

การสำรวจแนวทางใหม่: เนปาลนำร่องใช้แนวทาง บูรณาการข้อมูล



ข้อจำกัดความรับผิดชอบ: เนื้อหาที่นำเสนอใน Stats Brief ไม่ใช่การแสดงออกถึงความคิดเห็นใด ๆ ของสำนักงานเลขาธิการสหประชาชาติที่เกี่ยวข้องกับสถานะทางกฎหมายของประเทศ ดินแดน เมืองหรือพื้นที่ใด ๆ หรือของหน่วยงานหรือเกี่ยวกับการกำหนดพรมแดนหรือเขตแดน ซึ่งเมื่อมีการระบุถึง "ประเทศหรือพื้นที่" จะครอบคลุมหมายถึง ประเทศ ดินแดน เมือง หรือพื้นที่ และบรรณานุกรมและข้อมูลอ้างอิงอื่น ๆ ได้รับการตรวจสอบทุกครั้งที่สามารถทำได้ โดยองค์การสหประชาชาติไม่รับผิดชอบต่อความพร้อมใช้งานหรือการทำงานของ URL นอกจากนี้ ความคิดเห็น ตัวเลข ค่าประมาณที่กล่าวถึงในสิ่งพิมพ์นี้ไม่ควรถือว่าเป็นการสะท้อนความคิดเห็นหรือถือเป็นการสนับสนุนจากองค์การสหประชาชาติ รวมถึงการกล่าวถึงชื่อบริษัทและผลิตภัณฑ์เชิงพาณิชย์ด้วย

Stats Brief นี้จัดทำโดย Jose Ramon Albert, Hossein Hassani และ Ram Hari Gaihre จากกลุ่มที่ปรึกษาของโครงการ ภายใต้การดูแลของ Afsaneh Yazdani นักสถิติ แผนกสถิติของ ESCAP และ Petra Nahmias หัวหน้าส่วนสถิติ ประชากรและสังคม แผนกสถิติของ ESCAP ซึ่งผู้เขียนได้จัดทำ Stats Brief นี้โดยอ้างอิงจากรายงานเรื่อง “การเสริมสร้างความสามารถทางสถิติของประเทศสำหรับการใช้แนวทางบูรณาการข้อมูล (Strengthening the National Statistical Capacity to Implement Data Integration Approaches)”¹ ผู้เขียนขอขอบคุณสำนักงานสถิติแห่งชาติเนปาล โดยเฉพาะอย่างยิ่ง Munni Kumari Chaudhary รองหัวหน้าสถิติกองแผนและบริหารทรัพยากรบุคคล และ Rishi Ram Sigdel ผู้อำนวยการส่วนฝึกอบรม สำหรับการสนับสนุนอย่างเต็มที่ตลอดโครงการ และผู้เขียนขอขอบคุณ Sovannaroth Tey สำหรับข้อมูลและการสนับสนุนอันล้ำค่าในโครงการ หากต้องการข้อมูลเพิ่มเติมของนโยบายนี้ สามารถติดต่อสอบถามได้ที่:

Rachael Beaven

ผู้อำนวยการกองสถิติ

คณะกรรมการเศรษฐกิจและสังคมแห่งเอเชียและแปซิฟิก (ESCAP)

Email: stat.unescap@un.org

หมายเลขติดตาม: พฤษภาคม 2566 | ฉบับที่ 33

เครดิตรูปภาพ: iStock-849234128.jpg

¹ <https://www.unescap.org/kp/2023/strengthening-national-statistical-capacity-implement-data-integration-approaches-pilot>.

สารบัญ

สารบัญ.....	4
สรุป.....	5
บทนำ.....	6
โครงการเนปาล.....	7
การเปิดใช้งานการเข้าถึงข้อมูล.....	8
การเตรียมพร้อมการประมวลผลแหล่งข้อมูล.....	9
การเชื่อมโยงแหล่งข้อมูลและผลลัพธ์ที่อาจเกิดขึ้น.....	10
บทสรุป ข้อจำกัด และแนวทางต่อไป.....	10
กล่องข้อมูล	
กล่องที่ 1 – แหล่งข้อมูลที่เกี่ยวข้องในการนำร่อง.....	8
กล่องที่ 2 – ขั้นตอนหลักก่อนการประมวลผลข้อมูล.....	9

