้น มาตรฐานจะช่วยให้ไม่สับสนและใช้งานร่วมกันได้ มาตรฐานนั้นมีส่วนของ data model คือแบบจำลองอย่างเป็นทางการของระบบข้อมูลขององค์กร และ data dictionary ซึ่งเป็นส่วนของการนิยามข้อมูลต่างๆ ที่มีใช้ในระบบสารสนเทศอย่างเป็นทางการ เป็นเอกสาร เพื่อให้เกิดความชัดเจนในระบบที่ซับซ้อนขององค์กร ประเด็นหลักที่สามคือเรื่องความมั่นคง

ปลอดภัยของข้อมูล ซึ่งหากไม่มีส่วนนี้แล้วจะทำให้ข้อมูลเสียหายหรือเกิดความผิดพลาดได้จากอุบัติเหตุหรือความประสงค์ร้ายต่างๆ ซึ่งมีทั้งในส่วนของความมั่นคงปลอดภัยทางกายภาพ เช่นการขโมยหรือทำลายอุปกรณ์ที่ใช้เก็บข้อมูล และทางตรรกเช่นการเจาะระบบผ่านทางเครือข่ายที่เรียกว่า hack นั่นเอง ส่วนประเด็นหลักที่สี่ซึ่งเป็นประเด็นสุดท้ายของคาบนี้คือเกี่ยวกับสถาปัตยกรรมในการจัดวางระบบข้อมูล ว่าจะเป็นแบบรวมศูนย์คือรวมไว้ในที่เดียว หรือเป็นแบบกระจายคือแบ่งส่วนอยู่ในหลายๆ ที่ในองค์กร ซึ่งแต่ละแบบต่างมีข้อดีและข้อเสียที่องค์กรต้องคำนึงถึง

เหตุที่ต้องคำนึง "สถาปัตยกรรม" ของข้อมูล ราวกับจะสร้างอาคารสูงสามสิบชั้นสักหลัง เป็นเพราะเมื่อระบบสารสนเทศเป็นขององค์กรขนาดใหญ่ที่ซับซ้อน จำต้องมีการวางแผน การออกแบบ การจัดสร้าง การปฏิบัติงาน และการบำรุงรักษาอย่างเป็นระบบ เหล่านี้รวมถึงข้อมูลที่เป็นหัวใจสำคัญ เป็นแก่นของระบบสารสนเทศด้วย หากไม่มีการออกแบบและวางแผนในลักษณะของสถาปัตยกรรมที่เหมาะสม เมื่อจัดสร้างระบบขึ้นจะสะเปะสะปะ ทำให้เกิดความผิดพลาดยุ่งยาก ความเสียหายที่จะตามจะยากต่อการแก้ไข

สิ่งที่ต้องคำนึงถึงคือเมื่อเป็นองค์กรขนาดใหญ่ ข้อมูลก็มีจำนวนมากมหาศาลและมีความซับซ้อนหลากหลายขึ้นเป็นเงาตามตัวไปด้วย ข้อมูลจำนวนมากเหล่านี้มีธรรมชาติที่ซับซ้อนจำต้องอาศัยสิ่งที่เรียกว่า Enterprise Data Architecture ตาม Slide 4 เข้ามาจัดการ หลักการนี้ประกอบด้วยคำสามคำ คำแรกคือ enterprise ซึ่งหมายถึงสิ่งที่จะทำนั้นเป็นระดับองค์กรซึ่งมีขนาดใหญ่ มีความซับซ้อน ทำให้ข้อมูลมีจำนวนมาก ซับซ้อน หลากหลายตามไปด้วย ธุรกิจระดับร้านขายของชำหรือที่เรียกว่าโช่วยนั้นคงไม่อาจนับเป็น Enterprise ได้ ต้องเป็นองค์กรหรือธุรกิจที่มีการแบ่งส่วนงานย่อยหลายส่วน โดยแต่ละส่วนทำหน้าที่เฉพาะด้านเพื่อสนับสนุนองค์กร เช่น ฝ่ายบัญชีการเงิน ฝ่ายผลิต ฝ่ายพัสดุและสินค้าคงคลัง ฝ่ายประชาสัมพันธ์ ฝ่ายขาย เหล่านี้จึงจะนับเป็นองค์กรที่ต้องการสถาปัตยกรรมข้อมูล



คำที่สองคือ data ซึ่งหมายรวมถึงข้อมูลที่เป็นข้อมูลดิบ คือได้มาตรงๆ ไม่มีการเสริมแต่ง เช่นจำนวนเงินที่ได้จากขายสินค้าไปชิ้นหนึ่ง กับข้อมูลสุก คือผ่านการประมวลแล้ว เช่นจำนวนเงินของการขายสินค้าของสาขาในรอบเดือนที่ผ่านมา เช่นนี้เป็นต้น ข้อมูลสุกนี้เรียกอย่างเป็นทางการว่าสารสนเทศ นั่นเอง คำที่สามคือ architecture แปลว่าสถาปัตยกรรม เป็นโครงสร้างของข้อมูลในภาพรวม หากเปรียบกับการก่อสร้าง ถ้าสร้างกระต๊อบหรือเพิง ก็ไม่ต้องมีพิธีรีตรองอย่างไร ถ้าสร้างบ้านขึ้นเดี่ยว หนึ่งห้องนอนหนึ่ง

ห้องน้ำก็อาจมีแบบแปลนง่ายๆ ช่างรับเหมาสามารถเขียนหรือกำหนดขึ้นมาได้ แต่หากจะสร้างโรงแรม คฤหาสน์ หรืออาคารสักสามสิบชั้น จำต้องมีการวางแผน การออกแบบ การกะประมาณสิ่งต่างๆ ให้มันใจก่อนสร้าง เพราะหากสร้างไปแล้วจะแก้ไขปรับเปลี่ยนอย่างไรย่อมไม่อาจทำได้ตามอำเภอใจ เช่นเปลี่ยนใจว่าไม่เอาสามสิบชั้น จะให้เป็นห้าสิบชั้นขณะที่ก่อสร้างไปได้สิบชั้น เช่นนี้ย่อมไม่อาจทำได้เพราะเสาเข็มและฐานรากที่วางเอาไว้ใต้ดินนั้นไม่อาจรับน้ำหนักอาคารถึงห้าสิบชั้นได้ และไม่อาจรื้อถอนขึ้นมาทำใหม่ได้โดยง่าย เช่นนี้เป็นต้น ในความเป็นจริงแล้ว นอกจาก "สถาปัตยกรรม" ของข้อมูลองค์กรยังต้องมี "วิศวกรรม" ในการ "ก่อสร้าง" ระบบข้อมูลที่ซับซ้อนนี้ขึ้นมาตามที่ได้วางแผนออกแบบไว้ในสถาปัตยกรรมด้วย แต่หากมีสถาปัตยกรรมการออกแบบที่เหมาะสมแล้ว ก็สามารถสานต่อไปในเชิงวิศวกรรมของการก่อสร้างหรือการจัดทำระบบข้อมูลที่ซับซ้อนนั้นขึ้นมาให้เหมาะสมได้อย่างเป็นระบบ ไม่ทำให้เกิดความวุ่นวายยุ่งยากต่อการนำไปใช้ประโยชน์



เพื่อให้เข้าใจความซับซ้อนของระบบข้อมูลขนาดใหญ่ จะยกข้อมูลระดับมหาศาลของโลกเป็นตัวอย่างสักสองแห่ง ตาม Slide 5 แห่งข้อมูลขนาดใหญ่ของโลกแห่งแรกคือบริษัท Google ซึ่งในเดือนกันยายน พ.ศ. 2550 ประมาณกันว่ามีโฆษณาและการค้นหาผ่านระบบของบริษัทไม่ต่ำกว่า 20 petabyte ต่อวัน ซึ่งเทียบเท่ากับ สองหมื่น terabyte หรือราว 20 ล้าน gigabyte ต่อวันที่เดียว โดยหากเปรียบเทียบจะเห็นว่าเทียบเท่ากับ hard disk ที่มีขายกันในท้องตลาดขนาด 1 terabyte ถึงสองหมื่นลูกที่จะต้องผ่านระบบของบริษัทในแต่ละวันทีเดียว

ส่วนแหล่งข้อมูลขนาดใหญ่อีกแหล่งหนึ่งคือของบริษัท YouTube ซึ่งเป็นเว็บไซต์ที่ให้ผู้ใช้บริการ upload และชมวิดีโอได้อย่างอิสระ โดยในเดือนกรกฎาคม พ.ศ. 2552 ประมาณกันว่ามีวิดีโอกว่า 100 ล้านเรื่องและเพิ่มทวีขึ้นอยู่ตลอดเวลาอย่างรวดเร็ว โดยหากสมมติว่าวิดีโอแต่ละเรื่องมีขนาด 30 megabyte จะเป็นข้อมูลระดับ 3 petabyte หรือในราว 3 ล้าน gigabyte ที่เดียว

สถิติต่างๆ ของ Youtube ใน Slide 6 บ่งว่าในบรรดาวิดีโอที่ 120 ล้านเรื่องนั้น มีการ upload ขึ้นไปยังเว็บไซต์ของบริษัทกว่าวันละสองแสนเรื่อง คิดเป็นปริมาณวิดีโอที่ upload ขึ้นไปในแต่ละนาทีกว่า 13 ชั่วโมง มีการรับชมวิดีโอที่ upload ขึ้นไปในแต่ละนาทีกว่า 200 ล้านเรื่อง มีสมาชิกที่เข้าชมวิดีโอมากกว่า 300 ล้านคน หรือราว 5% ของประชากรของโลก วิดีโอที่นั่นแบ่งเป็นเพลง 20% บันเทิง 15% เป็นวิดีโอส่วนตัว 14% เป็นเรื่องบันเทิงและเบาสมอง 13% เรื่องกีฬา 7% เป็นความรู้ 6% เรื่องเกี่ยวกับรถยนต์ 5% และเป็นภาพยนตร์เพียง 5% ซึ่งจะเห็นได้ว่าปัจจุบันสังคมโลกโดยเฉพาะในกลุ่มที่มีศักยภาพทางระบบสารสนเทศสูงนั้นได้เริ่มต้นเข้าสู่สื่อที่เป็นวิดีโอกันมากขึ้นอย่างเห็นได้ชัด

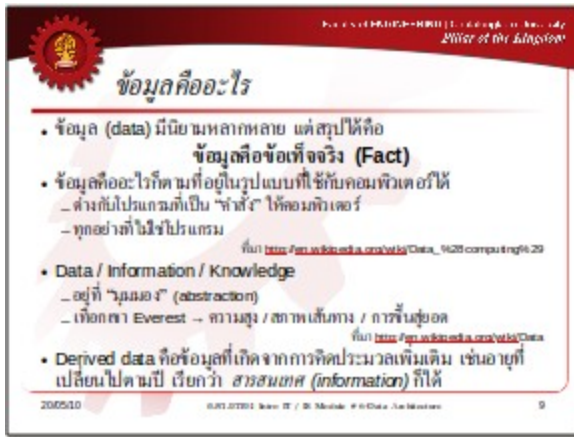


การสร้างระบบข้อมูลที่เหมาะสมขึ้นมาเพื่อรองรับข้อมูลอันมหาศาลและซับซ้อนมากขององค์กรได้นั้น จำต้องอาศัยแบบแผนที่เหมาะสม โดยเฉพาะการมองข้อมูลอย่างเป็นระบบ มีสถาปัตยกรรมของข้อมูล ซึ่งจำต้องคำนึงถึงองค์ประกอบหลายส่วนด้วยกันตาม Slide 7 โดยเริ่มแรกต้องเข้าใจธรรมชาติของข้อมูล ให้รู้ว่าข้อมูลที่มากมายซับซ้อนนั้นเป็นอย่างไร รู้ถึงลักษณะหรือหลักสำคัญเกี่ยวกับข้อมูล รู้ว่าข้อมูลเป็นโครงสร้างพื้นฐานแบบหนึ่ง เข้าใจคุณสมบัติของข้อมูลที่ดี และสุดท้ายจึงกล่าวถึงองค์ประกอบสำคัญของระบบข้อมูลที่ดี

การออกแบบระบบข้อมูลในเชิงโครงสร้างพื้นฐานนั้นจำต้องเข้าใจธรรมชาติของข้อมูลดังใน Slide 8 เป็นอย่างดี ว่าข้อมูลนั้นคืออะไร มาจากไหน มีรูปแบบอย่างไรได้ จะเก็บข้อมูลให้ดีขึ้นได้อย่างไร เก็บที่ใด เมื่อเก็บแล้วสามารถเรียกมาใช้ได้หรือไม่ หากเรียกใช้ไม่ได้ก็ไม่เกิดประโยชน์อันใด ข้อมูลที่หามาเก็บไว้นั้นจะเรียกมาใช้ได้รวดเร็วเพียงใด ข้อมูลมีพร้อมให้ใช้หรือไม่ และใครมีสิทธิ์ใช้ข้อมูลใดได้อย่างไรบ้าง และสุดท้ายคือหากข้อมูลเกิดความเสียหายหรือสูญหายจะกู้คืนได้หรือไม่ เหล่านี้ผู้ที่ออกแบบระบบข้อมูลนั้นต้องมีความเข้าใจพอสมควรจึงสามารถออกแบบระบบให้เหมาะสมได้



