

การวิจัยเชิงปริมาณทางสังคมศาสตร์ในโลกใหม่: บทสำรวจเบื้องต้น*

Social Science Quantitative Research in the New World: Preliminary Exploration

เรวัต แสงสุริยงค์ (Rewat Sangsuriyong)**

บทคัดย่อ

การวิจัยเชิงปริมาณด้านสังคมศาสตร์บนโลกเก่ากำลังถูกอิทธิพลของโลกใหม่บังคับให้ต้องปรับตัวไปใช้ข้อมูลขนาดใหญ่บนโซเชียลมีเดีย เนื่องจากมนุษย์ได้เข้าไปสร้างสังคมเสมือนจริงและข้อมูลที่เกี่ยวข้องกับการกระทำทางสังคม ปรากฏการณ์ทางสังคม และหลักฐานที่เป็นดิจิทัลอย่างรวดเร็ว หลากหลาย คลุมเครือ ผันผวน แปรผัน หนืด และจำนวนมหาศาล ดังนั้นจึงจำเป็นต้องมีการเรียนรู้ใหม่เกี่ยวกับนวัตกรรมด้านการรวบรวมข้อมูล เก็บข้อมูล วิเคราะห์ข้อมูล และนำเสนอข้อมูล

วิธีการหาความรู้พื้นฐานความเชื่อของนักสังคมศาสตร์ อาจต้องมีการทบทวน/ปรับแต่งในบางส่วนให้เหมาะกับโลกใหม่ เพราะมีวิธีการ/เทคโนโลยีใหม่ที่เกิดขึ้นมาสำหรับใช้เป็นเครื่องมือในการเข้าถึง รวบรวม และวิเคราะห์ข้อมูลขนาดใหญ่ ซึ่งอาจทำให้ได้คำตอบใหม่/สนับสนุน หรือนำผลการศึกษาไปเปรียบเทียบกับวิธีการพื้นฐาน และการวิจัยแบบประยุกต์ ที่ได้ข้อมูลมาจากโลกเก่าต่อไป

คำสำคัญ: โลกใหม่ โซเชียลมีเดีย อินเทอร์เน็ต ระเบียบวิธีการวิจัย ข้อมูลขนาดใหญ่ สังคมศาสตร์

* เพื่อเผยแพร่ผลงานวิชาการ

** ผู้ช่วยศาสตราจารย์ประจำภาควิชาสังคมวิทยา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา e-Mail: rewat@buu.ac.th

Assistant Professor, Sociology Department, Faculty of Humanities and Social Sciences, Burapha University.

Abstract

When it comes to the new world, quantitative approaches in traditional style of social sciences in the old era has been compelled to use big data on cyberspace. Since human beings have created the virtual society and various aspects of information including social action, social phenomena, and evidence that is digitalized in the manner of rapidity, variety, variant, ambiguity, instability, volatility and giant. Therefore, it is necessary to new learning approaches for innovation in data collection, data analysis and visualization.

Knowledge seeking approaches of social sciences should have been reviewed, adjusted, partially, in order to have better fitted with the new world because approaches and technologies are available to as tools to access, assess and analyze big data, which could provide opportunities to find new answers or to support or compare results of studies with basic research and applied research obtained from data in the old world.

คำสำคัญ: New world, Cyberspace, Internet, Research methodology, Big data, Social sciences

บทนำ

พัฒนาการและความก้าวหน้าของเทคโนโลยีสารสนเทศและการสื่อสารทำให้เกิดคำว่า “ไซเบอร์สเปซ (cyberspace)” เพื่อใช้อธิบายพื้นที่บนโลกใหม่ที่เกิดจากเครือข่ายคอมพิวเตอร์ขนาดใหญ่ หรือ “อินเทอร์เน็ต (internet)” ที่มีการขยายตัวอย่างต่อเนื่องและซับซ้อน และได้กลายมาเป็นวัฒนธรรมของประชาชนที่ได้รับความนิยม (popular culture) มีและใช้ในชีวิตประจำวันอย่างแพร่หลาย แต่คำว่า ไซเบอร์สเปซ ได้รับความนิยมน้อยมาก เพราะมีการประดิษฐ์คำขึ้นมาใช้กันใหม่เป็นจำนวนมาก เช่น โลกเสมือนจริง (virtual world) โลกดิจิทัล (digital world) และพื้นที่สารสนเทศ (information space) เป็นต้น ขณะที่คำว่า อินเทอร์เน็ต ยังได้รับความนิยมใช้กันอย่างแพร่หลาย

การวิจัยเชิงปริมาณ (quantitative research) เป็นวิธีการวิจัยวิธีการหนึ่งที่ได้รับคามนิยมอย่างแพร่หลายในหลายสาขาวิชาด้านสังคมศาสตร์ และเป็นวิธีการวิจัยที่นักวิจัยและวิจารณ์ถึงความไม่เหมาะสมอย่างต่อเนื่องในการนำมาใช้กับการค้นหาความรู้ด้านสังคมศาสตร์ในหลายเรื่องจากกลุ่มนักวิจัยที่นิยมการวิจัยเชิงคุณภาพ (qualitative research) และคงยังมีต่อไป และเมื่อสังคมมนุษย์วิวัฒนาการเข้าสู่ยุคสังคมสารสนเทศ (information society era) การวิจัยเชิงปริมาณต้องเผชิญกับการเปลี่ยนแปลงในหลายด้านซึ่งอาจเป็นจุดให้โจมตีเพิ่มขึ้นได้อีก เช่น พื้นที่การวิจัย ปริมาณและรูปแบบของข้อมูล การเลือกใช้ข้อมูล และการเลือกใช้สถิติในการวิเคราะห์ข้อมูล ดังนั้นจึงต้องทำการค้นหาความรู้เบื้องต้นจากอินเทอร์เน็ตเพื่อพรรณนาและวาดภาพให้เห็นว่า สังคมใหม่คืออะไร ข้อมูลที่มีอยู่ในโลกใหม่เป็นอย่างไร การหาความรู้บนโลกใหม่กับในโลกเก่าแตกต่างกันหรือไม่อย่างไร การวิเคราะห์ข้อมูลในโลกใหม่เป็นอย่างไร อะไรคือปัญหาและความท้าทายของการวิจัยในโลกใหม่ และอะไรคือผลกระทบของการวิจัยจากโลกใหม่

คำว่า “โลกใหม่ (new world)” เป็นคำอุปมาอุปไมย (metaphor) ถูกนำมาใช้เรียกสิ่งต่างๆ ที่มีความแตกต่างกันอย่างหลากหลาย ทั้งด้านดินแดน ด้านชีววิทยา ด้านพันธุ์พืชและสัตว์ และเครื่องดื่ม รวมถึงด้านความคิดและการกระทำ จึงไม่ใช่เรื่องแปลกหรือใหม่ที่บทความนี้จะใช้คำว่า “โลกใหม่” แทน “อินเทอร์เน็ต” และคำอื่นๆ ที่กล่าวมาแล้ว เพราะในระบบอินเทอร์เน็ตหรือไฮเบอร์สเปซเป็นพื้นที่จัดเก็บข้อมูลที่ถูกแปลงไปจากสรรพสิ่งบนโลกมนุษย์และสามารถทำให้เสมือนจริงขึ้นมาได้เหมือนที่เห็นอยู่บนโลกมนุษย์ มีการสร้างตัวตนหรือสิ่งแทนตัวตน (avatar) เช่น สมนาม (alias) สัญลักษณ์ (symbol) สัญลักษณ์ (icon) เนื้อหา (content) เข้าไปรวมกลุ่มเป็นสังคมขนาดเล็กและขนาดใหญ่ มีการสร้างกิจกรรมด้านเศรษฐกิจ การเมือง ศาสนา ฯลฯ และมีการปฏิสัมพันธ์ทางสังคม (social interaction) เสมือนอยู่บนโลกทางกายภาพ (physical world)

คงมีคำถามว่า *โลกใหม่* เกิดขึ้นมาอย่างไร และเมื่อใด? ความจริงแล้วโลกใหม่เกิดขึ้นในพื้นที่หน่วยความจำ (memory) ของเครื่องคอมพิวเตอร์ทางด้านตะวันตกของประเทศสหรัฐอเมริกาบนทวีปอเมริกา ซึ่งเคยเป็นดินแดนใหม่หรือโลกใหม่ที่ชาวยุโรปเดินทางมาพบ * เมื่อนานมาแล้ว แต่การเริ่มเกิดโลกใหม่ในคริสต์ศตวรรษที่ 20 ไม่ได้เกิดมาจากนวัตกรรมด้านการเดินทางเหมือนในอดีต เพราะเกิดมาจากการทำงานของอิเล็กทรอนิกส์ (electron) ในรูปของสัญญาณแอนะล็อก (analog) และสัญญาณดิจิทัล (digital) ในยุคสงครามเย็นระหว่างค่ายเสรีประชาธิปไตยกับค่ายสังคมนิยมคอมมิวนิสต์ เมื่อวันที่ 29 ตุลาคม ค.ศ. 1969 เวลา 22.30 น. ภายใต้โครงการ The Advanced Research Projects Agency Network: ARPAnet ของกระทรวงกลาโหมของสหรัฐอเมริกา ที่เริ่มทดลองสร้างเครือข่ายคอมพิวเตอร์ขนาดเล็ก ๆ เพื่อใช้ในการรับส่งข้อมูลทางการทหาร ผลการทดลองการรับส่งข้อมูลผ่านระบบเครือข่ายคอมพิวเตอร์ประสบความสำเร็จเกินคาด แต่มีขอบเขตการใช้งานอยู่ในวงจำกัด จนถึงปี ค.ศ. 1991 จึงได้อนุญาตให้นำไปใช้เพื่อวัตถุประสงค์ทางการค้า ทำให้เกิดระบบเครือข่ายคอมพิวเตอร์เชื่อมโยงกันทั่วโลก (international network) หรือที่เรียกกันสั้นๆ ว่า อินเทอร์เน็ต (internet) (เรวัต แสงสุริยงค์, 2541) และกลายมาเป็นโลกไร้พรมแดนที่ไม่ได้มีรูปทรงสี่เหลี่ยมทรงกลม/รี/แป้น ไม่มีพื้นดิน ไม่มีสิ่งมีชีวิตอาศัยอยู่ แต่มีรูปทรงแบนราบ (flat) (ฟรีดแมน, 2554)