สรุป

แนวทางบูรณาการข้อมูลกำลังเป็นที่นิยมอย่างมากในนักสถิติทางการ (official statistician) และผู้ผลิตข้อมูล ซึ่งในการตระหนักถึงศักยภาพของตนอย่างเต็มที่ จำเป็นต้องมีความพร้อมทางองค์กร (institutional capacity) และความพร้อมทางเทคนิค (technical capacity) ให้มากขึ้น สำหรับในส่วนของ การเสริมสร้างความสามารถของระบบสถิติในเอเชียและแปซิฟิก ESCAP ได้เผยแพร่แนวทางของแปซิฟิกเอเชียในเรื่องการบูรณาการข้อมูลสำหรับสถิติทางการ² และริเริ่มชุมชนนักปฏิบัติการเกี่ยวกับการบูรณาการข้อมูล (Data Integration Community of Practice)³ ซึ่งจะมีพื้นที่ในการแบ่งปันความรู้ ประสบการณ์ และการเรียนรู้ในเรื่องดังกล่าว นอกจากนี้ใน พ.ศ. 2565 ESCAP ได้ร่วมมือกับสำนักงานสถิติแห่งชาติ (National Statistical Office: NSO) ของเนปาล ริเริ่มโครงการเพื่อใช้แนวทางที่เหมาะสมในเนปาลสำหรับปรับปรุงความสามารถของระบบสถิติของประเทศ (National Statistical System: NSS) ในการบูรณาการข้อมูลและผลิตข้อมูลที่ดียิ่งขึ้น เพื่อวัตถุประสงค์วัดเป้าหมายการพัฒนาที่ยั่งยืน (Sustainable Development Goals: SDG) โดย Stats Brief นี้ได้สรุปภาพรวมและคำแนะนำทางเทคนิคของโครงการ รวมถึงสนับสนุนการเสริมสร้างขีดความสามารถแก่ NSO เนปาลและพันธมิตรด้านข้อมูลของ NSO เนปาล

² <https://www.unescap.org/kp/2021/asia-pacific-guidelines-data-integration-official-statistics>.

³ <https://www.unescap.org/our-work/statistics/communities>.

บทนำ

สำนักงานสถิติแห่งชาติ (National Statistical Office: NSO) ทั่วโลกกำลังขยายขอบเขตการผลิตสถิติของตนเพื่อตอบสนองต่อบริบทการดำเนินการที่เปลี่ยนแปลงไป โดยได้ยกระดับความต้องการสถิติในพันธสัญญาทั่วโลก (global commitment) เช่น วาระการพัฒนาที่ยั่งยืน ค.ศ. 2030 (2030 Agenda for Sustainable Development) และแผนพัฒนาระดับประเทศ (national development plan) ให้มีคุณภาพ ทันเวลา และระดับย่อย (granular) ซึ่ง NSO กำลังตอบสนองต่อข้อจำกัดด้านงบประมาณและอัตราการตอบที่ลดลงโดยเปลี่ยนจากวิธีการเก็บรวบรวมข้อมูลแบบดั้งเดิม (traditional data collection) ที่รวมถึงสำมะโนและการสำรวจตัวอย่าง ไปสู่การใช้ข้อมูลเพื่อการบริหาร (administrative data) และแหล่งข้อมูลใหม่ที่ดียิ่งขึ้น และด้วยความก้าวหน้าทางเทคโนโลยีและความพร้อมใช้งานของแหล่งข้อมูลใหม่ ทำให้ปัจจุบัน NSO มีทางเลือกมากขึ้น

สำหรับเอเชียและแปซิฟิก ตามรายงานความคืบหน้าล่าสุดของ SDG พบว่าตัวชี้วัด SDG จำนวน 128 ตัวจาก 231 ตัวชี้วัด มีความพร้อมใช้งานของข้อมูล (data availability) อย่างเพียงพอ โดยในรายงานระบุว่า *ข้อมูลที่ไม่เพียงพอต่อการนำไปใช้งานต่อสำหรับ 51 จาก 169 เป้าหมาย บ่งชี้ให้ระบบสถิติ (statistical system) ในภูมิภาคเพิ่มความพยายามในการดำเนินการเป็นสองเท่าเพื่อเติมเต็มช่องว่างของข้อมูล⁴* ซึ่งในความพยายามที่จะเสริมสร้างการผลิตสถิติทางการ NSO ควรสำรวจแนวทางใหม่ ๆ โดยการบูรณาการข้อมูล รวมไปถึงการรวมกันของแหล่งข้อมูลที่แตกต่างกัน เช่น สำมะโน การสำรวจตัวอย่าง ระเบียบเพื่อการบริหาร (administrative records) และข้อมูลขนาดใหญ่ (big data) ซึ่งเป็นหนทางหนึ่งในการเชื่อมช่องว่างของข้อมูล และการบูรณาการข้อมูลจะช่วยให้ความกระจ่างในพื้นที่ที่คลุมเครือและสร้างข้อมูลเชิงลึกที่ไม่สามารถหาได้จากแหล่งข้อมูลเดียว