**Slide 9** ข้อมูลคืออะไร ในวงการระบบสารสนเทศนั้น นิยามของข้อมูลมีหลากหลาย แต่พอสรุปได้ว่า ข้อมูลคือข้อเท็จจริงคือ fact ซึ่งมีอยู่รอบตัว สามารถรับรู้ได้ทั้งจากอายตนะทั้งห้า หรือจากระบบอุปกรณ์รับหรืออ่านข้อมูลทั้งหลาย เช่นจากบาร์โค้ดที่ข้างกล่องผลิตภัณฑ์ แต่หากพิจารณาในเชิงเทคนิค ข้อมูลนั้นเป็นอะไรก็ได้ที่อยู่ในรูปแบบที่สามารถนำมาใช้กับคอมพิวเตอร์ได้ ซึ่งแตกต่างจากสิ่งที่เรียกว่าโปรแกรมที่เป็น "คำสั่ง" สำหรับสั่งให้คอมพิวเตอร์หรือระบบสารสนเทศทำงานต่างๆ ดังนั้นในเชิงวิชาการอาจมองได้ด้วยซ้ำไปว่าข้อมูลนั้นเป็นอะไรก็ได้ที่ไม่ใช่โปรแกรม และในระบบเองตัวโปรแกรมหรือคำสั่งนั้นสามารถพิจารณาเป็นข้อมูลได้ มีโปรแกรมซึ่งอ่านโปรแกรมอื่นเป็นข้อมูลเข้าไปเพื่อประมวลผลบางอย่างได้ด้วย เช่นการแปลงโปรแกรมที่อยู่ในภาษาสำหรับใช้กับเครื่องหนึ่งให้เป็นภาษาที่จะใช้กับอีกเครื่องหนึ่ง เป็นต้น



นอกเหนือจาก "ข้อมูล" แล้วยังมี "สารสนเทศ" และ "ความรู้" อีกด้วย อันว่า data information และ knowledge สามคำนี้มีความหมายไม่แน่นอน สารสนเทศจากระบบหนึ่งอาจพิจารณาได้เป็นข้อมูลสำหรับป้อนเข้าไปยังอีกระบบหนึ่งก็เป็นได้ ดังนั้นจึงขึ้นกับมุมมองของการใช้งาน เรียกว่าเป็น abstraction เหมือนกับขยะที่เราทิ้งไปอาจเป็นรายได้ของคนเก็บขยะเช่นนี้เป็นต้น ตัวอย่างของข้อมูล สารสนเทศ และความรู้ คือเกี่ยวกับเทือกเขา Everest ที่สูงที่สุดในโลกซึ่งนักไต่เขาชอบไปปีนเพื่อสร้างชื่อเสียงหรือเพื่อความท้าทาย ผู้เรียนอาจมองว่าความสูงของยอดเขาเป็นข้อมูล ส่วนสภาพเส้นทางสภาพอากาศ เป็นสารสนเทศที่ต้องเอาข้อมูลพื้นฐานมา

ประมวลขึ้น ส่วนความรู้คือวิธีจะขึ้นถึงยอดเขาได้ดีที่สุด ซึ่งนักไต่เขาขึ้นยอดที่เคยขึ้น Everest มาแล้วจะมีประสบการณ์มีการศึกษาข้อมูลและสารสนเทศอย่างลึกซึ้ง จนเกิดเป็นความรู้อันหาค่าเปรียบมิได้ไปทีเดียว กลายเป็นผู้ชำนาญหรือผู้เชี่ยวชาญไป

บางครั้งข้อมูลอาจได้มาจากการปรับเปลี่ยนข้อมูลอื่นเพียงเล็กน้อย เช่นนี้เรียกข้อมูลนั้นว่า derived data คือเป็นข้อมูลที่ได้มาจากข้อมูลอื่นโดยไมยาก เช่นหาอายุคนจากการเอาวันที่วันนี้ลบด้วยวันเกิดของผู้นั้น เป็นต้น ซึ่งหากกระบวนการแปลงหรือการประมวลนี้ซับซ้อน อาจเรียก derived data ว่าเป็น information ก็ได้

ข้อมูลมาจากไหนได้บ้าง ใน **Slide 10** นั้นกล่าวถึงแหล่งข้อมูล ซึ่งอาจเป็นข้อมูลที่มีมาจากภายนอกระบบ เช่นอายุของบุคลากร หรือเป็นข้อมูลที่กำหนดขึ้นไว้แล้วในระบบ เช่นปีหนึ่งมี 12 เดือน เป็นต้น ข้อมูลโดยทั่วไปมาจากแป้นพิมพ์ซึ่งยังคงเป็นอุปกรณ์หลักที่ใช้ป้อนข้อมูลโดยมนุษย์ ทั้งในส่วนของการธุรกิจหรือข้อมูลส่วนบุคคลก็ตาม แต่ในปัจจุบันข้อมูลสามารถได้มาจากระบบอุปกรณ์อ่านหรือรับข้อมูลได้โดยตรง เช่นการสแกนด้วยแสงเลเซอร์จากบาร์โค้ดที่เป็นเส้นๆ อยู่ข้างกล่องของผลิตภัณฑ์ ทำให้ร้านค้าสามารถคิดเงินในการซื้อสินค้าของลูกค้าและควบคุมปริมาณสินค้าในระบบสินค้าคงคลังได้ หรือเป็นข้อมูลที่ได้จาก sensor ซึ่งเป็นอุปกรณ์ตรวจจับต่างๆ เช่นวัดอุณหภูมิ วัดปริมาณรถยนต์ในท้องถนน วัดกระแสลมเพื่องานการพยากรณ์อากาศ หรือเป็นอุปกรณ์สมัยใหม่เช่น กล้องดิจิทัลซึ่งเป็นอุปกรณ์รับภาพถ่ายหรือวิดีโอมา



เป็นข้อมูลของระบบ

ประเด็นสำคัญคือข้อมูลนั้นควรมาจากแหล่งกำเนิดของข้อมูลนั้น หากข้อมูลมีการฝากส่งกันมาหลายทอด อาจทำให้เกิดความคลาดเคลื่อนผิดเพี้ยนไปได้เนื่องด้วยข้อจำกัดของช่องทางการส่งผ่านข้อมูลเหล่านั้น ควรคำนึงถึงว่าข้อมูลเป็นของ "ผู้ให้" และ "ผู้ใช้" ข้อมูลเหล่านั้น สิ่งหนึ่งที่เป็นข้อบกพร่องของระบบสารสนเทศคือ ฝ่ายสารสนเทศมักมองว่าข้อมูลเป็นของฝ่าย ซึ่งเป็นแนวคิดที่ไม่ถูกต้อง ฝ่ายระบบสารสนเทศนั้นเป็นเพียงผู้ปกป้องรักษาข้อมูลเท่านั้น หากใช้เจ้าของข้อมูลที่แท้จริงไม่ หากไม่ได้รับความร่วมมืออันดีจากเจ้าของข้อมูล โอกาสที่ข้อมูลจะผิดนั้นจะมีสูง ดังนั้นการที่นักคอมพิวเตอร์ถูกขอให้ไปช่วยที่ฝ่ายบัญชีเพราะข้อมูลที่ได้รับมาไม่ถูกต้องนั้น เป็นสิ่งที่ไม่สมควรทำอย่างยิ่ง เพราะข้อมูลจะถูกต้องหรือไม่ขึ้นอยู่กับฝ่ายบัญชีจะต้องวางแผนหาแนวทางในการหาและนำข้อมูลที่ถูกต้องเข้าสู่ระบบ ทางเทคนิคของการนำเข้าสู่ข้อมูลนั้นฝ่ายระบบสารสนเทศอาจช่วยได้แต่การจะรู้ว่าข้อมูลใดถูกต้อง และรู้ว่าจะตรวจสอบหรือแก้ไขอย่างไรหากข้อมูลไม่ถูกต้องนั้น เป็นหน้าที่โดยตรงของเจ้าของข้อมูลนั้นๆ กล่าวสั้นๆ คือข้อมูลต้องมาจากแหล่งของมันจึงเป็นข้อมูลที่น่าเชื่อถือมีโอกาสน่าเชื่อถือมีคุณภาพดี แต่ถ้าจำเป็นต้องผ่านขั้นตอนหรือเส้นทางที่ยอกระชั้น จำต้องมีกระบวนการตรวจสอบว่าการส่งผ่านข้อมูลเหล่านั้นมีความน่าเชื่อถือเพียงใด ข้อมูลที่ผ่านแต่ละขั้นตอนนั้นมีความถูกต้องสมบูรณ์หรือไม่

ข้อมูลที่ได้มานั้นสามารถอยู่ในรูปแบบที่หลากหลาย **Slide 11** ในสมัยก่อนคอมพิวเตอร์ใช้ในงานด้านวิทยาศาสตร์หรืองานวิจัย ข้อมูลเป็นตัวเลขที่ได้จากการทดลองเป็นหลัก คอมพิวเตอร์เป็นเครื่องคิดเลขที่ซับซ้อน ต่อมาจึงพัฒนาเทคโนโลยีมาใช้กับธุรกิจจึงเกิดข้อมูลที่เป็นตัวอักษรเพื่องานประมวลผลทางพาณิชย์ ใส่ชื่อคนชื่อสินค้าได้ เป็นต้น แต่ในปัจจุบันเทคโนโลยีพัฒนาไปอีกขั้นหนึ่งสามารถรับข้อมูลที่เป็นสื่อประสมเช่นภาพ เสียง เพลง หนังส วิดีโอ ด้วยอุปกรณ์รับข้อมูลที่พัฒนาขึ้น เช่น กล้องดิจิทัล



ในสมัยก่อนข้อมูลเหล่านี้เป็นข้อมูลแบบ analog แต่ปัจจุบันพัฒนาขึ้นมาให้อยู่ในรูปแบบของ digital อันว่านาล็อกนั้นหมายถึงข้อมูลที่เป็นค่าต่อเนื่อง ตัวอย่างเช่นเทอร์โมมิเตอร์ที่ใช้วัดอุณหภูมิหรือที่เรียกกันทั่วไปว่าปรอท จะมีประเภทที่บรรจุอัลกอซอลไว้ภายใน ใส่ลิ้วให้มองเห็นเป็นเส้นและอ่านค่าอุณหภูมิจากการขยายตัวของเส้นอัลกอซอลที่จะดันให้เส้นนั้นสูงหรือต่ำอยู่ในหลอดพลาสติกของปรอทวัดนั้น ถ้าเป็นทางการแพทย์จะใช้สารปรอทจริงๆ มองเห็นเป็นเส้นสีเงินบางๆ จะเห็นว่าค่าที่อ่านได้เป็นค่าประมาณ ต่อเนื่องกันไป ส่วนข้อมูลที่เป็นดิจิทัลจะมีค่าเป็นตัวเลขชัดเจนเช่น

เทอร์โมมิเตอร์สมัยใหม่ที่อ่านอุณหภูมิออกมาเป็นตัวเลขโดยตรง แต่ในความเป็นจริงแล้วค่าที่อ่านได้นั้นเป็นค่าอนาล็อก แต่มีวงจรอิเล็กทรอนิกส์ทำการแปลงค่าอนาล็อกนี้ให้อยู่ในรูปของตัวเลขดิจิทัลเพื่อให้สามารถนำไปประมวลผลในระบบสารสนเทศที่เป็นระบบดิจิทัลโดยพื้นฐานได้สะดวก

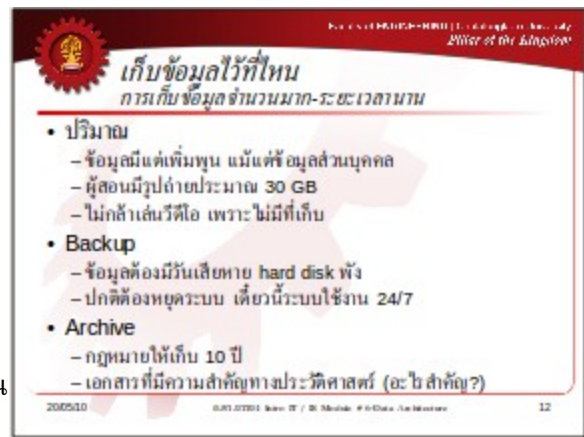
ข้อมูลอาจเป็นเลขจำตัวเต็ม เช่นจำนวนคนไข้ในโรงพยาบาล หรืออาจเป็นทศนิยมเช่น อัตราการเกิดในไทยซึ่งแม้จะเป็นจำนวนคนต่อหน้าที่แต่เป็นค่าเฉลี่ยทั้งประเทศ เช่น 17.9 คนต่อประชากรพันคน เป็นต้น