การค้นพบและการนำเอาความรู้ใหม่ที่ได้จากธรรมชาติมาสร้างเป็นนวัตกรรม ทำให้โลกมนุษย์และโลกใหม่มีสิ่งที่เหมือนกันคือ มีสิ่งที่เกิดจากธรรมชาติ (กระแสไฟฟ้า) และสิ่งที่มนุษย์สร้างขึ้น (ข้อมูล) แต่โลกใหม่มีบริบทที่ห้อมล้อม (context) ไปด้วยเทคโนโลยีที่เป็นระบบอิเล็กทรอนิกส์ (electronic) โดยมีเทคโนโลยีการสื่อสาร (communication technology) เชื่อมโยงให้เข้ามารวมเป็นพื้นที่เดียวกัน มีการยืดและหด (elasticity) ขนาดของพื้นที่อยู่ตลอดเวลา แต่มีอัตราส่วนของการยืดขยายมากกว่าการหดลงหลายเท่าตัว

ข้อมูลในโลกใหม่

ข้อมูลเป็นสิ่งที่อยู่จริง (ontology) ในธรรมชาติและเป็นความจริงแท้ที่นักวิจัยต้องนำมาวิเคราะห์สนับสนุนคำตอบในการค้นหาความรู้ ข้อมูลโดยตัวของมันเองมีความเป็นธรรมชาติหรือเป็นกลางและไม่มี ความหมาย แต่มนุษย์หรือนักวิจัยเป็นผู้สร้างความหมายให้กับข้อมูลเมื่อนำไปใช้งาน (Doorn, 2014) หากย้อนไปดูประวัติศาสตร์ของมนุษย์เราจะพบว่าข้อมูลที่มนุษย์เริ่มสร้างขึ้นก็เป็นเรื่องเกี่ยวกับธรรมชาติ มีการบันทึกไว้ใน

* เลฟ เอริกสัน (Leif Ericsson) เดินทางถึงเมื่อ ค.ศ. 1000 และคริสโตเฟอร์ โคลัมบัส (Christopher Columbus) เดินทางถึงเมื่อ ค.ศ. 1492

พื้นที่ (storage) ที่เป็นธรรมชาติ แต่ข้อมูลที่มนุษย์ทำการบันทึกไว้มีความหมายแตกต่างกันไปตามความคิดและการตีความ (interpretative) ของมนุษย์แต่ละคน ดังนั้น ข้อมูลจึงเป็นทั้งความจริง (reality) เมื่อมันอยู่อย่างอิสระจากการรับรู้ของมนุษย์ และเป็นสิ่งที่มนุษย์สร้างขึ้น (socially constructed) และมโนคติ (idea) เมื่อเข้ามาอยู่ในการรับรู้ของมนุษย์

วิวัฒนาการของสังคมแต่ละยุคได้สร้างและมอบเทคโนโลยีให้เป็นมรดกกับสังคมในยุคต่อมาเสมอ ในยุคสังคมดั้งเดิม (traditional society) มนุษย์ได้สร้างภาษาพูดและภาษาเขียนไว้ให้สังคมสมัยใหม่ (modern society) นำเอาไปสร้างเทคโนโลยีการพิมพ์ด้วยเครื่องจักรและส่งต่อให้สังคมหลังสมัยใหม่ (post-modern society) ใช้เป็นข้อมูลและสารสนเทศในโลกใหม่ การเริ่มบันทึกข้อมูลของมนุษย์ที่บันทึกไว้บนพื้นที่ที่มีอยู่ตามธรรมชาติ เช่น ผนังถ้ำ แผ่นดินเหนียว ใบไม้ ฯลฯ กระจายอยู่ทั่วไป ไม่มีการจัดเก็บอย่างเป็นระบบ และไม่สามารถนับจำนวนได้ว่ามีจำนวนเท่าไร การรวบรวมและสะสมข้อมูลไว้อย่างเป็นระบบเกิดขึ้นครั้งแรกในยุคกรีกโบราณที่ห้องสมุดอเล็กซานเดรีย (สร้างเมื่อศตวรรษที่ 3 ก่อนคริสต์ศักราช) ทำให้เริ่มนับจำนวนได้อย่างคร่าวๆ คือ มากสุดประมาณ 400,000 ม้วน และน้อยสุดประมาณ 40,000 ม้วน หรืออาจน้อยที่สุดประมาณหลักพัน[†]

เมื่อสังคมมนุษย์วิวัฒนาการเข้าสู่สังคมอุตสาหกรรม การผลิตและการบันทึกข้อมูลด้วยเครื่องจักรบนสื่อหลายรูปแบบ เช่น หนังสือ เทปเสียง รูปภาพ และฟิล์มภาพยนตร์ ทำได้อย่างรวดเร็วและเป็นจำนวนมากจนเกิดปรากฏการณ์ “การท่วมท้นของสารสนเทศ (information explosion)”[‡] และเป็นที่มาของแนวคิดการหาพื้นที่ที่มีจำนวนน้อยที่สุดสำหรับจัดเก็บสารสนเทศที่เกิดขึ้นจำนวนมหาศาล ผลที่ตามมาคือ ทำให้เกิดการแปลงรูปข้อมูล/สารสนเทศและนำเข้าไว้ในคอมพิวเตอร์เพิ่มขึ้นอย่างรวดเร็ว (Forbes, 2013) การนับจำนวน/ปริมาณข้อมูลและสารสนเทศที่บันทึกอยู่ในสื่อแบบเก่า (non-electronic media) และช่วงแรกของสื่อแบบใหม่ (electronic media) ทำได้เพียงการนับแบบรวมเป็น ชุด (volume) การนำมาใช้ประกอบการค้นหาความรู้ต้องทำการแปลงรูปและประมวลผลด้วยความสามารถของมนุษย์ที่มีอยู่อย่างจำกัด จนถึงปี ค.ศ. 1975 กระทรวงไปรษณีย์และโทรคมนาคมของประเทศญี่ปุ่นได้เริ่มนำเอาวิธีการนับจำนวนคำ (amount of words) มาใช้ในการเก็บข้อมูลการไหลเวียนของสารสนเทศภายในประเทศญี่ปุ่น ทำให้สามารถวัดข้ามสื่อทั้งหมดได้ และใน ค.ศ. 1981 สำนักงานสถิติกลางประเทศอังกฤษได้เริ่มทำการวิจัยการนำจำนวนสารสนเทศด้วย บิต (bits) และ ทำเป็นคู่มือการคำนวณสารสนเทศของประเทศ (National Accounting of Information) ออกมาเผยแพร่ (Hilbert, 2012) เมื่อเกิดการแพร่กระจายและแทรกซึมของอินเทอร์เน็ตเข้าสู่ทุกอณูของสังคมและความสามารถในการประมวลผลแบบกลุ่มเมฆ (cloud computing) ทำให้เกิดกระบวนการแปลงและผลิตข้อมูลขนาดใหญ่/อันมหาศาล (big data) เก็บไว้ในโลกใหม่ในรูปของตัวเลข 2 ตัว คือ 0 กับ 1 หรือที่เรียก ข้อมูลดิจิทัล (digital data) หากมนุษย์ไม่ได้วางรากฐานวิธีการนับจำนวน/ปริมาณข้อมูลและสารสนเทศไว้ให้คนยุคสังคมสารสนเทศนำไปพัฒนาต่อยอดใช้กับข้อมูลขนาดใหญ่คงเกิดปัญหาความไม่รู้ว่ปัจจุบันมีข้อมูลและ

[†] จากคำถามของเดเมทริอุสแห่งฟาเลอร์ม (Demetrius of Phalerum) ลูกศิษย์ของอริสโตเติลที่ตั้งคำถามว่า ห้องสมุดแห่งอเล็กซานเดรียมีหนังสือกี่พันเล่ม (Hilbert, 2012)

[‡] เริ่มปรากฏอยู่ในพจนานุกรมภาษาอังกฤษออกซฟอร์ด (Oxford English Dictionary: OED) ค.ศ. 1941 ว่า “a rapid increase in the amount of information available, (now) esp. as a result of the increased use, availability, and sophistication of information technology” (Forbes, 2013; OED, 2015)

สารสนเทศมีจำนวน/ปริมาณเท่าไร เพราะตั้งแต่โลกใหม่ก่อตัวขึ้น ค.ศ. 1969 ใน ค.ศ. 1986 มีข้อมูลแอนะล็อกประมาณร้อยละ 99.2 ค.ศ. 2002 ข้อมูลดิจิทัลเริ่มมีมากกว่าข้อมูลแอนะล็อก และ ค.ศ. 2007 มีข้อมูลดิจิทัลประมาณร้อยละ 94 (Hilbert & Lopez, 2011) การศึกษาของไอซีดี (International Data Corporation: IDC) พบว่า ค.ศ. 2013 มีข้อมูลอยู่ในโลกใหม่ประมาณ 4.4 เซตตะไบต์ (Zettabyte: ZB) หรือ 4.4 พันล้านล้านล้านตัวอักษร มีการเพิ่มขึ้นของข้อมูล 2 เท่าในทุกๆ 2 ปี และคาดการณ์ว่า ค.ศ. 2020 จะมีข้อมูลประมาณ 44 เซตตะไบต์ หรือ 44 พันล้านล้านล้านตัวอักษร (IDC, 2014) ซึ่งเป็นข้อมูลที่มาจากแหล่งต่างๆ ดังนี้ (Cheung, 2011; Mrazek, Biglan & Hawkins, n.d.)