การบูรณาการข้อมูลดูเหมือนจะมีแนวโน้มดี แต่ต้องเอาชนะความท้าทายเพื่อดึงศักยภาพออกมาใช้ได้อย่างเต็มที่ โดยเฉพาะอย่างยิ่งในแง่ของความพร้อมทางองค์กรและความพร้อมทางเทคนิค โดยความพร้อมทางองค์กรมักหมายถึงความพร้อมของ ก) แหล่งข้อมูลที่เกี่ยวข้อง ข) พื้นฐานทางกฎหมายที่เข้มแข็งซึ่งสนับสนุนการใช้และการเข้าถึงข้อมูลและคำอธิบายข้อมูล (metadata) เพื่อวัตถุประสงค์ทางสถิติและกฎหมายที่คุ้มครองข้อมูลส่วนบุคคล ค) กระบวนการที่เหมาะสมในการทำงานร่วมกับผู้ถือข้อมูลต้นทาง (data holder) (ทั้งจากภาครัฐและเอกชน) ง) โครงสร้างพื้นฐานสารสนเทศ (ICT Infrastructure) จ) ทรัพยากรบุคคลรวมถึงความเชี่ยวชาญและทักษะของบุคคล

ความพร้อมทางเทคนิคยังเป็นความท้าทายเช่นกันซึ่งรวมถึง ก) ทักษะในการจัดการกับการขาดการทำงานร่วมกัน ในแง่ของการขาดตัวระบุที่ไม่ซ้ำกัน (unique identifier) แนวคิดที่แตกต่างกัน การจำแนกประเภท คุ่มรวมรูปแบบข้อมูล ระยะเวลาอ้างอิง และอื่น ๆ ข) ทักษะในการจัดการปัญหาคุณภาพของข้อมูล เช่น ข้อมูลสูญหาย (missing data) ค่าผิดพลาด (erroneous value) และความไม่สอดคล้องกันภายในและระหว่างแหล่งข้อมูล และ ค) ความเชี่ยวชาญในการเชื่อมโยงระเบียบ (record linkage) และการจับคู่ทางสถิติ (statistical matching) นอกจากนี้ เจ้าหน้าที่ของ NSO ยังจำเป็นต้องมีความเข้าใจเชิงลึกและความรู้ในแหล่งข้อมูลที่เกี่ยวข้อง ซึ่งสามารถหาได้จากคำอธิบายข้อมูลที่เหมาะสม ตลอดจนการมีสื่อสารกับผู้ถือข้อมูลต้นทางในระดับเชี่ยวชาญอย่างสม่ำเสมอ และทักษะอื่น ๆ ที่ NSO ยังจำเป็นต้องมี คือ ทักษะด้านอารมณ์ (soft skill) เช่น การเจรจาต่อรอง การสร้างความสัมพันธ์ และการสื่อสาร

⁴ <https://www.unescap.org/kp/2023/asia-and-pacific-sdg-progress-report-2023>.

แม้ว่าความสามารถที่จำเป็นสำหรับการบูรณาการข้อมูลจะมีขอบเขตกว้าง แต่ NSO อาจมีทักษะและประสบการณ์เหล่านี้แล้วสำหรับการบูรณาการข้อมูลในระดับมหภาค เช่น ในการรวบรวมบัญชีประชาชาติ (National Accounts) NSO อาจสามารถใช้ต่อยอดในการบูรณาการข้อมูลระดับจุลภาค (micro-level data) ได้โดยที่มีขีดจำกัดในเรื่องความสามารถ นอกจากนี้ การสำรวจการประเมินความสามารถในการบูรณาการข้อมูล (Data Integration

Capacity Assessment Survey) ในระดับภูมิภาค ที่จัดทำขึ้นในปี พ.ศ. 2563 โดย ESCAP แสดงให้เห็นถึงช่องว่างด้านความสามารถอย่างมีนัยสำคัญ โดยพบว่า มีเพียงร้อยละ 45 ของประเทศที่ตอบแบบสอบถามนั้นมีทักษะที่จำเป็น และร้อยละ 77 ของประเทศที่ตอบแบบสอบถาม มีความต้องการที่จะเสริมสร้างความสามารถอย่างมาก ซึ่งรวมถึงประเทศเนปาลด้วย