ประเด็นเกี่ยวกับรูปแบบของข้อมูลที่สำคัญอีกประเด็นหนึ่งคือเรื่องของหน่วย เนื่องจากในคอมพิวเตอร์จะมีการเก็บข้อมูลแยกจากการกำหนดหน่วยของข้อมูลนั้น เช่นดูข้อมูลจากจอภาพเป็น 22 กิโลกรัม ในความเป็นจริงแล้วข้อมูลที่เก็บไว้เป็น 22 ส่วนหน่วยกิโลกรัมนั้นเป็นโปรแกรมที่จัดทำมาใส่เพิ่มเมื่อเวลาแสดงผล ลักษณะเช่นนี้ทำให้องค์กรอวกาศของสหรัฐคือ นาซาต้องสูญเสียยานสำรวจดาวอังคารราคา 125 ล้านเหรียญไปดวงหนึ่งเมื่อพ.ศ. 2542 ราวสิบปีที่ผ่านมา โดยดาวเทียม

ดวงนี้เป็นความร่วมมือกับองค์การอวกาศของยุโรปทำให้เกิดการป้อนข้อมูลที่มีหน่วยเป็นอังกฤษ คือเป็นไมล์ เข้าไปในระบบของยานที่ใช้หน่วยเมตริกที่เป็นสากล คือเป็นกิโลเมตร

ความผิดพลาดเหล่านี้ทำให้เกิดการแพร่หลายของมาตรฐาน XML ซึ่งจะระบุ "ข้อมูลเกี่ยวกับข้อมูล" เช่นหน่วยหรือความหมายของข้อมูล ผนวกไปกับตัวข้อมูลด้วย เช่นแทนที่จะบอกว่า "10" ก็บอกรวมหมดว่า "10 อายุ เดือน" เป็นต้น หากไม่มีการกำกับข้อมูลไว้เป็นอย่างดี ไม่ว่าจะด้วยมาตรฐานอย่าง XML หรือไม่ จะไม่สามารถแยกแยะหรือตีความข้อมูลได้ เช่นข้อมูลวันเดือนปี ซึ่งมีหลากหลายรูปแบบ เช่นระบบของสหรัฐจะขึ้นด้วยเดือนตามด้วยวันที่และปี ส่วนแบบสากลหรือแบบของไทยจะขึ้นด้วยวันที่ตามด้วยเดือนและปี หรือการระบุเป็นพ.ศ.แทนค.ศ. เป็นต้น

เมื่อมีข้อมูลแล้วปัญหาที่ตามมาคือจะเก็บไว้ที่ไหน อย่างไร นานแค่ไหน **Slide 12** ข้อมูลในปัจจุบันนับวันมีแต่จะกองงยเพิ่มพูนขึ้นเรื่อยๆ ไม่เพียงแต่ข้อมูลขององค์กรหรือข้อมูลทางธุรกิจเท่านั้น แม้แต่ข้อมูลส่วนบุคคลเองยังมีแนวโน้มที่จะขยายตัวอย่างรวดเร็วเนื่องจากข้อมูลสื่อประสมที่สร้างขึ้นได้ง่ายๆ อย่างผู้สอนเองมีข้อมูลของภาพที่ถ่ายจากกล้องดิจิทัลของครอบครัวอยู่ในราว 30 กิกกไบต์ ทำให้ยากลำบากต่อการดูแลรักษาให้มีความคงทนถาวร ปลอดภัยจากการที่ฮาร์ดดิสก์จะเกิดพังเสียหายเป็นอย่างมาก ต้องสำเนาไว้หลายชุด ต้องซื้อฮาร์ดดิสก์เพิ่มเติมทุกปี และทำดัชนีรายการว่าชุดไหนอยู่ที่ใด เพื่อกันการสับสน ทำให้ผู้สอนไม่กล้าจะเล่นกล้องถ่ายวิดีโอ เพราะเกรงจะหาที่เก็บไม่ได้



สังกรรมประการหนึ่งของการเก็บข้อมูลคืออุปกรณ์ที่ใช้เก็บนั้นมีอายุ อย่างแผ่นซีดีหรือวีดีโอดูภาพยนต์นั้น เมื่อวานขณะเตรียมเอกสารการสอนนี้อยู่ เปิดดูแผ่นเก่าๆ พบว่าสารที่เคลือบไว้เพื่อเก็บข้อมูลเริ่มเสื่อม พลิกดูเห็นเป็นดวงๆ ไม่สามารถใช้งานได้อีกต่อไป ส่วนฮาร์ดดิสก์นั้นก็มีการใช้งานเช่นกัน ขึ้นกับคุณภาพของการผลิตแต่สำหรับอุปกรณ์ระดับ

ชาวบ้านหรือผู้บริโภคทั่วไปนั้นหากใช้งานปกติก็อยู่ในราว 5 ถึง 10 ปีเท่านั้น ดังนั้นจะเก็บรูปภาพลูกๆ ตอนเด็กเอาไว้ทำวิดีโอฉายวันแต่งงานนั้น อยู่ไม่ถึงหรือครึ่ง ต้องมีการทำสำเนาข้อมูลเป็น backup หรือสำรองเอาไว้เป็นระยะ ในทางธุรกิจก็เช่นกัน อาจสำเนาไว้ทุกวันหรือทุกเดือน

ในสมัยก่อนการสำเนาข้อมูลต้องปิดระบบ แต่ในปัจจุบันระบบสารสนเทศมีการให้บริการแบบต่อเนื่องไม่มีหยุดเช่นเครื่อง ATM ในลักษณะของ 24/7 คือตลอดยี่สิบสี่ชั่วโมงทุกวันตลอดเจ็ดวันในสัปดาห์โดยไม่มีวันหยุด ดังนั้นระบบสมัยใหม่จะมีการสำเนาข้อมูลได้ในขณะที่ระบบยังทำงานอยู่

นอกจากข้อมูลจะมีปริมาณที่เพิ่มพูนขึ้นอย่างรวดเร็ว ยังมีแนวโน้มว่าจะถูกเก็บอย่างถาวรไม่เหมือนในสมัยก่อนที่หน่วยบันทึกข้อมูลมีราคาแพงมีขนาดจำกัด ส่วนมากจะลบข้อมูลที่ไมใช้แล้วทิ้งไปเช่นข้อมูลผู้ป่วยย้อนหลังเกินห้าปีเพราะถือว่าอยู่ในกระดาษในแฟ้มเอกสารปกติอยู่แล้ว แต่ในปัจจุบันการดำเนินกิจกรรมต่างๆ อาศัยข้อมูลจากระบบสารสนเทศอิเล็กทรอนิกส์มากขึ้นเรื่อยๆ ทำให้พึ่งพาความถาวรของข้อมูลที่อยู่ในกระดาษไม่ได้อีกต่อไป

การเก็บข้อมูลที่นับวันมีแต่จะเพิ่มพูนขึ้นไว้เป็นระยะยาวเช่นนี้กำลังเริ่มจำเป็นสำหรับธุรกิจและองค์กรโดยทั่วไป นอกจากทำตามกฎหมายที่ต้องเก็บข้อมูลเพื่อใช้ในทางศาลเป็นเวลา 10 ปีแล้ว ยังมีเรื่องของการนำข้อมูลเก่าย้อนหลังไปนานๆ เพื่อการวิเคราะห์หาแนวโน้มต่างๆ เช่นการเพิ่มลดของประชากรในกลุ่มอายุต่างๆ เพื่อการวางแผนระดับชาติเรื่องแรงงาน อุตสาหกรรมหรือการประกันสุขภาพ รวมไปถึงการนำข้อมูลเก่ามาใช้ในการคาดเดาถึงพฤติกรรมของลูกค้าเพื่อผลทางการค้าด้วย เหล่านี้เป็นต้น

นอกจากนั้นการเก็บข้อมูลบางประเภทยังเกี่ยวเนื่องกับความสำคัญของข้อมูลนั้นในเชิงประวัติศาสตร์ ในลักษณะนี้จะเป็นการเก็บข้อมูลแบบถาวร หรือที่เรียกว่า archive คือไม่มีการโยนทิ้งหรือลบทิ้ง ตัวอย่างเช่นข้อมูลสำมะโนประชากร ซึ่งสามารถใช้ติดตามร่องรอยของบุคคลสำคัญในประวัติศาสตร์ได้ เป็นต้น ปัญหาที่จะทราบได้อย่างไรว่าข้อมูลใดมีความสำคัญทางประวัติศาสตร์ ภาพถ่ายทั่วไปของครอบครัวหนึ่งอาจจับภาพสำคัญทางการเมืองหรือทางประวัติศาสตร์ไว้ได้ก็เป็นได้และจะมีค่าสูงเมื่อเวลาผ่านไปยาวนาน เช่นเด็กเลี้ยงควายที่ถ่ายไว้เมื่อสามสิบปีที่แล้วกลายเป็นนายกรัฐมนตรีไปในขณะนี้เป็นต้น

การเก็บข้อมูลที่ดีต้องมีลักษณะที่เหมาะสมต่อการใช้งาน **Slide 13** เช่นถ้าเป็นข้อมูลที่มีปริมาณไม่มาก อาจเก็บไว้เป็นแฟ้มข้อมูลธรรมดาที่ไม่มีโครงสร้างพิเศษ แฟ้มเหล่านี้อาจเก็บข้อมูลที่เป็นชิ้นๆ ชัดเจน เช่น รูปภาพหนึ่งรูปต่อหนึ่งแฟ้ม หรือภาพยนตร์สักเรื่อง แต่ทั้งนี้แฟ้มข้อมูลเหล่านี้ต้องนำไปประกอบเป็นระบบข้อมูลที่ซับซ้อนต่อไปด้วย



เมื่อข้อมูลมีจำนวนมากขึ้น การใส่รวมไว้ในแฟ้มเฉยๆ จึงไม่สะดวกต่อการค้นหา จึงต้องใช้แฟ้มที่มีโครงสร้างพิเศษ มีดัชนีหรือ index สำหรับช่วยในการสืบค้นข้อมูลขึ้นที่ตรงการในแฟ้ม เหมาะสำหรับการประมวลผลแบบโต้ตอบคือให้บริการลูกค้าเป็นรายๆ ไปได้ แต่เมื่อข้อมูลมีปริมาณมากขึ้น มีความหลากหลายของการใช้งาน มีระบบงานสารสนเทศมากมายเพื่อการทำกิจกรรมขององค์กร ข้อมูลเหล่านี้จึงถูกบรรจุไว้ในซอฟต์แวร์พิเศษกลายเป็นระบบฐานข้อมูลซึ่งมีระบบดัชนีที่ซับซ้อน มีซอฟต์แวร์สนับสนุนการสืบค้นอย่างพลิกแพลงและหลากหลายได้

เพื่อเป็นตัวอย่างเกี่ยวกับประโยชน์ของการเก็บข้อมูลไว้เป็นระยะเวลานาน จึงให้ผู้เรียนได้เห็นตัวอย่างความต้องการสารสนเทศที่อาจเป็นไปได้ดังใน **Slide 14** ตัวอย่างแรกคือกระทรวงสาธารณสุขต้องการติดตามการแพร่กระจายของโรคไข้หวัดนกย้อนหลังไปนานๆ นับยี่สิบปี เช่นนี้หากข้อมูลการเจ็บป่วยล้มตายจากโรคซึ่งรวมทั้งคนและสัตว์ หากอยู่ในรูปของเอกสารกระดาษ การสนองตอบความต้องการทางสารสนเทศนี้คงใช้เวลาหลายเดือนเพราะต้องไปรวบรวมข้อมูลจากหน่วยงานต่างๆ มาบรรจุลงในระบบ เขียนโปรแกรมเพื่อประมวลผลออกมา ลักษณะการเช่นนี้เป็นกับความต้องการในตัวอย่างที่สองเช่นกัน คือหากมหาวิทยาลัยต้องการศึกษาถึงพัฒนาการการเปลี่ยนแปลงของหลักสูตรต่างๆ ที่มีอยู่ในคณะต่างๆ ของมหาวิทยาลัยเพื่อการวางแผนและการพัฒนาหลักสูตรของมหาวิทยาลัย เช่นนี้ก็ต้องการข้อมูลซึ่งมักอยู่ในรูปของเอกสารกระดาษจากที่ต่างๆ กระจัดกระจายมากมาย ทั้งสองตัวอย่างนี้จะเห็นได้ว่าเอกสารกระดาษที่เกี่ยวข้องนั้นจะกระจายอยู่ตามหน่วยงานที่เป็นเอกเทศเป็นอิสระจากกันพอสมควร หากเป็นคำสั่งหรือความต้องการจากหน่วยเหนือ เช่นจากรัฐมนตรี ปลัดกระทรวง หรืออธิการบดี ก็พอจะเป็นไปได้แม้จะใช้เวลาบางอย่างที่กล่าวแล้ว แต่หากเป็นหน่วยวิจัยการแพร่ของโรคในกองใดกองหนึ่งของกระทรวง หรือฝ่ายวิชาการของคณะใดคณะ