1. ข้อมูลด้านสังคม (Social everything) ได้แก่ เครือข่ายการติดต่อ (networking) การแสดงความคิดเห็น (commenting)
2. ข้อมูลด้านการใช้อินเทอร์เน็ต (Internet users) ได้แก่ การค้นหาแบบออนไลน์ (online searches) การเข้ามาดูหน้าเว็บแบบออนไลน์ (online page-view)
3. ข้อมูลด้านการบริหาร (Administrative) ได้แก่ ผู้มาโรงพยาบาล (hospital visits) ใบเสร็จรับเงินจากการขาย (sales receipts) การเฝ้าติดตามการจราจร (traffic monitoring)
4. ข้อมูลด้านการค้า (Commercial) ได้แก่ การใช้โทรศัพท์มือถือ (cell phone usages) การทำธุรกรรมด้วยบัตรเครดิต (credit card transactions) ข้อมูลการทำประกัน (insurance records) การค้นหาสินค้า (product searches)
5. ข้อมูลด้านสุขภาพ (Health information) ได้แก่ บันทึกข้อมูลทางการแพทย์แบบอิเล็กทรอนิกส์ (electronic medical records) การเฝ้าติดตามทางการแพทย์ (medical monitoring)
6. ข้อมูลด้านดาวเทียม (Satellite imagery) ได้แก่ การตั้งถิ่นฐานของมนุษย์ (human settlements) พื้นดิน (land surface) การเปลี่ยนแปลงของระบบนิเวศ (ecosystem change)
7. ข้อมูลด้านเฝ้าติดตาม (Monitoring systems) ได้แก่ สุขภาพทางกาย (physical health) สุขภาพทางใจ (mental health) สุขภาพด้านพฤติกรรม (behavioral health) การศึกษา (education)

ในช่วง 10 ปีที่ผ่านมา ข้อมูลพฤติกรรมของมนุษย์ (human behavior) ที่ได้จากเว็บ (web) อุปกรณ์เคลื่อนที่ (mobile device) อุปกรณ์ตรวจจับ (sensors) การทำธุรกรรม (transaction) กิจกรรมเครือข่ายทางสังคม (social network activity) และการทำให้เป็นดิจิทัล (digitization) จากข้อมูลที่บันทึกไว้ในสื่อแบบเก่า ถูกสร้างและแปลงรูปเป็นข้อมูลทางสังคมขนาดใหญ่ (socially-generated big data) สามารถใช้เครื่องจักรสกัดออกมา (extracted) ได้ง่ายและรวดเร็วกว่าความสามารถของมนุษย์/ข้อมูลที่บันทึกไว้ในสื่อแบบเก่า ทำให้นักสังคมศาสตร์สามารถทดสอบสมมติฐานและศึกษาปรากฏการณ์ต่างๆ ได้อย่างไม่เคยทำมาก่อน (Pennsylvania State University, n.d.; Stony Brook University, n.d.) ดังตัวอย่างจากแหล่งข้อมูลขนาดใหญ่ของภาครัฐและเอกชนที่นำมาเผยแพร่ให้บริการฟรี ดังนี้ (Marr, 2014)

1. ข้อมูลขนาดใหญ่ของภาครัฐ เช่น Data.gov, Data.gov.uk, European Union Open Data Portal และ Health and Social Care Information Centre เป็นต้น

2. ข้อมูลขนาดใหญ่ของภาคเอกชน เช่น Amazon Web Services, Facebook Graph, Google Trends และ New York Times

ข้อมูลขนาดใหญ่หรือจำนวนมหาศาลดังที่กล่าวมามีทั้งข้อมูลที่เกิดจากโลกเก่าที่ทำการแปลงให้เป็นดิจิทัล (H2M/digitization) และเกิดจากโลกใหม่ที่เป็นข้อมูลดิจิทัลโดยกำเนิด (M2M/born digital) เข้ามาอยู่รวมกันในไซเบอร์สเปซด้วย 3 กระบวนการ หรือที่เรียกว่า การทำให้เป็นข้อมูลคอมพิวเตอร์ (datafication) ดังนี้ (UNECE Statistics Wikis, 2013; Doorn, 2014)

1. ข้อมูลที่เกิดจากกระบวนการของตัวกลางหรือข้อมูลด้านการบริหาร (Process-mediated data/Administrative data) เป็นข้อมูลจากการทำงานในระบบเก่า (traditional business system) ที่เกิดจากการบันทึก (record) การติดตาม (monitor) เหตุการณ์/กิจกรรมทางด้านสังคม การเมือง เศรษฐกิจ ของภาครัฐ/เอกชน/ประชาชน เช่น การติดต่อระหว่างกัน การเลือกตั้ง การซื้อ/ขาย/บริการ

2. สารสนเทศของมนุษย์ (Human-sourced information) เป็นเครือข่ายทางสังคมในรูปแบบต่าง (social networks) ที่ได้มาจากการบันทึกประสบการณ์ของมนุษย์ สิ่งที่เป็นบันทึกไว้ในหนังสือ งานด้านศิลปะ ภาพถ่าย/เสียง/วิดีโอในอดีต

3. ข้อมูลที่เกิดจากการสร้างของเครื่องจักร (Machine-generated data) เป็นข้อมูลจากทุกสิ่งที่เป็นอินเทอร์เน็ต (internet of things: IoT) ที่ได้มาจากการใช้การตรวจจับ (sensors) และเครื่องจักร (machine) เพื่อทำการวัดและบันทึกเหตุการณ์หรือสถานการณ์ในโลกมนุษย์/โลกทางกายภาพ

ข้อมูลในโลกใหม่มีความไม่แน่นอน (uncertainty) ความไม่สอดคล้องกัน (inconsistency) ความไม่สมบูรณ์ (incompleteness) ความคลุมเครือ (ambiguities) ความซ่อนเร้น (latency) ความไม่จริงใจ (deception) และเป็นสิ่งจำลอง (model) หรือมีลักษณะ 7Vs คือ เกิดอย่างรวดเร็ว (velocity) มีรูปแบบที่หลากหลาย (variety) มีทั้งที่เป็นจริงและไม่จริง (veracity) มีความผันผวน (volatility) มีความแปรผัน (variability) มีความหนืด (viscosity) และมีจำนวน (volume) มาก (Desouza & Smith, 2014) ข้อมูลในโลกใหม่มีหลายรูปแบบมีทั้งที่เป็นเสียง ภาพ วิดีโอ ตัวเลข และตัวอักษร แต่ทั้งหมดนี้จำแนกได้เป็น 2 รูปแบบ (format) ดังนี้ (Mewawalla, 2012; Lam, 2014)

1. ข้อมูลที่มีโครงสร้าง (structured data) ได้แก่ ไฟล์แบบตัวอักษร (text file) ตารางคำนวณ (spreadsheet) ฐานข้อมูล (database) ไฟล์แบบเอ็กเซล/XML ข้อมูลจากอุปกรณ์ตรวจจับ (sensor data)

2. ข้อมูลที่ไม่มีโครงสร้าง (unstructured data) ได้แก่ เสียง (voice) ภาพยนตร์ (movie) เอกสาร (document) จดหมายอิเล็กทรอนิกส์ (e-mail) สื่อสังคม (social media) รูปภาพ (image) ภาพถ่ายจากดาวเทียม (satellite imagery)

ห่วงโซ่ในการพัฒนาเทคโนโลยีของมนุษย์ในแต่ละยุคมีผลกระทบต่อการเปลี่ยนวิธีการค้นหาความรู้ของมนุษย์ด้วย กล่าวคือ หลังการปฏิวัติอุตสาหกรรม มนุษย์มีเครื่องมือที่มีประสิทธิภาพในการทำให้มนุษย์ใช้ร่างกาย (human body) อย่างมีประสิทธิภาพในการค้นหาความรู้ แต่หลังจากการปฏิวัติดิจิทัล เทคโนโลยีที่เกิดมาจากการปฏิวัติอุตสาหกรรมได้รับการพัฒนาให้มีความสมรรถนะมากขึ้น ทำให้มนุษย์ใช้สมอง (human brain) ในการคำนวณ (computing) การวิเคราะห์ (analytic) และการสร้างภาพข้อมูล (visualization) ในกระบวนการ

ค้นหาความรู้ได้ดีกว่าในอดีต (Lam, 2015) รวมถึงลดการใช้งานความสามารถของกล้ามเนื้อในการค้นหาความรู้ เช่น การเดินทางลงพื้นที่สัมภาษณ์และสำรวจข้อมูล การหันมาใช้ความสามารถของเทคโนโลยีสารสนเทศและการสื่อสารช่วยรวบรวมข้อมูล ทำให้นักวิจัยได้ข้อมูลจากหลายแหล่งมากขึ้นและไกลมากขึ้น