โครงการเนปาล

ในเดือนมิถุนายน พ.ศ. 2565 ESCAP ได้ริเริ่มโครงการในชื่อ “การเสริมสร้างความสามารถทางสถิติของประเทศสำหรับการใช้แนวทางบูรณาการข้อมูล เพื่อผลิตและปรับปรุงข้อมูลสำหรับการวัดตัวชี้วัด SDG (Strengthening the National Statistical Capacity to Implement Data Integration Approaches to Produce and Improve Data for Measuring SDG Indicators)” โดยโครงการนี้มีวัตถุประสงค์เพื่อร่วมมือกับ NSO เนปาลอย่างใกล้ชิดในการฝึกปฏิบัติสำหรับการนำร่องการบูรณาการข้อมูลจากแหล่งข้อมูลหลายแหล่ง รวมถึงสำมะโนการสำรวจ และข้อมูลเพื่อการบริหาร และโครงการนี้มีเป้าหมายเพื่อพัฒนาขีดความสามารถ และให้ความช่วยเหลือด้านเทคนิคและกล่องเครื่องมือไอที (IT toolbox) สำหรับใช้ในกระบวนการเต็มรูปแบบของการฝึกบูรณาการข้อมูล

ขั้นตอนแรกคือการฝึกกระบวนการบูรณาการข้อมูลที่สามารถเสร็จสิ้นได้ภายในเวลาและทรัพยากรที่จำกัดของโครงการ โดยคำนึงถึงลำดับความสำคัญและความพร้อมใช้งานของประเทศตนเอง และแหล่งข้อมูลที่สามารถเข้าถึงได้ ในครั้งนี้ มีการนำการศึกษาแบบนั่งโต๊ะ (desk study) มาใช้ และมีการรวบรวมมุมมองของผู้มีส่วนได้ส่วนเสียของประเทศผ่านการ

ประชุมเชิงปฏิบัติการขั้นต้น (inception workshop) และการอภิปรายหลายครั้ง

การฝึกที่กำหนดไว้มีวัตถุประสงค์สำหรับบูรณาการแหล่งข้อมูลที่เกี่ยวข้องเพื่อให้สามารถคำนวณ ก) สัดส่วนของประชากรที่ได้รับ ความคุ้มครองตามระบบและมาตรการการคุ้มครองทางสังคม (social protection floors/systems) จำแนกตามสถานะความยากจน โดยจะปรับปรุงข้อมูลสำหรับการติดตามตัวชี้วัด SDG 1.3.1⁵ และข) ผลผลิตสถิติที่เกี่ยวข้องสำหรับใช้ในการตรวจสอบสถานะความยากจนของครัวเรือน ซึ่งการฝึกที่เกี่ยวข้องกับการบูรณาการแหล่งข้อมูลเฉพาะจำนวนสี่แหล่งได้อธิบายไว้ในกล่องที่ 1

มีการให้ความช่วยเหลือด้านเทคนิคและ ICT เพื่ออำนวยความสะดวกในการเข้าถึง จัดเตรียม และเชื่อมโยงแหล่งข้อมูล ตลอดจนสร้างสถิติที่เฉพาะเจาะจง ตามมาด้วยการอบรมเชิงปฏิบัติการ ณ กาฐมาณฑุ ประเทศเนปาล ในเดือนกุมภาพันธ์ พ.ศ. 2566⁶ เพื่อเพิ่มความสามารถทางเทคนิคในแนวทางบูรณาการข้อมูลทั่วไป และในส่วนที่เกี่ยวข้องกับการฝึกที่ได้กำหนดไว้

⁵ ตัวชี้วัด SDG 1.3.1: สัดส่วนของประชากรที่ได้รับ ความคุ้มครองตามระบบและมาตรการการคุ้มครองทางสังคม จำแนกตามเพศ และแบ่งเป็น เด็ก ผู้ว่างงาน ผู้สูงอายุ ผู้พิการ หญิงตั้งครรภ์ เด็กเกิดใหม่ ผู้ที่ได้รับบาดเจ็บจากการทำงาน ผู้ที่ยากจนและเปราะบาง

⁶ <https://www.unescap.org/events/2023/national-training-workshop-implementing-data-integration-approaches-official-statistics>