หนึ่งเพื่อการพัฒนาหลักสูตรของคณะ เช่นนี้โอกาสสนองความต้องการนี้จะเป็นไปแทบไม่ได้เลย เพราะต้องขอความร่วมมือจากหน่วยงานในระดับเดียวกันหรือระดับสูงกว่า ไม่สามารถทำได้โดยง่าย

ตัวอย่างที่สามก็เช่นกัน เป็นการวิเคราะห์ถึงการย้ายถิ่นของประชากรไทยย้อนหลังไปยี่สิบปี เช่นนี้เป็นความต้องการที่ไม่มีผลต่อการบริหารจัดการบ้านเมืองหรือผลกำไรทางธุรกิจแต่อย่างใด แต่เป็นการศึกษาในเชิงประวัติศาสตร์ที่จะมีผลในระยะยาวของการพัฒนาเมืองใหญ่และการวางรากฐานทางเศรษฐกิจและสังคมของประเทศ เช่นนี้ระบบสารสนเทศแทบจะช่วยเหลืออะไรไม่ได้เลย

สำหรับตัวอย่างสุดท้ายนั้นต่อจะเป็นความจริง สามารถทำได้ขึ้นสักหน่อยเมื่อเทียบกับสามตัวอย่างข้างต้น หากมีการฟ้องร้องเกี่ยวเนื่องกับธุรกรรมของบริษัท ก็สามารถนำข้อมูลในรูปของเอกสารกระดาษ ค้นหาขึ้นมาเพื่อประโยชน์ทางศาลได้ แต่หากในอนาคตมีการผลักดันให้สามารถใช้เอกสารอิเล็กทรอนิกส์แทนเอกสารกระดาษอย่างเป็นทางการเป็นรูปธรรม เช่นนี้ก็จำเป็นที่บริษัทจะต้องหาวิธีเก็บรักษาข้อมูลจำนวนมากไว้เป็นเวลาอย่างน้อยยี่สิบปีตามความต้องการทางกฎหมายอย่างหลีกเลี่ยงไม่ได้ แต่การเก็บในลักษณะนี้เป็นระบบที่ชัดเจนอยู่แล้วด้วยกระดาษดังนั้นการแปลงให้อยู่ในรูปของการเก็บระยะยาวโดยใช้อิเล็กทรอนิกส์จึงยังเป็นที่สงสัยหากมีการวางแผน ออกแบบ และจัดสร้างอย่างเหมาะสม

คราวนี้เมื่อมีข้อมูลแล้ว เก็บรักษาไว้ดีแล้วไม่ว่าระยะสั้นระยะยาว คำถามคือสามารถเรียกมาใช้ได้หรือไม่ ดีแค่ไหน

**Slide 15** เหมือนดูหนังสือสอบอย่างหนักแต่แล้วเมื่อเข้าห้องสอบก็ทำอะไรไม่ได้ทำสอบไม่ได้ พอออกจากห้องสอบก็กลับจำคำตอบได้จนแทบจะเสียดิไป เช่นนี้เรียกว่าไม่สามารถค้นคืนข้อมูลที่มีอยู่ออกมาใช้ได้ นักเรียนนักศึกษาจึงมักมีระบบช่วยจำหลากหลายเช่น ไก้จิกเด็กตายบนปากอ่าง ซึ่งหมายถึงอักษรกลาง ก จ ด ต ถ ฎ ฏ บ ป อ แก้วตัวนั่นเอง

ซึ่งการจะเรียกใช้ข้อมูลได้อย่างเป็นผลนี้ต้องมีการเตรียมการ การออกแบบระบบข้อมูลไว้อย่างเหมาะสมก่อนนั่นเอง อย่างตัวอย่างอักษรกลางนั้นก็ต้องมีการเตรียมการ การหาคำพ้องเพื่อให้จำได้ง่าย การท่องจำคำกลอนวรรคนั้นไว้อย่างเป็น



ระบบ เพราะไม่มีเพียงอักษรกลางเท่านั้น ไม่ใช่เอาข้อมูลมาสุ่มๆ กันแล้วเวลาจะใช้จะค้นหาวิธีที่เหมาะสมซึ่งอาจจะไม่สามารถทำได้เพราะข้อมูลไม่อยู่ในรูปแบบที่เหมาะสมต่อการสร้างดัชนีสืบค้น

ข้อมูลที่มีนั้นจะเป็นประโยชน์ต้องไม่เพียงสามารถนำมาใช้ได้ แต่ต้องแน่ใจว่าได้มาโดยเร็ว ทันใช้งาน และการสืบค้นข้อมูลมาใช้นั้นต้องครบ ไม่ได้มาเพียงส่วนเดียวหรือส่วนน้อย และสุดท้ายคือข้อมูลที่ค้นมาได้เพื่อการใช้งานนั้นต้องตรงกับความต้องการการใช้งานด้วย ไม่ใช่สักแต่ไปหยิบไปคว้าอะไรที่ไม่เกี่ยวข้องมาให้ อย่างบางครั้งก็เกิดกับผลจากการสืบค้นของ

Google

เหล่านี้อาจเรียกได้ว่าปัจจุบันเกิดเป็นศาสตร์ใหม่ที่เริ่มเป็นที่นิยมอย่างมากคือศาสตร์ด้านการค้นคืนหรือการสืบค้นสารสนเทศ หรือ Information Retrieval ที่กลายเป็นศาสตร์ที่มีความซับซ้อน กระตุ้นให้เกิดความนิยมอย่างมากด้วยความสำเร็จอย่างสูงของบริษัท Google ยักษ์ใหญ่แห่งการสืบค้น

สูงขึ้นไปจากการสืบค้นคือการใช้ประโยชน์ข้อมูลที่มีอย่างพิสดาร ซับซ้อน เป็นประโยชน์ต่อกิจการอย่างสูง เช่นการทำเหมืองข้อมูล หรือ Data Mining ซึ่งเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างข้อมูลที่มีอยู่จำนวนมากขององค์กรเพื่อค้นพบประเด็นใหม่ๆ ที่ไม่คาดคิด เช่นอาจพบว่าการติดเชื้อของโรคคว่ำบา เกิดจากวัวที่ไม่ได้เกิดจากการผสมเทียม ซึ่งเกินความคาดหมายของนักวิจัย เช่นนี้เป็นต้น

คำถามอีกข้อหนึ่งที่สำคัญเกี่ยวกับธรรมชาติของข้อมูลซึ่งต้องคำนึงถึงในการออกแบบระบบข้อมูลที่ดี คือข้อมูลที่ต้องการใช้นั้นต้องการเร็วแค่ไหน ข้อมูลของเมื่อวานนี้เร็วพอไหม หรือต้องเป็นของภายในชั่วโมงนี้ หรือภายในนาทีนี้ หรือวินาทีนี้เลยหรือไม่ ซึ่งเหล่านี้มีผลกระทบต่อการใช้งานข้อมูล

**Slide 16** เรื่องความทันสมัยของข้อมูลที่ต้องการใช้นั้นมีผลอย่างสูงต่อระบบข้อมูล โดยเฉพาะในส่วนของความเร็วที่ข้อมูล หากเป็นข้อมูลสรุปประจำเดือน สักวันที่ 20 ค่อยไปตามก็ยังไม่ดี แต่หากต้องการข้อมูลที่ทันการตลอดทุกนาที ก็ต้องมีวิธีการตามเข้าไปเก็บข้อมูลตรงจุดที่เกิดข้อมูลโดยตรง เช่นข้อมูลจรรยา หากเป็นรายงานข่าวที่ช่างกล้องไปจับมาได้แล้วส่งมอเตอร์ไซค์กลับมาที่ห้องส่งก็คงเหมาะกับการรายงานสรุปการจรรยาประจำวัน แต่หากต้องการให้ผู้บริโภคใช้ประโยชน์สารสนเทศขณะขับรถอยู่ก็ต้องเอากล้องไปติดไว้ตามถนนตามสี่แยกต่างๆ ด้วยจึงได้ข้อมูลที่ทันต่อเหตุการณ์สมัยก่อนเทคโนโลยียังไม่ดีนัก อุปกรณ์ไม่ซับซ้อนและมีราคาสูง ดังนั้นข้อมูลจึงต้องเดินทางทางกายภาพ เช่นมาเป็นเอกสารกระดาษ มายังเครื่องคอมพิวเตอร์ มีเจ้าหน้าที่คอยป้อนข้อมูลเหล่านั้นเก็บสะสมไว้เป็นชุด เป็นกะ เรียกว่า batch ซึ่งเมื่อมีจำนวนมากเพียงพอจึงค่อยใส่เข้าไปประมวลผลในระบบ ในลักษณะนี้ผลลัพธ์หรือรายงานที่ได้จะเป็นสารสนเทศแบบรายสัปดาห์ เช่นรายวัน รายเดือน หรือรายปี ตัวอย่างเช่นใบเรียกเก็บเงินค่าโทรศัพท์ประจำเดือนหรือใบแจ้งดอกเบี้ยเงินฝากทุก 6 เดือน เป็นต้น



ต่อมาเมื่อเทคโนโลยีพัฒนาขึ้น อุปกรณ์ถูกลง แพร่หลายมากขึ้น ทำให้สามารถนำจอภาพและแป้นพิมพ์ไปวางไว้หน้าเจ้าหน้าที่ พนักงาน และลูกค้า ให้เขาเหล่านั้นป้อนข้อมูลที่จำเป็นและเกี่ยวข้องเข้าสู่ระบบได้โดยตรง สามารถประมวลผลและได้รายงานผลลัพธ์ออกมาทันต่อเหตุการณ์ รวมถึงอุปกรณ์สมัยใหม่สามารถเชื่อมต่อกัน เช่นระบบชุมสายโทรศัพท์สามารถเชื่อมโยงส่งข้อมูลการใช้โทรศัพท์แต่ละเครื่องตรงเข้าสู่ระบบฐานข้อมูลลูกค้าได้ เช่นนี้ลูกค้าสามารถเข้ามาที่เว็บไซต์ของบริษัทโทรศัพท์เพื่อเรียกดูการใช้งานโทรศัพท์ของเขาได้ทันทีหลังจากใช้โทรศัพท์เสร็จ ตัวอย่างที่เห็นได้ชัดอีกอย่างคือการถอนเงินจากตู้อัตโนมัติ ATM ซึ่งระบบต้องรู้ทันทีว่ามียอดเงินเหลือเท่าไร ถอนไปแล้วเท่าไรในวันนั้นและเหลือให้ถอนได้เท่าไร ลักษณะของข้อมูลแบบ interactive นี้ทำให้มีระบบการใช้โทรศัพท์แบบ prepaid หรือเติมเงินได้ เพราะระบบสารสนเทศมีข้อมูลที่ทันต่อเหตุการณ์ สามารถคิดหักเงินที่เติมไว้ไปตลอดเวลาที่ใช้โทรศัพท์ได้ทันที แต่ระบบสารสนเทศบางประเภทก็ต้องการความรวดเร็วในการได้มาซึ่งข้อมูลอย่างสูง เช่นการควบคุมเครื่องสูบสารเคมีกำลังสูงที่ต้องได้ข้อมูลระดับของสารเคมีในถังเก็บอย่างทันทีเพื่อหยุดเครื่องสูบหากสารเคมีเต็ม

ธรรมชาติของข้อมูลที่สำคัญอีกประการหนึ่งซึ่งมักมองข้ามไปคือเกี่ยวกับความมั่นคงปลอดภัยของข้อมูล **Slide 17** ใครจะใช้ข้อมูลได้บ้างนั้น อย่างแรกก็ต้องมีการกำหนดสิทธิ์การใช้งานหรือ usage authorization อย่างเป็นทางการ ไม่เช่นนั้นจะเป็นกรณีที่เจ้าหน้าที่แก้ไขข้อมูลพรรคการเมืองที่เป็นข่าวไปเมื่อไม่กี่ปีมานี้ หรือเจ้าของข้อมูลเข้าไปแก้อายุตัวเองเพื่อให้เกษียณอายุช้าลง เช่นนี้อาจเกิดขึ้นได้หากไม่มีการระบุสิทธิ์และควบคุมดูแลการใช้สิทธิ์ อย่างในตัวอย่างนี้แม้จะเป็นเจ้าของข้อมูลแต่ไม่อาจพลการแก้ไขข้อมูลของตนเอง