การหาความรู้ในโลกใหม่

เคยมีคำถามว่า โลกใหม่หรือไซเบอร์สเปซ เป็นเครือข่ายของคอมพิวเตอร์ที่เชื่อมต่อกันอยู่ทุกที่หรือเป็นปรากฏการณ์ (phenomena) ที่เกิดจากการประยุกต์ใช้งานระหว่างหรือภายในเครือข่ายคอมพิวเตอร์และการติดต่อกันผ่านระบบเครือข่ายคอมพิวเตอร์ที่เรียกว่า พื้นที่เสมือนจริง (virtual) และมีข้อสันนิษฐานว่าเป็นสิ่งที่ไม่มีตัวตน (intangible) เป็นสิ่งที่ถูกมองว่ามีลักษณะพิเศษ ความเข้าใจที่ไม่ถูกต้องเกี่ยวกับโลกใหม่จะเป็นตัวนำไปสู่ความสับสนเชิงภววิทยา (ontology) แต่นักปรัชญาส่วนใหญ่ต่างยอมรับกันอย่างกว้างขวางว่าภววิทยาในการศึกษาโลกใหม่มีความหลากหลาย กล่าวคือ มีทั้งสิ่งที่มีอยู่จริง (existence) สิ่งที่มีชีวิต (being) ความเป็นจริง (reality) และ/หรือแต่ละสิ่งที่มีอยู่ในโลกนี้ (things) ดังนั้น ภววิทยาของโลกใหม่จึงเป็นทุกสิ่งที่เป็นองค์ประกอบที่ถูกนำมาประยุกต์ใช้และทำให้เกิดปรากฏการณ์ต่างๆ ในโลกใหม่ เช่น ซอฟต์แวร์ อินเทอร์เน็ต จดหมายอิเล็กทรอนิกส์ เครื่องคอมพิวเตอร์ สื่อสังคมออนไลน์ ความจริงเสมือน (virtual reality: VR) ทุกสิ่งที่เป็นอินเทอร์เน็ต (IoT) และการประมวลผลแบบกลุ่มเมฆ เป็นต้น (Koepsell, 2000)

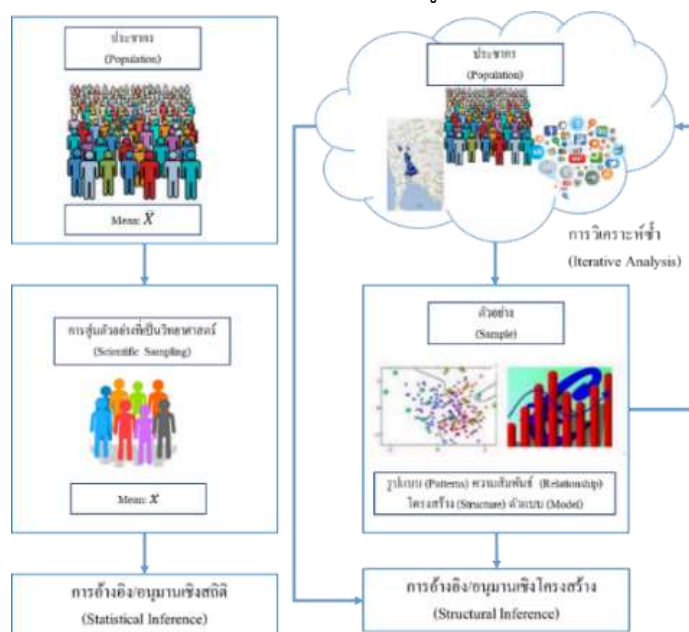
วิธีการหาความรู้ (epistemology) ในอดีตของนักวิจัยต้องลงพื้นที่รวบรวมข้อมูลปฐมภูมิ (primary data) ข้อมูลทุติยภูมิ (secondary data) และข้อมูลตติยภูมิ (tertiary data) เพื่อนำมาใช้ในการสร้างและทดสอบทฤษฎี การรวบรวมข้อมูลบางอย่างนักวิจัยไม่ต้องสร้างความสัมพันธ์ทางสังคมกับผู้ให้ข้อมูล แต่มีข้อมูลจำนวนมากที่นักวิจัยต้องใช้ความสัมพันธ์ทางสังคมเพื่อให้ได้มาซึ่งข้อมูล ความง่ายและความยากของการได้มาซึ่งข้อมูลยังมีความเกี่ยวข้องกับประเภทของสังคม กล่าวคือ คนในสังคมชนบทสามารถสร้างความสัมพันธ์ทางสังคมได้ง่ายกว่าคนในสังคมชนเมือง และสังคมเมือง ในปี ค.ศ. 2014 มีมนุษย์อาศัยอยู่ในเมืองประมาณร้อยละ 54 และคาดการณ์ว่าใน ค.ศ. 2050

จะมีคนอาศัยอยู่ในเมืองประมาณร้อยละ 66 (United Nations, 2014) ซึ่งสัมพันธ์กับการพัฒนาและการเปลี่ยนแปลงด้านสังคม เศรษฐกิจ การเมือง และพฤติกรรมของมนุษย์ที่มีการนำเอาเทคโนโลยีดิจิทัลที่เชื่อมโยงกับอินเทอร์เน็ต (IoT) มาใช้เพิ่มมากขึ้น ปรากฏการณ์ความเป็นเมืองและความเป็นดิจิทัลดังกล่าวย่อมมีผลต่อความยากลำบากในการเก็บรวบรวมข้อมูลในการหาความรู้ของนักวิจัยมากขึ้น ประกอบกับข้อมูลที่อยู่บนโลกใหม่เป็นปรากฏการณ์ที่มีสภาพแวดล้อมที่แตกต่างไปจากปรากฏการณ์บนโลกเก่าหรือโลกมนุษย์ และน่าจะทำให้นักวิชาการได้ประสบการณ์และความรู้ใหม่ที่เหมือนหรือไม่เหมือนกับความรูที่มีอยู่ในอดีต

การค้นหาความรู้ (knowledge discovery) มีขั้นตอนหลัก คือ การรวบรวมข้อมูล (gathering data) การทำให้ข้อมูลอยู่ในรูปแบบที่สามารถใช้วิเคราะห์ได้ (usable format) การรวมเข้า (consolidating) เป็นข้อมูลชุดเดียวกัน (dataset) ทำการวิเคราะห์ (analyzing) และการตีความ (interpretation) (Waterman & Bruening, 2014) ซึ่งการวิจัยแบบดั้งเดิม (ในโลกเก่า) ส่วนใหญ่เริ่มจากการกำหนดปัญหาการวิจัย (research problems) ทำการรวบรวมข้อมูล (collection data) ด้วยการกำหนดขอบเขตประชากร (population frame)

เลือกตัวอย่าง (sampling) เพื่อเป็นตัวแทนในการให้ข้อมูล จากนั้นจึงออกแบบการเก็บข้อมูลเป็นข้อมูลเชิงปริมาณ (quantitative) หรือเชิงคุณภาพ (qualitative) หรือเชิงผสมผสาน (mixed-methods) จากประสบการณ์หรือประสาทสัมผัส คือ การได้เห็นและการได้ยินของนักวิจัยในการเข้าถึงและรวบรวมข้อมูลและทำการวิเคราะห์ข้อมูลด้วยตัวนักวิจัยเอง ส่วนการวิจัยเชิงปริมาณทางสังคมศาสตร์แนวใหม่ (ในโลกใหม่) ที่ใช้ข้อมูลจากโลกใหม่มีกระบวนการที่ไม่แตกต่างจากการวิจัยแบบดั้งเดิมมากนัก เพราะมีกระบวนการที่สำคัญ คือ เริ่มต้นจากการกำหนดปัญหาการวิจัย รวบรวมข้อมูลจากการเข้าถึงข้อมูล (data access) หรือรวบรวมข้อมูล (collect data) ที่มีรูปแบบที่หลากหลายจากชุดข้อมูลเดียวหรือหลายชุดข้อมูล (single or multiple datasets) เพื่อนำมาวิเคราะห์ด้วยคอมพิวเตอร์ที่มีสมรรถนะสูง (high performance computing: HPC) เช่น ระบบประมวลผลแบบกลุ่มเมฆ เพื่อหารูปแบบ (pattern) ความสัมพันธ์ (relationship) โครงสร้าง (structure) และตัวแบบ (model) และทำการวิเคราะห์ซ้ำ (iterative analysis) กับกลุ่มประชากร ก่อนนำไปตีความเพื่อตอบคำถามการวิจัย (Cheung, 2011; Rovai, 2013; Lam, 2014) โดยมีกระบวนการดังภาพที่ 1

ภาพที่ 1 แสดงกระบวนการวิเคราะห์ข้อมูลแนวเก่ากับแนวใหม่



ปรับปรุงมาจาก: Cheung, 2011

โดยภาพรวมจากกระบวนการวิจัย การวิจัยด้านสังคมศาสตร์แนวใหม่มีกระบวนการค้นหาความรู้ไม่แตกต่างไปจากการค้นหาความรู้แบบดั้งเดิมมากนัก แต่มีแนวโน้มที่จะเกิดการเปลี่ยนรูปแบบในการดำเนินการวิจัยในบางขั้น เช่น การใช้ข้อมูลหรือทำข้อมูลให้เป็นข้อมูลคอมพิวเตอร์ที่อยู่ในรูปของข้อมูลดิจิทัลมากขึ้น ใช้ข้อมูลขนาดใหญ่บนอินเทอร์เน็ตมากขึ้น มีข้อมูลแบบไร้โครงสร้างมากกว่ามีโครงสร้าง ใช้เทคโนโลยีสารสนเทศและการสื่อสารหรือเครื่องจักรในกระบวนการรวบรวมข้อมูลและการวิเคราะห์ข้อมูลมากขึ้น ใช้ที่เก็บข้อมูลขนาดใหญ่มากกว่าขนาดเล็ก มีการนำข้อมูลมาเก็บไว้ในบึงข้อมูล (data lake)⁵ หรือใช้การกรองข้อมูลจากข้อมูลขนาดใหญ่

⁵ เทคโนโลยีเก็บข้อมูลขนาดใหญ่ที่สามารถเก็บข้อมูลดิบได้หลายประเภทหรือหลายรูปแบบ

มาเก็บไว้ในฐานข้อมูล (database) หรืออาจใช้ข้อมูลจากเหมืองข้อมูล (data mining) ขององค์การที่มีอยู่แล้ว มาทำการวิเคราะห์หาความรู้ในเชิงทิศทาง (trend) ของสิ่งต่างๆ ที่เกิดขึ้น

การวิเคราะห์ข้อมูลในโลกใหม่

ความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ทำให้เกิดความก้าวหน้าในการค้นหาความรู้เพิ่มมากขึ้น ปัจจุบันนักวิจัยสามารถใช้โปรแกรมคอมพิวเตอร์วิเคราะห์ข้อมูลขนาดใหญ่ให้ได้ผลลัพธ์มากกว่าการเข้าใจ (hindsight) ปัญหาหลังจากปรากฏการณ์นั้นเกิดขึ้นแล้วว่า เกิดอะไรขึ้น (what happened?) หรือทำไมถึงเกิดขึ้น (why it happened?) โดยนำข้อมูลเข้าสู่กระบวนการเรียนรู้ของเครื่องจักร (machine learning) เพื่อค้นหาความรู้ที่ลึกซึ้งหรือเป็นแก่นแท้ (insight) ของปรากฏการณ์ทำให้ทราบว่าเกิดเหตุการณ์อะไร (what will happened?) ทำให้ได้ความรู้ที่สามารถนำไปใช้ในอนาคตได้อย่างมีเหตุผล (foresight) นักวิจัยหรือ ผู้ที่เกี่ยวข้อง สามารถตัดสินใจหรือจะทำการอย่างไรกับเหตุการณ์ที่จะเกิดขึ้น (how can we make it happen?) ได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม ข้อมูลขนาดใหญ่เพียงอย่างเดียวไม่สามารถนำไปใช้ตอบปัญหาการวิจัยหรือแก้ไข ปัญหาของสังคมได้ทันที แต่ข้อมูลขนาดใหญ่มีประโยชน์มากเพราะสามารถนำไปช่วยตอบคำถามการวิจัยและ แก้ไขปัญหาสังคมได้ และข้อมูลขนาดใหญ่โดยตัวมันเองก็ไม่เหมาะที่นักวิจัยจะนำมาใช้ (consumption) เลยทันที แต่สิ่งที่เหมาะกับการนำมาใช้ในการวิจัย คือ สารสนเทศ (information) ที่ได้มาจากข้อมูล (Wu, 2013; Grimmer, 2015) เพราะข้อมูลขนาดใหญ่มีทั้งข้อมูลที่เป็นจัดเก็บไว้เป็นไฟล์ (file) ตาราง (table) และจัดเก็บไว้ในฐานข้อมูล (in database) มีทั้งข้อมูลแบบใกล้เคียงเวลาจริง (near real-time) และข้อมูลใกล้เคียงเวลาจริง (real-time) เป็นข้อมูลแบบมีโครงสร้าง กึ่งมีโครงสร้าง และไม่มีโครงสร้าง การนำข้อมูลดิบ (raw data) หรือ ข้อมูลขนาดใหญ่มาใช้ประโยชน์ต้องทำการลดทอนข้อมูล (data reduction) ให้เป็นข้อมูลที่มีขนาดเล็กกว่า และนำไปวิเคราะห์ด้วยวิธีการต่างๆ ดังนี้ (Wu, 2013)

การวิเคราะห์เชิงพรรณนา (Descriptive Analytics) ใช้สถิติพื้นฐาน (descriptive statistics) จาก คำสั่งคำนวณค่าผลรวม (aggregate functions) ได้แก่ นับ (counts) รวม (sums) เฉลี่ย (averages) ร้อยละ (percentages) ต่ำสุด (min) สูงสุด (max) และการคำนวณทางคณิตศาสตร์ เช่น $+$ $-$ $\times$ $\div$ ที่มีอยู่ในโปรแกรม ฐาน เพื่อสรุปข้อมูล (summarize) และบอกว่าเกิดอะไรขึ้นในอดีต (what happened in the past) เช่น การเขียนข้อความ (post) การแสดงความคิดเห็น (mentions) กลุ่มเพื่อน (fans) ผู้ติดตาม (followers) การเข้ามาดูหน้าเว็บ (page views) การชื่นชม (kudos) การให้คะแนนบวกหนึ่ง (+1s), การแสดงตนเพื่อบอกตำแหน่ง (check-ins) และการปักหมุดสิ่งที่สนใจ (pins)

1. การวิเคราะห์เชิงทำนาย (Predictive Analytics) ใช้ข้อมูลที่มีอยู่ (data we have) สร้างตัวแบบ (model) และใช้ตัวแบบทำนายข้อมูลที่ไม่มีอยู่ (data we don't have) ได้แก่ ตัวแบบการทำนาย (predictive modeling)^{**}, การเรียนรู้ของเครื่อง (machine learning)⁺⁺ และเหมืองข้อมูล (data mining)⁺⁺

^{**} วิธีการจัดกลุ่มจากการควบคุมข้อมูล (Group method of data handling) วิธีการของเบย์อย่างง่าย (Naive Bayes) วิธีการข้างเคียงที่มีค่าใกล้เคียงมากที่สุด (k-nearest neighbor algorithm) ตัวจำแนกหมู่มาก (Majority classifier) กลไกการสนับสนุนแบบเส้นสมมติ (Support vector machines) การตัดป่าแบบสุ่ม (Random forests) การเสริมกำลังต้นไม้ (Boosted trees) ต้นไม้แบบถดถอยและแบ่งประเภท (Classification and Regression Trees: CART) การวิเคราะห์เส้นโค้งด้วยวิธีถดถอยแบบปรับตัวเชิงพหุ (Multivariate adaptive

เพื่อคาดการณ์ (forecast) ว่าอะไรอาจจะเกิดขึ้นในอนาคต (what might happen in the future) หรือเป็นความน่าจะเป็นในธรรมชาติ (probabilistic in nature) ไม่ผูกติดกับมิติของเวลา (not limited to the time domain) เช่น การวิเคราะห์สื่อสังคมออนไลน์เป็นผลการวิเคราะห์ที่ไม่ได้เป็นไปตามเวลาตามธรรมชาติ (non-temporal in nature)

2. การวิเคราะห์เพื่อเสนอทางเลือก (Prescriptive Analytics) ใช้ข้อมูลที่มีอยู่ (existing data) ข้อมูลจากการกระทำ (action data) และข้อมูลย้อนกลับ (feedback data) มาสร้างตัวแบบและแนะนำการตัดสินใจจากผลลัพธ์ ตัวแบบที่ได้จากการวิเคราะห์เป็นตัวแบบการทำนายแบบพหุ (multiple predictive models) แต่ละตัวแบบเสนอทางเลือกสามารถทำนายความเป็นไปได้ของผลที่ตามมา (possible consequences) บนฐานของความแตกต่างในตัวเลือกของการกระทำโดยมีคำแนะนำด้วยเหตุผลที่ดีที่สุดของการกระทำที่ระบุไว้ก่อนในทุกๆ ผลลัพธ์สำหรับใช้ในการประกอบการตัดสินใจเลือกเพื่อให้ได้ผลลัพธ์สุดท้ายตามเป้าหมาย การสร้างตัวแบบการทำนายเชิงเสนอทางเลือกต้องมียุคประกอบในการทำนายมากกว่า 2 องค์ประกอบ คือ

2.1 ความสามารถในการนำไปปฏิบัติได้ (Actionable): ผู้บริโภคข้อมูลต้องสามารถนำไปใช้งานได้บนฐานของการทำนายผลลัพธ์ของตัวแบบ

2.2 ระบบย้อนกลับ (Feedback System): ตัวแบบต้องมีระบบย้อนกลับที่มีแนวทางในการปรับเปลี่ยนผลลัพธ์บนฐานของนำไปปฏิบัติได้ หมายความว่าตัวแบบการทำนายต้องมีความฉลาดพอที่จะเรียนรู้ความสัมพันธ์ที่ซับซ้อนระหว่างการกระทำของผู้ใช้และการปรับแต่งผลลัพธ์จากข้อมูลย้อนกลับ

การวิเคราะห์ข้อมูลขนาดใหญ่ตามที่กล่าวมามีการใช้ข้อมูลที่แตกต่างกัน กล่าวคือ การวิเคราะห์เชิงพรรณนาสามารถใช้ข้อมูลทั้งหมดที่เราต้องการ ส่วนการวิเคราะห์เชิงทำนายต้องมีข้อมูลอย่างเพียงพอสำหรับการสร้างตัวแบบและตรวจสอบ (validate) ตัวแบบในการทำนาย และการวิเคราะห์เชิงเสนอทางเลือกนอกจากมีข้อมูลในการตรวจสอบตัวแบบในการทำนายแล้ว ยังต้องมีข้อมูลย้อนกลับสำหรับนำมาใช้กับแต่ละประเภทของการกระทำ/ทางเลือก แนวทางการนำการวิเคราะห์ที่กล่าวมาไปประยุกต์ใช้ในการวิจัย คือ การวิเคราะห์เชิงพรรณนาเหมาะจะนำไปใช้ในการวิเคราะห์ด้านสังคม (social analytics) จากสื่อสังคมออนไลน์ การวิเคราะห์

regression splines: MARS) เครือข่ายประสาท (Neural Networks), ACE and AVAS, การประมาณค่าแบบกำลังสองน้อยที่สุด (Ordinary Least Square) ตัวแบบเชิงเส้นทั่วไป (Generalized Linear Models: GLM) การถดถอยแบบโลจิสติก (Logistic regression) ตัวแบบการเพิ่มทั่วไป (Generalized additive models) การถดถอยแบบแกร่ง (Robust regression) และการถดถอยแบบกึ่งพารามิเตอร์ (Semiparametric regression)