ความมั่นคงปลอดภัยของข้อมูลก็เป็นสิ่งสำคัญ หากไม่มีระบบและการออกแบบที่ดีย่อมเปิดโอกาสให้ผู้ไม่ประสงค์ดีเข้ามาขโมยข้อมูลหรือทำลายเปลี่ยนแปลงให้เกิดความเสียหายได้ เช่นการแก้ไขข้อมูลการเงินในบัญชีธนาคาร ซึ่งในกรณีนี้ไม่อาจ

มั่นใจเต็มที่ได้ในกำบัง จำต้องมีวิธีตรวจสอบถึงการเปลี่ยนแปลงแก้ไขโดยไม่มีสิทธิ์อำนาจเหล่านี้ได้



นอกจากนั้นเมื่อข้อมูลส่วนบุคคลของลูกค้าและผู้คนเข้าไปอยู่ในระบบสารสนเทศมากขึ้นเรื่อยๆ ย่อมเกิดปัญหาความเป็นส่วนตัวของข้อมูลเหล่านั้น ตัวอย่างเช่นการที่ผู้เรียนได้รับจดหมายโฆษณาขายของทางอีเมลเหล่านั้นเนื่องมาจากบริษัทที่เก็บข้อมูลส่วนตัวที่ผู้เรียนไปเป็นสมาชิกได้นำข้อมูลเหล่านั้นไปขายหรือแจกจ่ายหรือไม่ปกป้องให้ดีพอ เช่นนี้จึงเป็นการละเมิดสิทธิ์ความเป็นส่วนตัวของผู้บริโภค

การปกป้องข้อมูลเหล่านี้ไม่อาจทำในภายหลังได้โดยง่าย จำต้องมีการคำนึงและออกแบบ จัดสร้างเพื่อไว้ตั้งแต่ต้นจึงจะได้ระบบข้อมูลที่มีความน่าเชื่อถือได้

ประเด็นสุดท้ายของธรรมชาติของข้อมูลคือหากข้อมูลเกิดความเสียหายสามารถกู้คืนได้หรือไม่ **Slide 18** ผู้เรียนลองนึกถึงสภาพการณ์ที่ผู้เรียนตื่นมาไปทำงานในตอนเช้าแล้วพบว่าที่ทำงานถูกม็อบวางเพลิงวอดวายไปหมดเมื่อคืน รวมไปถึงห้องเซิร์ฟเวอร์และระบบสารสนเทศทั้งหมดขององค์กร

สำหรับเครื่องเซิร์ฟเวอร์ ซอฟต์แวร์ต่างๆ โปรแกรมและระบบเครือข่าย หากมีงบประมาณ มีเงินทุนก็สามารถจัดซื้อจัดหาทดแทนได้ แต่ข้อมูลขององค์กรนั้นเป็นสมบัติที่มีค่ามหาศาล หากเสียหายไปจะไม่สามารถกู้คืนได้โดยง่าย อาจต้องใช้เวลานานในการคอยติดตามคืนจากลูกค้า จากหลักฐานเอกสารกระดาษต่างๆ ซึ่งอาจถึงกับไม่สามารถทำได้ เช่นนี้จะเห็นว่าหากข้อมูลเกิดความเสียหายจะทดแทนได้ยากมาก ดังนั้นต้องออกแบบระบบข้อมูลเพื่อไว้สำหรับสภาพการณ์ฉุกเฉินยิ่งยวดเหล่านี้ด้วย



ในศาสตร์ด้านระบบสารสนเทศนี้เรียกว่าเป็นหลักการของความต่อเนื่องไม่ล่มสลายของธุรกิจ หรือ Business Continuity คือธุรกิจต้องสามารถดำเนินต่อไปได้แม้เมื่อเกิดความเสียหายต่อข้อมูล

เมื่อเข้าใจธรรมชาติของข้อมูลแล้ว ควรเรียนรู้เกี่ยวกับสังกรรมของข้อมูล ซึ่งถึงได้ว่าเป็นแก่นความรู้ที่ต้องใช้กำกับสติในการวางแผน การออกแบบ และการจัดสร้างระบบข้อมูล รวมไปถึงการปฏิบัติงานในการใช้ข้อมูล และการบำรุงรักษาข้อมูลให้อยู่ในสภาพพร้อมใช้งานได้เสมอ

สังกรรมพื้นฐานของข้อมูลใน **Slide 19** คล้ายกับของศาสนาพุทธที่ใช้ยึดเหนี่ยวจิตใจ นั่นคือกฎแห่งกรรม หมายถึงทำสิ่งใดไว้ย่อมได้ผลนั้นกลับคืน ดังนั้นหลักของข้อมูลคือ "ป้อนขยะ ได้ขยะ" หรือ Garbage In – Garbage Out เมื่อปล่อยให้ข้อมูลที่ไม่ถูกต้องเข้าไปสู่ระบบ หรือ "ป้อนขยะ" ก็จะได้ผลลัพธ์ที่ไม่ถูกต้องคืนมา หรือ "ได้ขยะ" นั่นเอง ตัวอย่างเช่นหากข้อมูลการสมัครสอบเข้าอุดมศึกษาหรือที่เรียกว่าระบบ Admission ไม่ถูกต้อง ผลที่ได้คือนักเรียนจะเสียสิทธิ์ในการเข้าสถาบันที่ตนต้องการ เช่นนี้เป็นต้น

จะพบว่าข้อมูล Admission นั้นสามารถมีจุดที่จะผิดพลาดได้มากมายตลอดเส้นทางของข้อมูล เช่นผิดตั้งแต่รหัสเลขประจำตัวประชาชน 13 หลัก เลขนี้ถูกกำหนดให้ตั้งแต่เด็กถูกแจ้งเกิดแล้ว เพียงแต่ในขณะนั้นยังไม่มียบัตรประชาชน แต่ถ้าดูใน

ทะเบียนบ้านจะพบเลข 13 หลักนี้ ดังนั้นผู้ปกครองมักสับสนเมื่อนำบุตรเข้าเรียนอนุบาลหรือประถม เพราะต้องกรอกเลข 13 หลักนี้จึงผลการออกของตัวผู้ปกครองเองลงไป ดังนั้นเมื่อต้องการใช้สอบเข้ามหาวิทยาลัย ข้อมูลที่ผิดพลาดที่แฝงตัวอยู่ในระบบนี้จึงสำแดงฤทธิ์เดชออกมาทำให้นักเรียนพลาดโอกาสในการเข้าศึกษาต่อได้เพราะข้อมูลไม่ถูกต้อง ดังนั้นระบบสารสนเทศอย่าง Admission จึงต้องมีกระบวนการวิธีและขั้นตอนที่จะขัดเกลาตรวจสอบให้แน่ใจว่าข้อมูลที่เข้ามาสู่ระบบและที่อยู่ในระบบในทุกขั้นตอนนั้นไม่มีการผิดพลาดหรือผิดเพี้ยนไปเป็นอันขาด ซึ่งเป็นส่วนที่ยู่งยากที่สุดของระบบงาน Admission ของไทยทีเดียว

นอกจากเลขรหัส 13 หลักซึ่งไม่น่าจะผิดแล้ว ชื่อและสกุลของนักเรียนก็ผิดพลาดได้ นับแต่สะกดผิด อาจใส่คำนำหน้าเข้าไป



เช่น "นายขวัญ" ซึ่งอาจหมายถึง ดช. นายขวัญ หรือเป็น นายขวัญ รวมถึงยุคนี้มีการเปลี่ยนชื่อและนามสกุลกันด้วย วัตถุประสงค์ต่างๆ กัน ชื่อและรหัสโรงเรียนก็เป็นข้อมูลที่ผิดพลาดมากเช่นกัน นักเรียนนั้นเมื่อเรียนในโรงเรียนมาหกปีหรือสิบสองปี ชื่อโรงเรียนนั้นก็ซึมซาบอยู่แล้วในสายเลือด รู้ดีว่า "นี่โรงเรียนนั้น" หลับตาเดินไปโรงเรียนยังได้ แต่กลับสะกดชื่อโรงเรียนไม่ถูกหรือเรียกชื่อไม่ครบ รหัสประจำโรงเรียนยิ่งไม่ต้องพูดถึง แม้มีรายการเทียบให้ก็ยังคัดลอกมากรอกผิด เป็นต้น หรือคะแนนสอบ ทั้งที่เป็นผลการเรียนจากมัธยม หรือเป็นคะแนนที่ได้จากการตรวจข้อสอบวัดความรู้ต่างๆ มีจุดที่อาจผิดพลาดได้

มากมาย ดังนั้นระบบข้อมูลต้องมีสถาปัตยกรรม มีภาพรวม มีการวางแผนและการออกแบบที่เหมาะสม โดยปกติโรงเรียนทุกแห่งจะสอนนักเรียนให้หัดเอาเลขสองตัวมาบวกกัน แต่ไม่มีหลักสูตรใดที่สอนให้สงสัยว่าเลขสองตัวที่เป็นโจทย์นั้นมาได้อย่างไร ถูกต้องหรือไม่ นั่นคือการศึกษการสอนให้เรารู้ว่าข้อมูลที่ได้รับฟังมานั้นถูกต้องเสมอ จะเป็นด้วยว่าผู้ใหญ่บอกมา มาจากหน่วยงานที่น่าเชื่อถือระดับชาติ หรือฝรั่งให้มา เช่นนี้เป็นแนวทางสถาปัตยกรรมข้อมูลที่ผิดในขั้นพื้นฐานทีเดียว

ดังที่ได้กล่าวมาทั้งหมดจะเห็นว่าข้อมูลนั้นมีความสำคัญยิ่งยวดสำหรับระบบสารสนเทศ โดยทั่วไปจะมองข้ามกันในขณะนี้ออกแบบและจัดทำระบบสารสนเทศ แต่ข้อมูลนั้น **Slide 20** ถือได้ว่าเป็นโครงสร้างพื้นฐาน คือ infrastructure ได้แบบหนึ่ง เปรียบเสมือนเซิร์ฟเวอร์ เครือข่าย ระบบปรับอากาศ ตู้อุปกรณ และทั้งยังมีความสำคัญมากกว่าด้วยเพราะจะปล่อยให้เสียหายไม่ได้เป็นอันขาด ในสมัยก่อนนิยมมองว่าซื้อเครื่องเขียนโปรแกรม แล้วค่อยหาข้อมูลมาบรรจุ แต่ในปัจจุบันนั้นข้อมูลที่ซื้อมาอาจต้องใช้การพิจารณา การเตรียมการ การหาข้อมูล การป้อนข้อมูลเข้า เตรียมไว้ตั้งแต่ก่อนที่จะเริ่มซื้อเครื่องหรือเขียนโปรแกรมเสียอีก โครงการสร้างระบบสารสนเทศส่วนมากนั้นไม่สำเร็จเพราะการเขียนโปรแกรมได้ไม่ทันการ แต่ที่กระต่อนกระแต่นั้นเพราะมีระบบข้อมูลที่ไม่ดีพอนั่นเอง



ยกตัวอย่างบริการสืบค้นข้อมูลของบริษัท Google จะเห็นว่าหากเป็นเมื่อสิบหรือยี่สิบปีก่อน บริการนี้ไม่มีวันจะได้รับความนิยมเลย เพราะในสมัยนั้นอินเทอร์เน็ตและเว็บยังไม่แพร่หลาย ดังนั้นก็ไม่มีข้อมูลให้ Google ไปรวบรวมมาหรือเรียกว่า

crawl คือคืบคลานไปตามเว็บไซต์ต่างๆ แล้วเอามาทำดัชนีหรือ index เอาไว้ให้สืบค้นได้ ดังนั้นอาจมองได้ว่าเมื่ออินเทอร์เน็ตแพร่กระจายไปมากขึ้น ข้อมูลหลากหลายมหาศาลที่ผู้คนชนป้อนเข้าไปนั้นกลายเป็นโครงสร้างพื้นฐานอันสำคัญที่ทำให้บริษัทอย่าง Google สามารถรุ่มกิจการของตนลงไปแล้วตัดวงเอาของที่เป็นประโยชน์ขึ้นมาค้าขายได้ หากพิจารณาให้ดี เหมือนผู้คนที่ต่างก็เป็นเจ้าของข้อมูลที่ Google สืบค้นมาได้ และเราต่างก็เป็นลูกค้าที่มาบริโภคข้อมูลของเรากันเองโดยที่มี Google เป็น "นายหน้า" ในการแลกเปลี่ยนข้อมูลผ่านระบบสืบค้นของบริษัทนั่นเอง

ดังนั้นหากเปรียบเทียบข้อมูลของโลกที่ระดมกันใส่เข้าไปในอินเทอร์เน็ตเป็นเหมือนทะเลข้อมูลที่เป็นรากฐานหรือโครงสร้างพื้นฐานให้ Google ตักตวงไปใช้ ก็เหมือนกับน้ำประปาที่เป็นโครงสร้างพื้นฐานให้ร้านค้าแปดต่อแปดไปเปิดน้ำชงกาแฟขายนั่นเอง