^{††} การเรียนรู้แบบแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision tree learning) การเรียนรู้ตามกฎความสัมพันธ์ (Association rule learning) เครือข่ายประสาทเทียม (Artificial neural networks) การโปรแกรมตรรกะเชิงอุปนัย (Inductive logic programming) กลไกการสนับสนุนแบบเส้นสมมุติ (Support vector machines) การจัดกลุ่ม (Clustering) เครือข่ายแบบเบย์ (Bayesian networks) การเรียนรู้แบบเพิ่มประสิทธิภาพ (Reinforcement learning) การเรียนรู้แบบตัวแทน (Representation learning) การเรียนรู้แบบวัดความคล้ายคลึง (Similarity and metric learning) การเรียนรู้แบบประมวลคำศัพท์กระจาย (Sparse dictionary learning) และขั้นตอนเชิงพันธุกรรม (Genetic algorithms)

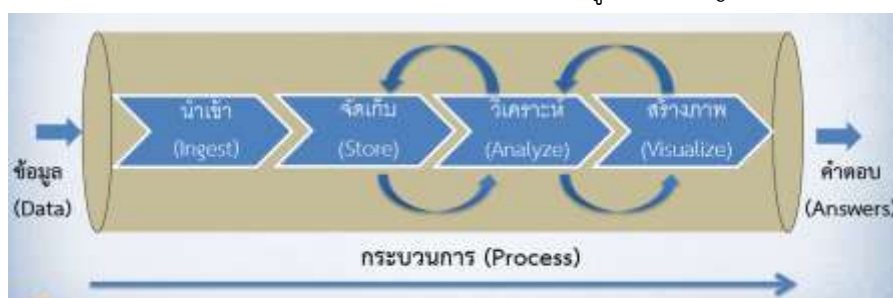
^{††} การตรวจสอบความผิดปกติ (Anomaly detection) ได้แก่ ค่าโด่ง (Outlier) ค่าความแปรปรวน (deviation) ค่าการเปลี่ยนแปลง (change) การเรียนรู้ตามกฎความสัมพันธ์ (Association rule learning) หรือตัวแบบการพึ่งพา (Dependency modelling) การจัดกลุ่ม (Clustering) การแบ่งกลุ่ม (Classification) การถดถอย (Regression) และการสรุป (Summarization)

เชิงทำนายเหมาะกับการวิเคราะห์ทิศทางเชิงเส้น (trend line) คะแนนอิทธิพลทางสังคม (social influence scoring) และการวิเคราะห์อารมณ์และความรู้สึก (sentiment analysis) และการวิเคราะห์เชิงเสนอทางเลือก เหมาะกับนำไปใช้ในการวิเคราะห์เพื่อประกอบการด้านการตัดสินใจ เช่น การบริหารองค์การ การบริหารนโยบาย และการตัดสินใจยอมรับหรือปฏิเสธสมมติฐานการวิจัย เป็นต้น

ปัญหาและความท้าทายของการวิจัยในโลกใหม่

การวิจัยในโลกใหม่ นักวิจัยด้านสังคมศาสตร์ที่นิยมใช้วิธีการเชิงปริมาณในการตอบข้อสงสัยหรือต้องการหาความรู้ใหม่ที่ต้องเผชิญกับความท้าทายหลายด้าน เช่น การทำความเข้าใจรูปแบบข้อมูลใหม่ การค้นหาข้อมูลแหล่งใหม่ วิธีการรวบรวมข้อมูลด้วยเครื่องมือและเทคโนโลยีใหม่ การใช้ระบบจัดเก็บข้อมูลและประมวลผลข้อมูลแบบใหม่ และการสร้างภาพข้อมูลในการนำเสนอใหม่ เพราะประเด็นความท้าทายที่กล่าวมามีความเกี่ยวข้องกับข้อมูลขนาดใหญ่ที่เป็นดิจิทัลและต้องใช้เทคโนโลยีการประมวลผลแบบกลุ่มเมฆในแต่ละกระบวนการ ดังภาพที่ 2

ภาพที่ 2 แสดงกระบวนการวิเคราะห์ข้อมูลขนาดใหญ่



ที่มา: Amazon, 2015

การใช้ข้อมูลขนาดใหญ่ไม่เพียงแต่จะมีปัญหาด้านปริมาณและโครงสร้างของข้อมูลเท่านั้น แม้ว่าจะมีข้อมูลที่เผยแพร่ให้เข้าถึงได้ทั่วไป (public data) แต่ยังมีข้อมูลอีกจำนวนมากที่เป็นข้อมูลที่ไม่เผยแพร่ให้นำไปใช้ทั่วไป (private data) ด้วยเหตุผลของความเป็นส่วนตัว (privacy) และความมั่นคง (security) เช่น ข้อมูลส่วนบุคคล ข้อมูลขององค์การ เป็นต้น รวมถึงข้อมูลของบางองค์การอาจเป็นข้อมูลที่ไม่มีความน่าเชื่อถือ หรือข้อมูลที่ได้มาอาจเป็นข้อมูลที่ไม่มีความถูกต้องหรือไม่มีความมาตรฐาน เป็นข้อมูลที่ไม่ถูกกฎหมายหรือทุจริต และเป็นข้อมูลที่มีความหมายไม่เหมือนกัน (Desouza & Smith, 2014; Waterman & Bruening, 2014) จึงทำให้มีคำถามเกิดขึ้นในการค้นหาความรู้ทางสังคมศาสตร์จากโลกใหม่หลายคำถามที่ทำให้นักวิจัยต้องเตรียมตอบข้อสงสัย ดังนี้ (Cheung, 2011)

1. ด้านประชากรและกลุ่มตัวอย่าง (population and sample) อะไรคือ ประชากร? เราสามารถสรุปข้อมูลจากตัวอย่างแล้วอ้างอิง (infer) ถึงลักษณะ (traits) ทางประชากรได้หรือไม่?
2. ด้านข้อมูล (data) ใครเป็นคนสร้างข้อมูล? สามารถตรวจสอบ (verified) และทดสอบซ้ำ (re-tested) ได้หรือไม่? ข้อมูลมีลักษณะสุ่ม (random) หรือไม่? อะไรคือความหมายที่แท้จริงของข้อมูล? ข้อมูลสามารถสะท้อนรูปแบบ (pattern) และโครงสร้าง (structure) ของสังคมได้หรือไม่?

3. ด้านการวิเคราะห์ (analytics) รูปแบบจากการวิเคราะห์อาจสะท้อนสิ่งที่กำลังเกิดขึ้นได้แต่การอ้างอิงถึงประชากร (reference population) อาจไม่ชัดเจนข้อมูลขนาดใหญ่ส่วนมากเป็นข้อมูลส่วนตัว (private) ชุมชนวิชาการสามารถเข้าถึงได้อย่างเท่าเทียมกันหรือไม่? สามารถนำไปทำการวิเคราะห์ซ้ำ (repeated analysis) ได้หรือไม่? การป้องกันการหลอกลวงจะอย่างไร?

4. ด้านสถิติ (statistics) ไม่สามารถใช้สถิติเชิงอ้างอิง (inferential statistics) ได้ แต่ใช้สถิติแบบไร้พารามิเตอร์ (non-parametric) ได้

ดูเหมือนว่าปัญหาใหญ่ของข้อมูลในโลกใหม่ คือ ความน่าเชื่อถือ แต่ปัญหานี้ไม่ใช่เรื่องใหม่ เพราะในกระบวนการหาความรู้ของมนุษย์ก็เผชิญปัญหานี้มาอย่างต่อเนื่อง การรวบรวมข้อมูลบนโลกเก่าเพื่อสร้างความรู้ใหม่ส่วนใหญ่ก็เริ่มจากข้อมูลจำนวนไม่มากจากพื้นที่เล็กๆ ภายใต้สภาพแวดล้อมหนึ่ง หลังจากนั้นจึงหาข้อมูลจากหลายแห่งมายืนยันข้อสรุป⁵⁵ หัวใจที่สำคัญในการใช้ข้อมูลขนาดใหญ่ คือ การไม่อนุมาน/อ้างอิง (inference) ไปสู่ประชากร (population) ขนาดใหญ่และสร้างข้อสรุปทั่วไป (generalization) ใช้ความสามารถของเทคโนโลยีคอมพิวเตอร์กำหนดขอบเขตของประชากร เลือกข้อมูลแบบสุ่ม และทำให้ข้อมูลที่มีโอกาสในการเลือกเท่าๆ กัน ปัญหาที่กล่าวมาอาจไม่เกิดขึ้นเลยหากปรากฏการณ์ (phenomenon) ที่เกิดจากการกระทำของมนุษย์ (actor) บนโลกใหม่ไม่มีข้อสงสัยหรือคำตอบที่ต้องค้นหา เมื่อปรากฏการณ์ในสังคมเสมือนจริงเป็นข้อมูลดิจิทัลขนาดใหญ่จึงทำให้เกิดช่องว่างในการหาความรู้ระหว่างศาสตร์ นักวิจัยด้านสังคมศาสตร์ไม่มีความเข้าใจและมีความสามารถในการนำข้อมูลบนโลกใหม่มาใช้ในการตอบข้อสงสัยหรือปรากฏการณ์ทางสังคมที่เกิดขึ้นบนโลกใหม่อย่างมีประสิทธิภาพ นักวิจัยด้านคอมพิวเตอร์และสถิติไม่มีความเข้าใจและมีมุมมองด้านสังคมบนโลกใหม่อย่างเพียงพอในการกำหนดโจทย์หรือปัญหาการวิจัย การหาความรู้จากข้อมูลเชิงปริมาณด้านสังคมศาสตร์บนโลกใหม่จึงต้องอาศัยทีมข้อมูลศาสตร์ (data science team) ทำงาน/วิจัยร่วมกัน

ผลกระทบของการวิจัยจากโลกใหม่

การแพร่กระจายและแทรกซึมของอินเทอร์เน็ตเข้าไปในทุกอณูของสังคมทั่วโลก ส่งผลให้ประชาชนจากทุกสถาบันทางสังคมสร้างข้อมูลเข้าไปเก็บไว้ในโลกเสมือนจริง นักคอมพิวเตอร์และนักสถิติที่มีความเชี่ยวชาญในการเข้าถึงและวิเคราะห์ข้อมูลไม่สามารถอธิบายหรือตีความผลการวิเคราะห์ที่ได้ให้เข้าใจอย่างลึกซึ้ง เพราะข้อมูลแต่ละชุดมีเนื้อแท้และความหมายเฉพาะ ต้องอาศัยผู้ที่มีความรู้ในแต่ละสาขาวิชา (domain knowledge) มาร่วมกับทีมงานนักข้อมูลศาสตร์ (data scientist) (Grimmer, 2015) นักสังคมศาสตร์ (social scientist) เป็นนักวิจัยสาขาวิชาหนึ่งที่เป็นที่ต้องการมากเช่นเดียวกับสาขาวิชาอื่นๆ ทำให้มีสถาบันทางการศึกษาในอเมริกาและยุโรปหลายแห่งพยายามปิดช่องว่างดังกล่าวด้วยการสร้างหลักสูตรให้ผู้เรียนมีความรู้ทั้งด้านสังคมศาสตร์ (social sciences) คอมพิวเตอร์ศาสตร์ (computational sciences) สถิติศาสตร์ (statistical

⁵⁵ ชาร์ลส์ ดาร์วิน (Charles Darwin) นักวิทยาศาสตร์ ได้ข้อมูลจากหมู่เกาะกาลาปากอสมาใช้พัฒนาทฤษฎีการคัดเลือกโดยธรรมชาติ (Theory of Natural Selection) และอัลเฟรด เรดคลิฟ - บราวน์ (Alfred R. Radcliffe - Brown) นักมานุษยวิทยา ได้ข้อมูลจากชาวเกาะและชาวพื้นเมืองมาพัฒนาทฤษฎีโครงสร้าง-หน้าที่ประโยชน์นิยม (Structural-Functionalism)

sciences) และวิทยาศาสตร์ (visual sciences) เพื่อผลิตนักวิทยาศาสตร์ด้านสังคมออกมาสู่ตลาดทั้งภาครัฐและเอกชนที่มีความต้องการเพิ่มขึ้นจำนวนมาก

การวิเคราะห์ด้านสังคมที่กำลังเป็นที่นิยมในยุคสังคมสารสนเทศและต้องการนักสังคมศาสตร์พันธุ์ใหม่ ได้แก่ การวิเคราะห์เครือข่ายทางสังคม (social network analysis) ตัวแบบเครือข่ายทางสังคม (social network modeling) เครือข่ายชุมชน (network communities) การเติบโตของพลวัตเครือข่ายทางสังคม (social network dynamic growth) ค่านิยมทางสังคม (social value) การแพร่กระจายของสารสนเทศ (information diffusion) พฤติกรรมการลงคะแนนเสียงและการเลือกตั้ง (voting behavior and elections) การกระทำร่วมกันและระบบความนึกคิด (partisanship and ideology) พฤติกรรมทางการเมือง (political activism) ความไว้วางใจ (trust) ความเอื้อเฟื้อ (altruism)ทุนทางสังคม (social capital) ความไม่เท่าเทียม และการเจรจาต่อรอง (inequality and bargaining) ความร่วมมือและความขัดแย้ง (cooperation and conflict) เป็นต้น (Pennsylvania State University, n.d.; Stony Brook University, n.d.; University of Cambridge, n.d.)

สรุป

การเปลี่ยนแปลงของเทคโนโลยีทางสังคมในศตวรรษที่ 21 มีผลทำให้มนุษย์มีพื้นที่ในการค้นหาความรู้เพิ่มขึ้น คือ โลกใหม่หรือไซเบอร์สเปซ มีปรากฏการณ์ใหม่ๆ ทางสังคมอีกจำนวนมากที่ยังไม่ได้ทำการศึกษาเพื่อตอบข้อสงสัย รวมถึงการนำเอาข้อมูลไปยืนยันหรือปฏิเสธทฤษฎีเก่าหรือการสร้างทฤษฎีใหม่ที่เหมาะสมกว่า เพื่อนำมาอธิบายปรากฏการณ์ทางสังคมที่เกิดจากการผสมผสานและเกิดคู่ขนานระหว่างโลกทางกายภาพและโลกเสมือนจริง อาจกล่าวได้ว่า การวิจัยเชิงปริมาณทางสังคมศาสตร์บนโลกเก่าและในโลกใหม่มีความแตกต่างกันไม่มากในด้านกระบวนการวิจัย เพราะมีการกำหนดปัญหาการวิจัย รวบรวมข้อมูล และการวิเคราะห์ข้อมูล รวมถึงการตอบปัญหาการวิจัยในรูปของรูปแบบ ความสัมพันธ์ โครงสร้าง และตัวแบบ เหมือนกัน แต่มีส่วนที่แตกต่างกันอย่างชัดเจนพอสรุปได้ดังนี้

1. ด้านข้อมูล เป็นเรื่องของจำนวนข้อมูลและมีการใช้ข้อมูลขนาดใหญ่มากขึ้น รูปแบบของข้อมูลที่มีความหลากหลาย (structure and unstructured) และกระจายอยู่ในหลายแห่ง (database) และการเปลี่ยนแปลงการใช้เทคโนโลยีในการรวบรวมข้อมูลและวิเคราะห์ข้อมูล
2. ด้านการออกแบบ เปลี่ยนแนวคิดการวิจัยแบบนิรนัย (deductive approach) ที่นิยมกันอย่างแพร่หลายไปสู่การวิจัยแบบอุปนัย (inductive approach) มากขึ้น ข้อสรุปที่ได้จะนำเอาทฤษฎีที่มีอยู่มาใช้อธิบายไม่ได้ นำผลการวิเคราะห์ไปใช้ยอมรับหรือปฏิเสธสมมติฐานที่ได้มาจากทฤษฎี หรืออาจเป็นเพียงการจารนัย (abductive approach) ในรูปของนำเสนอข้อมูล (data visualization) และการค้นหาหลักความสัมพันธ์แบบกลนัย (retroductive approach) เพื่อศึกษาโครงสร้างความสัมพันธ์ที่มีอิทธิพลต่อการทำนาย เท่านั้น
3. ด้านการวิเคราะห์ข้อมูล ความก้าวหน้าและความสามารถของเทคโนโลยีการคำนวณที่เรียกว่า การเรียนรู้ของเครื่องจักร ทำให้สามารถวิเคราะห์ข้อมูลได้ลึกและซับซ้อนกว่าการวิเคราะห์ด้วยสมองมนุษย์

(การวิจัยเชิงคุณภาพ) และเครื่องคอมพิวเตอร์ในอดีต (การวิจัยเชิงปริมาณ) รวมถึงสามารถทำการวิเคราะห์ได้อย่างต่อเนื่องเมื่อมีการเปลี่ยนแปลงของข้อมูลหรือเงื่อนไข โดยไม่ต้องเริ่มวิเคราะห์ใหม่

4. ด้านผลลัพธ์ ข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็วทำให้ไม่สามารถสร้างเป็นข้อสรุปทั่วไปสำหรับพัฒนาไปสู่การเป็นทฤษฎี ต้องใช้เครื่องจักรทำการวิเคราะห์และเรียนรู้ให้ได้คำตอบที่ใกล้เคียงกับปรากฏการณ์ทางสังคมในขณะนั้นมากที่สุด

นักสังคมศาสตร์ที่ทำวิจัยเชิงปริมาณที่ต้องการเปลี่ยนประสบการณ์ไปใช้ข้อมูลขนาดใหญ่ในการค้นหาคำตอบการวิจัยต้องปรับตัวและเรียนรู้เทคโนโลยีที่เป็นเครื่องมือในการเข้าถึงข้อมูล จัดเตรียมข้อมูล การลดทอนข้อมูล การวิเคราะห์ข้อมูล และการนำเสนอข้อมูลที่ออกแบบมาสำหรับใช้กับข้อมูลขนาดใหญ่ สิ่งที่สำคัญที่สุดในการวิจัยเชิงปริมาณด้วยข้อมูลขนาดใหญ่จากโลกใหม่ที่นักสังคมศาสตร์ต้องตระหนักก็คือ การเป็นตัวแทนของข้อมูล (data represent) เลือกเครื่องมือการวิเคราะห์ที่เหมาะสม (analytic tool) เข้าใจการประมาณการและความคาดเคลื่อน (estimates and errors) ของผลลัพธ์ การสร้างรูปแบบ ความสัมพันธ์ โครงสร้าง และตัวแบบจากข้อมูลในโลกใหม่อาจเป็นการย้อนยุคไปสู่การเริ่มพัฒนาความรู้ของมนุษย์ในอดีตที่นักทฤษฎีใช้วิธีการรวบรวมข้อมูลจากธรรมชาติและสังคมแล้วนำไปสร้างเป็นข้อสรุปชั่วคราวหรือทฤษฎีก่อนนำไปใช้ทดสอบด้วยข้อมูลจำนวนมากในเวลาต่อมา