ดังนั้นการที่จะมีโครงสร้างพื้นฐานทางข้อมูลที่ดี จำต้องมีข้อมูลที่ดีด้วย ข้อมูลจะดีได้นั้นมีคุณสมบัติสำคัญสามประการด้วยกัน ตาม **Slide 21** คือ "ถูกต้อง" "ครบถ้วน" และ "รวดเร็ว" สัจธรรมสามประการเกี่ยวกับข้อมูลที่สำคัญมาก โดยคำว่า "ถูกต้อง" นั้นพื้นฐานคือ accuracy คือตรงกับความเป็นจริง แต่แค่ไม่เพียงพอ ต้องคำนึงถึง consistency เพราะหากข้อมูลหลายชิ้นที่เกี่ยวข้องกัน แต่ละชิ้นถูกต้องแต่รวมกันแล้วไม่ถูกต้องเช่นนี้ก็ไม่ได้ เช่นชื่อของลูกค้า ในบัญชีหนึ่งเป็นอย่างหนึ่ง แต่ในอีกบัญชีเป็นอีกอย่างหนึ่งเพราะลูกค้าเปลี่ยนชื่อ



คำว่า "ครบถ้วน" นั้นมาจาก completeness คือแม้ว่าข้อมูลที่มีอยู่ในระบบจะถูกต้อง แต่ไม่ครบถ้วน ก็ไม่อาจเกิดประโยชน์ในภาพรวมแก่องค์กรได้ เพราะอาจเป็นข้อมูลที่ได้มาจากระบบงานย่อยแต่ไม่เพียงพอสำหรับที่จะใช้ในระบบงานอื่นหรือระบบงานในอนาคตได้ ซึ่งความครบถ้วนของข้อมูลนี้ยังเป็นปัญหาใหญ่ขององค์กรโดยทั่วไปเพราะการสร้างระบบสารสนเทศมักเกิดจากความต้องการระบบงานหรือบริการหรือการตอบคำถามความต้องการต่างๆ ขององค์กร ข้อมูลที่รวบรวมมาจึงมักเป็นเรื่องเฉพาะกิจ ไม่ได้มองระบบข้อมูลในภาพรวมขององค์กร เมื่อระบบสารสนเทศขององค์กรพัฒนาเติบโตไปเรื่อยๆ จะพบถึงการขาดแคลนหรือความไม่ครบถ้วน

ของข้อมูลนี้มากขึ้นเรื่อยๆ เพราะไม่ได้วางแผนออกแบบเผื่อไว้

ในส่วนของคำว่า "รวดเร็ว" นั้น แบ่งได้เป็นสามหลักการคือ in time หมายถึงข้อมูลที่ต้องนำมาใช้นั้นต้องมีให้ใช้ ไม่ใช่เมื่อจะใช้ค่อยไปหามา ซึ่งสิ่งนี้เป็นจุดอ่อนที่สำคัญของระบบสารสนเทศทั่วไปที่เมื่อมีความต้องการแล้วต้องรอฝ่ายไอทีพัฒนาระบบหาข้อมูล ฯลฯ สักสามเดือนหรือหกเดือนเสียก่อน จึงไม่ทันต่อความต้องการการใช้งาน โดยเฉพาะในโลกของการแข่งขันอย่างสูงเช่นนี้ ข้อมูลที่ได้มาเพื่อใช้ก็ต้องทันต่อเหตุการณ์ หรือ up to date ไม่ใช่ข้อมูลค้างเก่า แต่ควรส่งสะท้อนถึงความเป็นไปขององค์กรในปัจจุบันได้เป็นอย่างดี ตัวอย่างคือการเข้าไปดูสถิติต่างๆ ของประเทศไทย จะเห็นว่ามิหนวดยานเผยแพร่อยู่มากมายแต่มักเป็นข้อมูลของเมื่อสามปีหรือห้าปีที่แล้ว อยากได้ข้อมูลของปีนี้ต้องทำเรื่องทำหนังสือไปขอ เช่นนี้จะมีข้อมูลขึ้นเว็บไซต์ขององค์กรไปทำไม

นอกจากนั้นยังมีประเด็นเรื่อง real-time อีกด้วย คือหมายถึงข้อมูลนั้นต้องใช้เดี๋ยวนั้น ต้องเป็นข้อมูลล่าสุด "ทันกาล" เช่นข้อมูลที่ต้องใช้ในการหักเงินการใช้มือถือแบบ prepaid หรือเติมเงินเพราะเป็นระบบที่หักเงินค่าใช้จ่ายที่ได้จ่ายไว้แล้วไปตามเวลาที่โทรศัพท์คุยกัน ระบบทันกาลเหล่านี้เริ่มมีบทบาทมากขึ้นในระยะหลังเนื่องจากพัฒนาการของการนำระบบสารสนเทศมาใช้ให้บริการที่ทันสมัยภายใต้สภาพการณ์แข่งขันที่สูง

ความถูกต้องของข้อมูลนั้นถือได้ว่าเป็นปัจจัยแรกที่ต้องคำนึงถึงเพราะเกี่ยวข้องกับสัจธรรมของการ "ป้อนขยะ ได้ขยะ"

โดยตรง โดยแบ่งออกเป็นสามประเด็นด้วยกัน **Slide 22** กล่าวคือ ข้อมูลต้องมีค่าที่ตรงกับความเป็นจริงเช่นหากระบบระบุอายุของลูกค้าเป็นตัวเลขลงไป เช่นนี้อายุของลูกค้าจะไม่เปลี่ยนแปลงไปตามเวลาตามปีที่ผ่านไป

ในประเด็นที่สองเกี่ยวกับกลุ่มของข้อมูลซึ่งควรต้องสัมพันธ์กัน ไม่เกิดการแข่งหรือแปลกแยกไป เช่นในเรื่องอายุ หากเป็นดังตัวอย่างที่แล้ว อายุที่เป็นตัวเลขจะไม่สอดคล้องกับวันเดือนปีเกิดของลูกค้าเมื่อเวลาผ่านไป อีกตัวอย่างหนึ่งคือการระบุรหัสไปรษณีย์ไม่สอดคล้องกับจังหวัดของที่อยู่ ซึ่งอาจเกิดจากการป้อนข้อมูลที่ผิดพลาดของลูกค้าเพราะไปจำรหัสไปรษณีย์ของภูมิลำเนาเดิมจนเคยตัว นำมากรอกร่วมกับจังหวัดที่อาศัยอยู่ในปัจจุบัน เช่นนี้ระบบข้อมูลที่ดีควรมีการตรวจสอบความสัมพันธ์เหล่านี้ซึ่งทำได้ไม่ยากนัก

ส่วนประเด็นที่สามคือข้อมูลในระบบที่ซับซ้อนมักมีหลายสำเนา จะด้วยเป็นระบบย่อยที่ต่างก็มีข้อมูลของตนเองซึ่งอาจมีส่วนที่ไปทับซ้อนเหมือนกับข้อมูลของระบบย่อยอื่น หรืออาจเป็นการสำรองข้อมูลไว้ใช้ในกรณีฉุกเฉินเกิดความเสียหายกับข้อมูลหลัก หรืออาจเป็นการสำเนาข้อมูลหลายชุดแล้วกระจายไปอยู่ตามที่ต่างๆ เพื่อให้เกิดความสะดวกรวดเร็วในการใช้ข้อมูล แต่ไม่ว่าจะด้วยเหตุผลใด สำเนาทุกชุดของข้อมูลขึ้นเดียวกัน เช่นชื่อนามสกุล และที่อยู่ของลูกค้าผู้หนึ่งของธนาคาร ต้องถูกต้องตรงกันในทุกบัญชีที่ลูกค้าผู้นั้นเปิดไว้กับธนาคาร หรือในสัญญาบัญชีต่างๆ ที่ทำขึ้นกับธนาคาร เรื่องนี้ยากและซับซ้อนมากสำหรับระบบที่ไม่มีการวางแผนและออกแบบมาเป็นอย่างดี ระบบที่ประกอบด้วยระบบงานย่อยมากมาย จะเกิดปัญหาความไม่ถูกต้องตรงกันของข้อมูลที่มีหลายสำเนานี้เป็นอย่างมาก



เพื่อให้เข้าใจถึงประเด็นเกี่ยวเนื่องกับความถูกต้องของข้อมูลให้ผู้เรียนพิจารณากรณีศึกษาใน **Slide 23** จะพบว่าในแต่ละปีระบบ Admission สำหรับรับนักศึกษาเข้ามหาวิทยาลัย จะมีผู้สมัครที่กรอกชื่อหรือนามสกุลผิดกว่าแสนราย และในบรรดากระดาษคำตอบกว่าสองล้านใบในแต่ละปีนั้น จะมีการใช้ดินสอฝนเลขประจำสอบผิดประมาณห้าพันถึงหมื่นคนทุกปี หากเปรียบเทียบกับตัวเลขที่ได้จากการสำรวจต่างๆ เช่นค่านิยมในตัวนายกรัฐมนตรีที่ปรากฏให้เห็นตามผลการสำรวจหรือโพลต่างๆ จะพบว่าตัวเลขเหล่านั้นเป็นค่าโดยประมาณทั้งสิ้น แต่ในข้อมูลการสมัครเข้าสถาบันอุดมศึกษานั้นจะผิดไม่ได้

แม้แต่คนเดียว เพราะหมายถึงอนาคตของคนทุกคนที่จะพลาดโอกาสเข้าศึกษาต่อในมหาวิทยาลัย นั่นคือ ข้อมูลผิดพลาดเท่ากับนักเรียนเสียอนาคตทีเดียว แน่หนอนว่าอาจเห็นว่านักเรียนควรมีความรอบคอบและระมัดระวังมากกว่านี้ แต่เมื่ออยู่ในสนามสอบ นักเรียนบางคนจะตื่นเต้นอย่างมาก บางคนไม่ได้กรอกเลขประจำตัวสอบเลย มารู้อีกทีก็เมื่อออกจากห้องสอบแล้ว เป็นต้น ดังนั้นระบบข้อมูลที่ดีควรเพื่อในประเด็นเหล่านี้และหาทางเอื้ออำนวยให้การนำข้อมูลเข้านั้นกระทำได้อย่างถูกต้องมากขึ้น เช่นมีบัตรประจำตัวผู้สอบซึ่งมี barcode และมีระบบอ่านเข้าโดยอัตโนมัติ เป็นต้น ปัญหาใหญ่สองประการเกี่ยวเนื่องกับความถูกต้องของข้อมูลคือ ประการแรกจะสามารถทำให้ข้อมูลที่นำเข้าสู่ระบบนั้นถูกต้องได้อย่างไร และประการที่สองคือถ้าข้อมูลที่เข้าสู่ระบบมาแล้วเกิดไม่ถูกต้องจะทราบไหมและจะอย่างไร

ในส่วนที่เกี่ยวกับความครบถ้วนนั้น เป็นคุณสมบัติที่สำคัญสำหรับข้อมูลที่ดีของระบบสารสนเทศขนาดใหญ่ **Slide 24** ระบบเหล่านี้มักมีระบบงานย่อยมากมาย แต่ละระบบถือกำเนิดในเวลาและสภาวะของความต้องการที่แตกต่างกัน จึงมักมีความต้องการในด้านข้อมูลที่หลากหลาย แต่ละระบบจะขนขวายหาข้อมูลเฉพาะในส่วนที่ระบบงานย่อยนั้นต้องการเท่านั้น แต่สำหรับองค์กรแล้ว ต้องมีข้อมูลที่ครบถ้วน จะเอนเอียงไปสนองความต้องการของระบบงานใดไม่ได้ แต่ต้องสามารถสนองตอบทุกระบบงานทั้งที่มีอยู่และอาจมีขึ้นในอนาคตได้ด้วย จึงจะเรียกว่าระบบข้อมูลนั้นมีความครบถ้วน แน่แน่นอนว่าไม่จำเป็นที่บริษัทจะต้องรู้ทุกอย่างเกี่ยวกับลูกค้าของตน เช่น บริษัทโทรศัพท์ไม่จำเป็นต้องรู้เกี่ยวกับประวัติการรักษาพยาบาลของลูกค้า แต่ต้องรู้มาเพียงพอเพื่อสนองตอบงานของธุรกิจได้ครบถ้วนสมบูรณ์โดยเติมกำลังความสามารถที่จะจัดเตรียมหาไว้ได้ ตัวอย่างคือระบบกักกันโรคอาจต้องการประวัติการแพทย์ของผู้ป่วย แต่ไม่ต้องการเลขสวัสดิการการรักษาพยาบาล ในขณะที่ระบบสวัสดิการผู้ป่วยมีความต้องการที่สลับกัน เช่นนี้เป็นต้น ดังนั้นจึงเป็นหน้าที่ของผู้ออกแบบระบบข้อมูลให้เป็นโครงสร้างพื้นฐานที่ต้องจัดให้มีทั้งข้อมูลการแพทย์และข้อมูลสวัสดิการของผู้ป่วยไปในคราวเดียวกัน ไม่ใช่ค่อยๆ หาไปทีละเล็กละน้อย เพราะในการสัมภาษณ์หรือการเก็บข้อมูลผู้ป่วยนั้นอาจจะทำได้เพียงหนเดียวก่อนที่ผู้ป่วยจะหมดสติ ดังนั้นควรเก็บข้อมูลที่องค์กรคิดว่าน่าจะเป็นประโยชน์กับกิจกรรมขององค์กรในทุกภาคส่วนให้มากและครบถ้วนที่สุดในคราวเดียว