เมื่อพิจารณาจากหลายมุมมองและความคิดที่กล่าวมาทั้งหมดในบทความนี้ขอสรุปให้เห็นความแตกต่างของกระบวนการวิจัยเชิงปริมาณด้านสังคมศาสตร์ระหว่างโลกเก่าและโลกใหม่ได้พอสังเขปดังตารางที่ 1

ตารางที่ 1 เปรียบเทียบกระบวนการวิจัยด้านสังคมศาสตร์ในโลกเก่าและโลกใหม่

กระบวนการวิจัย	โลกเก่า	โลกใหม่
ประชากร	คน ชุมชน องค์กร สังคม ประเทศ	ข้อมูล สารสนเทศ สิ่งเสมือนจริง
การรวบรวมข้อมูล	แบบสอบถาม สัมภาษณ์ สัมภาษณ์	โปรแกรมคอมพิวเตอร์
การจัดการข้อมูล	ข้อมูลขนาดเล็ก มีโครงสร้าง	ข้อมูลขนาดใหญ่ ไม่มีโครงสร้าง
การวิเคราะห์ข้อมูล	วิเคราะห์จากตัวอย่างหรือประชากร	วิเคราะห์ซ้ำ (iterative analysis) ระหว่างตัวอย่างกับประชากร
การใช้สถิติ	สถิติพรรณนา สถิติมีพารามิเตอร์ สถิติไม่มีพารามิเตอร์	สถิติพรรณนา สถิติไม่มีพารามิเตอร์
การแสดงผลข้อมูล	ตารางตัวเลข แผนภูมิ	สร้างภาพข้อมูล (visualization)
การอ้างอิง	อนุมานเชิงสถิติ (statistical inference)	อนุมานเชิงโครงสร้าง (structural inference)

ดังนั้น นักสังคมศาสตร์ที่สนใจการทำวิจัยเชิงปริมาณอาจต้องเปลี่ยนพื้นที่สำหรับรวบรวมข้อมูลใหม่ เพราะการเก็บรวบรวมข้อมูลด้วยแบบสำรวจบนโลกเก่ามีความยากมากขึ้น ความเป็นเมืองที่ขยายตัวอย่างรวดเร็วทำให้ประชากรและกลุ่มตัวอย่างมีความเป็นคนเมืองมากขึ้น เกิดความแปลกหน้าในสังคม และส่วนใหญ่ไม่ค่อยให้ความร่วมมือในการให้ข้อมูล จึงต้องหาเครื่องมือและวิธีการรวบรวมข้อมูลด้วยเทคโนโลยีสมัยใหม่ นอกจากนี้ยัง

ต้องเรียนรู้การใช้เทคโนโลยีสำหรับวิเคราะห์ข้อมูลใหม่ เช่น การใช้เครื่องมือวิเคราะห์ข้อมูลด้วยการประมวลผลแบบกลุ่มเมฆ (cloud computing) เหมือนกับนักวิจัยเชิงปริมาณในอดีตที่เปลี่ยนจากการนับรอยขีด (tally) มาใช้เครื่องคิดเลข และเปลี่ยนมาใช้เครื่องคอมพิวเตอร์ ตามลำดับ 0101010101010101010101010101

เอกสารอ้างอิง

ภาษาไทย

เรวัต แสงสุริยงค์. (2541). “การใช้อินเทอร์เน็ตสนับสนุนการเรียนการสอนด้านมนุษยศาสตร์และสังคมศาสตร์”.

วารสารมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา. ปีที่ 6 ฉบับที่ 5 ประจำเดือนพฤศจิกายน 2540 - พฤษภาคม 2541.

เรวัต แสงสุริยงค์ (2545). “การวิจัยอิเล็กทรอนิกส์: มุมมองเชิงปริมาณ (E-research: Quantitative Perspective)”. วารสารการวิจัยบูรพาศึกษา มหาวิทยาลัยบูรพา. ปีที่ 1 ฉบับที่ 2.

ฟรีดแมน, โทมัส แอล. (2554). **ใครว่าโลกกลม = The world is flat : a brief history of the twenty-first century**. (รอฮีม ปรามาส, แปล). กรุงเทพฯ : วีเลิร์น.

ภาษาต่างประเทศ

Amazon. (2015). **Big Data and Analytics on AWS**. Accessed December, 2, 2016. Available form <http://www.slideshare.net/AmazonWebServices/big-data-and-analytics-on-aws>

Cheung, Paul. (2011). **Big Data, Official Statistics and Social Science Research: Emerging Data Challenges**. Accessed October, 18, 2015. Available form <https://www.worldbank.org/wb/Big-data-pc-2012-12-12.pdf>

Desouza, Kevin C. & Smith, KenDra L. (2014). **Big Data for Social Innovation**. Accessed December, 5, 2015. Available form <http://search.proquest.com/docview/1540467654?accountid=44783>

Doorn, Perter. (2014). **Big data in the humanities and social sciences**. Accessed October, 31, 2015. Available form <https://sciencenode.org/feature/big-data-humanities-and-social-sciences.php>

Forbes. (2013). **A Very Short History of Big Data**. Accessed October, 24, 2015. Available form <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/2/>

Grimmer, Justin. (2015). **We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together**. Accessed October, 9, 2015. Available form http://stanford.edu/~jgrimmer/bd_2.pdf

- Hilbert, Martin.(2012). “How to Measure “How Much Information”? Theoretical,Methodological, and Statistical Challenges for the Social Sciences. **International Journal of Communication** 6 (2012), 1042-1055. Accessed October, 23, 2015 Available form <http://ijoc.org/index.php/ijoc/article/viewFile/1318/746>
- Hilbert, Martin & Lopez, Priscila. (2011). “The World’s Technological Capacity to Store, Communicate, and Compute Information”. **Science**. 332 (2011), 60-65. Accessed October, 23, 2015 Available form <http://ijoc.org/index.php/ijoc/article/download/1562/742>
- IDC. (2014). **The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things**. Accessed October, 9, 2015. Available form <http://idcdocserv.com/1678>
- Koepsell, David R. (2000). **The Ontology of Cyberspace: Law, Philosophy, and the Future of Intellectual Property**. Illinois: Open Court.
- Lam, Dominic. (2014). **Big Data Challenges in Social Sciences & Humanities Research**. Accessed October, 9, 2015. Available form <http://www.datanami.com/2014/09/08/big-data-challenges-social-sciences-humanities-research/>
- Mar, Bernard. (2014). **Big Data: 20 Free Big Data Sources Everyone Should Know**. Retrieved November, 4, 2015. Accessed form <http://www.smartdatacollective.com/bernardmarr/235366/big-data-20-free-big-data-sources-everyone-should-know>
- Mrazek, Patricia, Biglan, Anthony Hawkins, J. David. (n.d.). **Community-Monitoring Systems: Tracking and Improving The Well-Being Of America’s Children and Adolescents**. Accessed November, 4, 2015. Available form <http://www.preventionresearch.org/CMSbook.pdf>
- Mewawalla, C. (2012). **Big Data**. Accessed October, 9, 2015. Available form <http://www.cmresearch.co.uk/resources/Sync+43+Big+Data.pdf>
- ODE. (2015). **Oxford English Dictionary**. Accessed October, 25, 2015. Available form <http://www.oed.com/viewdictionaryentry/Entry/95568>
- Pennsylvania State University. (n.d.). **Big Data Social Science**. Accessed October, 9, 2015. Available from <http://bdss.psu.edu/>
- Rovai, A.P., Baker, J.D. & Ponton, M.K. (2013). **Social Science Research Design and Statistics: A Practitioner’s Guide to Research Methods and SPSS Analysis**. [Kindle version]. Accessed October, 21, 2015. Available from <http://www.amazon.com>

Stony Brook University (n.d.). **Big Data for the Social Sciences**. Accessed October, 9, 2015.

Available from <http://www.stonybrook.edu/commcms/clusterhires/searches/bigdata.html>

United Nations. (2014). **World Urbanization Prospects**. Accessed November, 4, 2015. Available from <http://esa.un.org/unpd/wup/highlights/wup2014-highlights.pdf>

UNECE Statistics Wikis. (2013). **Big Data Home**. Retrieved October, 22, 2015. Accessed from <http://www1.unece.org/stat/platform/display/bigdata>

University of Cambridge. (n.d.). **Preparing social scientists for the world of big data**. Retrieved October, 9, 2015. Accessed from <http://www.cam.ac.uk/research/features/preparing-social-scientists-for-the-world-of-big-data>

Waterman, K. Krasnow and Bruening, Paula J. (2014). “Big data analytics: Risks and responsibilities”. **International Data Privacy Law**, 4(2), 89-95. Available from doi:<http://dx.doi.org/10.1093/idpl/ipu002>

Wu, Michael.(2013). **Big Data Reduction 1: Descriptive Analytics**. Accessed December, 5, 2015. Available from <http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-1-Descriptive-Analytics/ba-p/77766>

Wu, Michael. (2013). **Big Data Reduction 2: Understanding Predictive Analytics**. Accessed December,5,2015. Available from <http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-2-Understanding-Predictive-Analytics/ba-p/79616>

Wu, Michael. (2013). **Big Data Reduction 3: From Descriptive to Prescriptive**. Accessed December,5, 2015. Available from <http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-3-From-Descriptive-to-Prescriptive/ba-p/81556>