หากไม่มีสถาปัตยกรรม การวางแผนและออกแบบระบบข้อมูลอย่างดี จะทำให้แต่ละระบบงานย่อยไปขนขวายหาข้อมูลกันเอง และทำให้ข้อมูลที่องค์กรต้องการใช้ไม่ครบถ้วน และเกิดความซ้ำซ้อนในข้อมูลของแต่ละระบบงานย่อยอย่างมาก รวมถึงข้อมูลไม่ถูกต้องเพราะมีหลายสำเนา แยกกันดูแล

**Slide 25** สำหรับกรณีศึกษาเกี่ยวกับความครบถ้วนของข้อมูล จะให้ผู้เรียนพิจารณาระบบงานย่อยที่ความต้องการทางข้อมูลค่อยๆ เพิ่มพูนขึ้นทีละเล็กละน้อย ก่อความรำคาญ ความวุ่นวาย ความล่าช้า ในการปฏิบัติงานการเก็บข้อมูลและการพัฒนา



ระบบงานย่อยเหล่านั้น กรณีศึกษานี้ยกตัวอย่างที่เรียบง่ายมากๆ เพื่อความเข้าใจ เช่นระบบบริหารการศึกษาของโรงเรียนแห่งหนึ่ง ต้องการวิเคราะห์ถึงประเด็นโภชนาการของนักเรียน จึงต้องเก็บความสูงเฉลี่ยของนักเรียนในชั้นเป็นประจำ ดังนั้นข้อมูลที่ต้องการจึงเป็นความสูงของนักเรียนแต่ละคน คราวนี้ทางกระทรวงแนะนำว่าโภชนาการนั้นต้องแยกเป็นชายกับหญิงเพราะพัฒนาการในวัยเจริญพันธุ์นั้นแตกต่างกันในเพศทั้งสอง เช่นนี้ระบบสารสนเทศก็ต้องการข้อมูลเกี่ยวกับเพศของนักเรียนแต่ละคนด้วย แต่เมื่อโภชนาการนั้นขึ้นกับท้องถิ่น เพราะอาหารที่มีคุณค่าในแต่ละหมวดนั้นมีทางเลือกมากมาย บางชนิดก็หาได้ในท้องถิ่นนั้น บางชนิดหาไม่ได้ก็ไม่สามารถใช้อาหารที่ทดแทนกันได้

เช่นนี้ก็ต่อหาค่าเฉลี่ยความสูงตามภูมิภาคแล้ว ก็ต้องการข้อมูลที่อยู่ของนักเรียนแต่ละคนด้วย จะเห็นว่าการค่อยๆ เก็บข้อมูลในลักษณะนี้ย่อมเป็นความยุ่งยากลำบาก รวมทั้งทำให้เกิดความล่าช้าในการขยับความ

ต้องการจากแบบที่หนึ่งไปสองไปสาม ทำให้การพัฒนากระบวนการนั้นยุ่งยากซับซ้อน เสียเวลา และอาจเกิดความผิดพลาดได้ง่าย

การพิจารณาว่าข้อมูลใดจะเป็นประโยชน์แก่ระบบงานย่อยใด รวมไปถึงความต้องการในอนาคตนั้น เป็นเรื่องที่ยากมาก จำต้องผ่านการพิจารณาหาละเอียดโดยบุคลากรที่เข้าใจธรรมชาติและความต้องการขององค์กรหรือธุรกิจนั้นเป็นอย่างดี อย่าปล่อยให้เป็นที่ของนักคอมพิวเตอร์โดยเด็ดขาด และเมื่อมีข้อมูลเก็บเตรียมเอาไว้แล้ว การพัฒนาระบบงานย่อยตามความต้องการที่เพิ่มขึ้นหรือเปลี่ยนแปลงไป ย่อมจะทำได้สะดวกและรวดเร็ว

กรณีศึกษาอีกตัวหนึ่งเกี่ยวข้องกับความสัมพันธ์ของข้อมูลคือการที่ผู้ป่วยเปลี่ยนชื่อหรือนามสกุล ซึ่งเป็นเรื่องที่กระทำกันโดยทั่วไปโดยเฉพาะผู้ป่วยในต่างจังหวัดที่ถือโชคลางว่าหากเปลี่ยนชื่อแล้วจะช่วยให้พ้นจากความเจ็บป่วยได้ **Slide 26** ลักษณะการเช่นนี้หากผู้ป่วยเปลี่ยนชื่อกลางคัน ประวัติของเขาในปีที่แล้วจะไม่อาจมากระทบกับประวัติของเขาในปีนี้ได้เพราะชื่อไม่ตรงกัน เช่นนี้จะต้องมีวิธีการในการติดตามได้ว่ามีการเปลี่ยนชื่อ จำต้องเก็บชื่อไว้ทั้งของเดิมและของใหม่ โดยเฉพาะเมื่อการเปลี่ยนแปลงข้อมูลที่เป็นดิจิทัลในฐานข้อมูลของระบบจะเป็นการเปลี่ยนแปลงที่ไม่ทิ้งร่องรอยใดๆ ไว้เลย ดังนั้นผู้ออกแบบระบบข้อมูลต้องสร้างระบบที่จะทำให้เกิดการทิ้งร่องรอยเอาไว้



คุณสมบัตินี้สำคัญอีกประการหนึ่งของข้อมูลที่ดีนั่นคือต้องมี ความรวดเร็วทันต่อเหตุการณ์หรือความต้องการใช้งาน **Slide 27** ข้อมูลนั้นต้อง in time คือหมายถึงมีให้ใช้เมื่อต้องการใช้งาน ซึ่งไม่จำเป็นว่าข้อมูลต้องมีให้ใช้เสมอแม้เมื่อไม่ต้องการใช้งาน เพราะการทำให้ข้อมูลมีให้ใช้เช่นนั้นอาจเป็นค่าใช้จ่ายที่สูงสำหรับระบบงาน แต่ถ้าต้องการใช้ต้องมีใช้ ข้อมูลนั้นแม้จะมีให้ใช้เมื่อต้องการใช้ แต่อาจเป็นข้อมูลเก่าเก็บไม่ทันสมัย บางระบบงานอาจไม่เข้มงวดกับประเด็นนี้ แต่การที่ ข้อมูล up to date คือทันสมัยอยู่ตลอดนั้นจะทำให้เกิดการทิ้งร่องรอยที่ซับซ้อนและเื้อ้อำนวยกิจกรรมขององค์กรได้เป็น

อย่างมาก เช่น ยอดเงินในบัญชีที่จะถอนจาก ATM ซึ่งต้องเป็นยอดล่าสุดเสมอ ไม่เช่นนั้นจะทำให้เกิดผิดพลาดในการให้บริการได้

ในบางระบบอาจต้องการข้อมูลที่ทันสมัยระดับที่เข้มงวดมาก เพราะเป็นระบบที่ดำเนินการตอบสนองต่อสภาพทางกายภาพรอบตัว ระบบแบบ real time หรือระบบทันกาลนี้ ตัวอย่างได้แก่ ระบบสุบสารเคมีเข้าสู่ถังในโรงงานอุตสาหกรรม เช่นนี้จำเป็นต้องได้ข้อมูลเป็นระดับของสารเคมีในถัง ข้อมูลนี้ต้องรวดเร็วมาก เพราะหากช้า สารเคมีเต็มถังแล้วแต่ข้อมูลใหม่ยังไม่มา เครื่องสุบสารเคมีนั้นอาจหยุดไม่ทันทำให้สารเคมีล้นถังเกิดเป็นอันตรายได้

ตัวอย่างอีกอันหนึ่งที่ทำให้ระบบสารสนเทศมีบทบาทสำคัญอย่างมากในสังคมคือการนำระบบเหล่านี้ไปใช้ควบคุมเครื่องยนต์



กลไกที่ซับซ้อน เช่นการควบคุมการจุดระเบิดในเครื่องยนต์สันดาปภายในของรถยนต์ เพื่อให้เกิดอัตราเร่งที่เหมาะสมและประหยัดน้ำมัน เช่นนี้ระบบสารสนเทศหรือในที่นี้คือระบบควบคุมการทำงานของเครื่องยนต์ ต้องได้ข้อมูลเกี่ยวกับมุมการจุดระเบิดและความเร็วรอบเครื่องยนต์ รวมถึงข้อมูลต่างๆ เกี่ยวกับการใช้เครื่องยนต์และรถยนต์อีกมากมาย ในเวลาอันสั้นระดับเป็นหนึ่งในส่วนพันของวินาที เพื่อให้ระบบสามารถคำนวณองศาการจุดระเบิดที่เหมาะสมส่งคืนกลับไปให้กับเครื่องยนต์ได้ทันเวลา ซึ่งการควบคุมเครื่องยนต์กลไกที่ซับซ้อนนี้นำไปใช้ในระบบต่างๆ มากมายสุดจะจรรย์ เช่นควบคุมลิฟต์ไปจนถึงยานอวกาศ เครื่องบินรบสมัยใหม่นั้นไม่อาจบินได้โดยไม่มีระบบสารสนเทศคอยช่วยในการควบคุมการบินหรือเครื่องยนต์กลไกในเครื่องบินได้เลย

สถาปัตยกรรมของระบบข้อมูลที่ซับซ้อนนั้นประกอบด้วย 3 หลักการที่สำคัญ **Slide 28** คือมาตรฐาน การผนวกรวม และ การใช้ข้อมูลร่วมกัน

ต้องขอขำต่อผู้เรียนอยู่เสมอว่า ระบบข้อมูลที่ดีนั้นต้องมีการวางสถาปัตยกรรมหรือภาพรวมที่ชัดเจน เป็นลายลักษณ์อักษร และมีการออกแบบระบบให้ครอบคลุมทั้งสามประเด็นเหล่านี้ ไม่งั้นจะเกิดปัญหายุ่งยากตามมาเมื่อนำระบบข้อมูลนั้นไปใช้ประโยชน์อย่างกว้างขวางครอบคลุมองค์กรและปฏิบัติใช้ในระยะเวลา และในทางตรงกันข้าม หากไม่มีการนำระบบข้อมูลที่เหมาะสมมาบังคับใช้กับทั้งองค์กร ปล่อยให้มีการดำเนินการสร้างระบบงานย่อยของฝ่ายต่างๆ ขององค์กรกันไปอย่างเป็นอิสระเป็นเอกเทศ ความต้องการขององค์กรในภาพรวมและในระยะยาวจะเกิดปัญหาหลัก ระบบสารสนเทศที่เป็นเอกเทศในทางข้อมูลในลักษณะที่กระจายเหล่านั้น จะไม่อาจสนองตอบความต้องการขององค์กรได้อย่างดีหรือทันทั่วถึง



องค์ประกอบแรกของสถาปัตยกรรมข้อมูลนั้นคือมาตรฐานข้อมูล หรือ Data Standardization ตาม **Slide 29** ซึ่งในสมัยก่อนที่เทคโนโลยีและความเข้าใจของเราเกี่ยวกับระบบสารสนเทศยังเยาว์อยู่ ระบบงานย่อยต่างๆ จะแยกกันเป็นเอกเทศต่างฝ่ายต่างดำเนินการออกแบบ จัดทำ และดูแลใช้งานแยกกันไป ดังนั้นข้อมูลที่เป็นไปในทำนองเดียวกัน อย่างเช่นรหัส นิสิตนักศึกษากลับมีการแยกเป็นปริญญาตรีกับปริญญาบัณฑิต คือ โทและเอก เนื่องจากอยู่ภายใต้ความรับผิดชอบของคนละฝ่ายในมหาวิทยาลัยโดยทั่วไปโดยเฉพาะในมหาวิทยาลัยขนาดใหญ่ ทำให้การทำการประมวลผลหรือวิเคราะห์ข้อมูลข้ามระบบกันนั้นทำได้ยาก นิสิตนักศึกษาที่จบปริญญาตรีแล้วเข้าศึกษาต่อในระดับปริญญาโทของมหาวิทยาลัยเดียวกันจะกลายเป็นนิสิตนักศึกษาสองคนไปในระบบสารสนเทศของมหาวิทยาลัย ยิ่งหากนิสิตนักศึกษาสะกดชื่อผิดจะทำให้แทบจะไม่สามารถติดตามนิสิตนักศึกษาผู้นั้นข้ามจากระดับปริญญาตรีไปสู่ระดับปริญญาโทได้เลย เหล่านี้เป็นปัญหามาจากการที่ไม่มีมาตรฐานข้อมูลใช้ในมหาวิทยาลัยนั้น ซึ่งรวมไปถึงการนำข้อมูลจากมหาวิทยาลัยต่างๆ มาประมวลผลร่วมกันเพื่อประโยชน์ในการบริหารจัดการระดับชาติของกระทรวงศึกษาธิการด้วย มาตรฐานข้อมูลอาจกำหนดรูปแบบของรหัส นิสิตนักศึกษาให้เป็นแบบเดียวกันทั้งมหาวิทยาลัยเพื่อสะดวกต่อการเชื่อมโยงข้อมูลของนิสิตนักศึกษาได้อย่างเป็นระบบ หลักการทางสถาปัตยกรรมของระบบข้อมูลนี้ไม่ได้หมายถึงทุกอย่างทุกระบบงานย่อยจะต้องรวมศูนย์ ทำโดยฝ่ายหรือหน่วยงานเดียวของส่วนกลางขององค์กร แต่สามารถแยกกันทำ กระจายไปให้อยู่ในความรับผิดชอบของหน่วยงานย่อยขององค์กรได้อย่างอิสระ เพียงแต่ทุกระบบงานย่อยต้องใช้ข้อมูลที่ยึดตามมาตรฐานที่กำหนดไว้สำหรับองค์กรเพื่อเอื้ออำนวยการใช้งานระดับองค์กรและการแลกเปลี่ยนข้อมูลเพื่อการใช้งานร่วมกัน การแบ่งปันข้อมูลข้ามระบบ ไม่ใช่ว่าข้อมูลนิสิตนักศึกษา

แต่ละคนจะถูกเก็บโดยฝ่ายทะเบียน ฝ่ายกิจการนิสิต ฝ่ายนิสิตสัมพันธ์ รวมถึงสมาคมนิสิตเก่า เหล่านี้เป็นต้น  
กรณีศึกษาที่เห็นได้ชัดคือเมื่อพ.ศ. 2517 ศูนย์คอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัยใช้เวลาถึง 2 ปีเพื่อประสานและจัด



ทำมาตรฐานข้อมูลห้าด้านของมหาวิทยาลัยคือ นิสิต บุคลากร รายวิชา การเงิน และอาคารสถานที่ โดยอาศัยความร่วมมือช่วยเหลือจากหน่วยงานทุกภาคส่วนของมหาวิทยาลัยในการให้คำแนะนำ การแยกแยะและปรึกษาหารือถึงความจำเป็นและความหมายของข้อมูลแต่ละชั้นแต่ละประเภท ทั้งนี้เพื่อให้ทุกภาคส่วนของมหาวิทยาลัยมีภาษากลางในเชิงข้อมูลใช้ร่วมกัน แต่ปัญหาคือนักคอมพิวเตอร์ในสมัยนั้นมีความคิดก้าวหน้าเกินไปมาก ทำให้เทคโนโลยีและความเข้าใจของผู้บริหารงานและผู้ปฏิบัติการไม่เข้าใจและไม่ยอมรับในมาตรฐานนี้ จึงไม่มีใครนำไปใช้ปฏิบัติ แต่ตอนนี้เริ่มเห็นความสำคัญกันแล้ว ดังจะเห็นได้จากการที่รัฐบาลทำ

โครงการ GDX หรือ Government Data Exchange โดยหวังว่าจะสร้างมาตรฐานข้อมูลกลางของรัฐขึ้นมาเพื่อให้สามารถส่งผ่านแลกเปลี่ยนข้อมูลกันระหว่างหน่วยงานต่างๆ ของรัฐได้ แต่จนถึงปัจจุบันก็ยังไม่เห็นว่าจะประสบความสำเร็จแต่อย่างใด ซึ่งในส่วนของมาตรฐานข้อมูลนี้ เนื่องจากมีความสำคัญยิ่งยวดกับระบบข้อมูลที่ซับซ้อน จึงจะได้ขยายความในคาบเรียนถัดไป

มาตรฐานของข้อมูลดังที่ได้กล่าวไปแล้ว เป็นเพียงการสร้างกรอบการใช้ข้อมูลในระดับของมโนคติขององค์กร เป็นบทบัญญัติหรือแบบแผนในการทำงาน แต่การจะให้ส่วนงานย่อยขององค์กรที่ต่างก็มีข้อมูลของตนเอง ให้สามารถแบ่งปันและแลกเปลี่ยนข้อมูลกันใช้ รวมถึงการนำข้อมูลเหล่านั้นมารวมกันเพื่อสนองความต้องการโดยรวมขององค์กร เหล่านี้จำเป็นต้องอาศัยองค์ประกอบที่สองของสถาปัตยกรรมคือการผนวกรวมข้อมูล หรือ Data Integration ใน Slide 30

พัฒนาการทางสารสนเทศขององค์กรนั้นทำให้แต่ละฝ่ายที่ทำงานเป็นเอกเทศ ต่างเก็บข้อมูล "ของตน" คือของที่ต้องใช้ในส่วนที่ฝ่ายนั้นรับผิดชอบอยู่ เช่นฝ่ายการเงิน ข้อมูลเหล่านี้มักมีส่วนที่ซ้ำกัน และที่สำคัญคือส่วนที่ซ้ำกันนี้ต่างฝ่ายต่างจะพิจารณาว่าเป็น "ของตน" เพราะเป็นข้อมูลที่ฝ่ายนั้นลงทุนลงแรงเสียเวลาในการได้มา อัตราของแต่ละส่วนงานนี้ทำให้ข้อมูลที่ซ้ำซ้อนกันเหล่านี้ไม่ถูกต้องคือไม่สอดคล้องกัน หรือ inconsistent เพราะต่างฝ่ายต่างดูแลกันเอง ดังนั้นนิสิตนักศึกษา พนักงาน หรือลูกค้าขอเปลี่ยนที่อยู่ก็ขึ้นกับว่าไปขอเปลี่ยนที่ใดในองค์กร การเปลี่ยนแปลงนั้นจะไม่ถูกกระจายไปยังส่วนอื่นๆ ของนิสิตนักศึกษา พนักงาน หรือลูกค้าผู้นั้น ลักษณะเช่นนี้เป็นข้อผิดพลาดทางข้อมูลที่เกิดขึ้นในระดับประเทศด้วย โดยหากพนักงานของบริษัทเปลี่ยนชื่อหรือนามสกุล พนักงานผู้นั้นต้องตามเปลี่ยนไปยังระบบงานสารสนเทศมากมายหลายระบบทั้งในและนอกบริษัท ตั้งแต่ ฝ่ายบุคคล ฝ่ายผลิต ฝ่ายสวัสดิการพนักงาน แล้วออกไปเปลี่ยนที่กรมการปกครองซึ่งดูแลเลขรหัส 13 หลักประจำตัวประชาชน ไปเปลี่ยนที่กรมแรงงาน ธนาคารทุกแห่งและทุกบัญชีที่มีเปิดไว้ เปลี่ยนในใบขับขี่ที่กรมการขนส่ง ซึ่งทำให้ไม่ประหลาดใจที่ที่อยู่ในใบขับขี่เป็นอะไรที่ไม่มีใครสนใจเพราะโอกาสถูกต้องตรงความเป็นจริงนั้นมีน้อยเสียยิ่งกว่าน้อย แล้วไปเปลี่ยนที่บัตรเครดิต 3 ใบที่ใช้อยู่ รวมถึงบัตรส่วนลดของ



สหกรณ์และร้านค้าต่างๆ ด้วย

ข้อมูลที่ไม่ได้ผนวกรวมเข้าด้วยกันตามมาตรฐานที่วางไว้ทำให้เกิดปรากฏการณ์ที่นิสิตนักศึกษาที่จบปริญญาตรีแล้วเข้าศึกษาต่อปริญญาโทในมหาวิทยาลัยเดิมกลายเป็นคนอีกคนหนึ่งเพราะระบบทะเบียนนิสิตนักศึกษาทั้งสองส่วนแยกเป็นเอกเทศจากกัน ไม่อาจใช้ข้อมูลของคนคนเดียวกันร่วมกันได้

การแตกเป็นเสี่ยงๆ นี้องค์กรขนาดใหญ่ที่ไม่อาจผนวกรวมระบบงานย่อยหลากหลายมากมายเข้าด้วยกันได้ใช้วิธีแก้ปัญหาแบบกำปั่นทุบดินโดยการทุ่มทุนซื้อและจัดทำระบบรวมศูนย์อย่าง SAP ในราคานับหลายสิบล้านหรือหลายร้อยล้านบาท โดยการยกเลิกระบบงานหลักอื่นๆ อย่างระบบบัญชี ระบบสินค้าคงคลัง ระบบพนักงาน ระบบจัดซื้อ เหล่านี้เข้าด้วยกันเป็นระบบเดียว ซึ่งอาศัยการลงทุนมหาศาล ผู้บริหารระดับสูงเข้ามาสั่งการบงการ จึงทำให้เป็นไปได้ที่จะบังคับหักดิบหน่วยงานย่อยต่างๆ ให้เข้ามาตรฐานและระบบงานกลางนี้ได้

องค์ประกอบสุดท้ายที่สำคัญของสถาปัตยกรรมข้อมูลคือการใช้ข้อมูลร่วมกัน หรือการแบ่งปันข้อมูลกัน ที่เรียกว่า Data Sharing ตาม **Slide 31** เพราะข้อเท็จจริงประการหนึ่งคือข้อมูลพื้นฐานนั้นก่อกำเนิดขึ้นจากการปฏิบัติงานของหน่วยงานย่อยขององค์กร ซึ่งข้อมูลเหล่านี้สุดท้ายจำเป็นต้องนำมาใช้ร่วมกัน แบ่งปันกันระหว่างหน่วยงานต่างๆ ขององค์กร ตัวอย่างเดิมคือข้อมูลนิสิตนักศึกษาที่ขยับจากปริญญาตรีขึ้นมาศึกษาต่อในระดับปริญญาโทของมหาวิทยาลัยเดียวกัน หากมีการแบ่งปันข้อมูลใช้ร่วมกันระหว่างฝ่ายปริญญาบัณฑิตกับฝ่ายบัณฑิตศึกษา นิสิตนักศึกษาผู้นั้นจะสามารถคงความเป็นตัวเองไว้ได้เมื่อ



ขยับเลื่อนชั้นหลักสูตร ไม่กลายเป็นสองคนอย่างในระบบที่ไม่มีมีการแบ่งปันข้อมูลใช้ร่วมกัน

ในสมัยก่อนนั้นไม่อาจเรียนข้ามคณะได้เพราะไม่มีมาตรฐานข้อมูลกลางหรือการผนวกรวมระบบงานย่อยต่างๆ เข้าให้ทำงานกลมกลืนกัน ดังนั้นมหาวิทยาลัยส่วนมากจึงใช้วิธีสร้างระบบทะเบียนกลางขึ้นมาเพื่อแก้ปัญหาที่นี้ ซึ่งทำให้การทำงานรวมศูนย์สามารถใช้ข้อมูลร่วมกันได้แต่ในขณะเดียวกันหน่วยงานกลางที่เกิดขึ้นก็กลายเป็นโซห่วยที่เพิ่มขึ้นทำให้ขั้นตอนการทำงานยุ่งยากขึ้นเป็นเงาตามตัว

เมื่อระบบงานกระจายแพร่หลายไปในส่วนงานต่างๆ มากขึ้น

ระบบทะเบียน ระบบบุคลากร ระบบการเงิน รวมไปถึงระบบบริหารอาคารสถานที่ต่างๆ ทำให้เกิดความต้องการของผู้บริหารมหาวิทยาลัยที่จะทราบถึงสถานภาพต่างๆ ของมหาวิทยาลัยในภาพรวมเช่นด้านการเงินหรือเรื่องเงินสดในมือ เหล่านี้ทำให้เกิดแรงผลักดันของการผนวกรวมระบบเข้าด้วยกันเพื่อใช้ข้อมูลร่วมกัน แต่ความยากลำบากทำให้เกิดทางออกแบบเดียวกับหน่วยงานหรือองค์กรธุรกิจขนาดใหญ่ใช้กัน คือนำซอฟต์แวร์ขนาดใหญ่อย่าง SAP มาใช้โดยลงทุนมหาศาลนับหลายสิบล้านเพื่อให้สามารถตอบโจทย์ความต้องการของภาพรวมของมหาวิทยาลัยได้

บทถัดไปจะเป็นการขยายความในเรื่องของมาตรฐานข้อมูลทีถือเป็นบรรทัดฐานหรือพื้นฐานที่สำคัญของระบบสารสนเทศที่ซับซ้อนขององค์กรขนาดใหญ่ต่อไป