กล่องที่ 1 – แหล่งข้อมูลที่เกี่ยวข้องในการนำร่อง

- ผู้รับประโยชน์จากเบี้ยเลี้ยงประกันสังคม พ.ศ. 2565 (*Beneficiaries of Social Security Allowance 2022: BSSA*) เป็นการขึ้นทะเบียนที่ได้รับการพัฒนาและดูแลโดย Department of National ID and Civil Registration ของเนปาล ซึ่งรวมไปถึงข้อมูลเกี่ยวกับผู้รับประโยชน์จากเบี้ยเลี้ยงประกันสังคมสำหรับประชาชนกลุ่มเปราะบาง
- การสำรวจครัวเรือนยากจน พ.ศ. 2556 (*Poor Households Identification Survey 2013: PHIS*) เป็นการสำรวจขนาดใหญ่โดย Ministry of Land Management, Cooperatives and Poverty Alleviation ของเนปาล เพื่อระบุคนยากจนในการให้บริการตามเป้าหมาย (targeting service)
- สำมะโนประชากรและเคหะของประเทศ พ.ศ. 2554 (*National Population and Housing Census 2011: NPHC*) เก็บรวบรวมรายละเอียดข้อมูลเกี่ยวกับประชากร การเคหะ และความเป็นเจ้าของสินทรัพย์
- การสำรวจมาตรฐานการครองชีพของเนปาล (*Nepal Living Standards Survey 2010/11: NLSS*) เป็นการสำรวจหลายหัวข้อที่ดำเนินการโดย NSO เพื่อผลิตสถิติการบริโภคและข้อมูลอื่น ๆ ที่ความสัมพันธ์กับความยากจน

การเปิดใช้งานการเข้าถึงข้อมูล

ข้อกำหนดสำคัญสำหรับ NSO ในการดำเนินการบูรณาการข้อมูลคือ การได้รับสิทธิ์เข้าถึงข้อมูลและคำอธิบายข้อมูลที่จำเป็น ซึ่งในเนปาลนั้นมีพื้นฐานทางกฎหมายที่สนับสนุนการใช้และการเข้าถึงข้อมูลและคำอธิบายข้อมูลสำหรับวัตถุประสงค์ทางสถิติ ตลอดจนกฎหมายเกี่ยวกับการรักษาความลับและข้อมูลส่วนบุคคล นอกจากนี้ NSO ยังมีข้อตกลงผ่านหนังสือราชการกับ Department of National ID and Civil Registration และ Ministry of Land Management, Cooperatives and Poverty Alleviation ในการเข้าถึงข้อมูลเพื่อวัตถุประสงค์ทางสถิติ

อย่างไรก็ตาม ผู้ถือข้อมูลต้นทางที่เกี่ยวข้องไม่สมัครใจที่จะแบ่งปันข้อมูลระดับจุลภาค เนื่องจากความกังวลเรื่องความเป็นส่วนตัว รวมถึงความไม่แน่นอนเกี่ยวกับความเป็นไปได้ของโครงการ และเพื่อแก้ปัญหาและส่งเสริมการทำงานร่วมกัน จึงได้มีการจัดประชุมระหว่าง NSO ผู้ถือข้อมูลต้นทาง และสมาชิกของโครงการหลายครั้ง นอกจากการประชุมเชิงปฏิบัติการขั้นต้นของโครงการแล้ว ในระหว่างการประชุมได้เน้นย้ำถึงความมุ่งมั่นของ NSO ในการปกป้อง

ความเป็นส่วนตัวและการรักษาความลับตามกฎหมายทางสถิติ⁷ และหลักการพื้นฐานของสถิติทางการ (*Fundamental Principles of Official Statistics*)⁸ ด้วย นอกจากนี้ ยังได้แจ้งถึงวัตถุประสงค์ กระบวนการ และวิธีการ และผลลัพธ์ที่คาดว่าจะเกิดขึ้นของโครงการ รวมถึงผลประโยชน์ร่วมกันแก่ผู้ถือข้อมูลต้นทาง

เพื่อให้สามารถจัดการการฝึกบูรณาการข้อมูลได้ง่ายขึ้น และเพื่อส่งเสริมการแบ่งปันข้อมูลจากผู้ถือข้อมูลต้นทาง จึงจำกัดขอบเขตของการฝึกให้อยู่ในหนึ่งจังหวัด คือ คัณฑกี (Gandaki) ซึ่งเมื่อเทียบกับจังหวัดอื่นแล้ว จังหวัดนี้มีครัวเรือน/บุคคลในทุกแหล่งข้อมูลมากกว่า นอกจากนี้ เนื่องจากข้อมูลระดับย่อย (microdata) จากสำมะโนประชากรและการสำรวจมาตรฐานการครองชีพของเนปาล ครั้งล่าสุดไม่พร้อมใช้งาน จึงตัดสินใจใช้ฐานข้อมูลที่เก่ากว่า ซึ่งความพยายามทั้งหมดนี้ช่วยให้ผู้ถือข้อมูลต้นทางมีความเต็มใจในการแบ่งปันข้อมูลเพิ่มขึ้น

⁷ <http://rajpatra.dop.gov.np/welcome/book?ref=25085>.

⁸ <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

การเตรียมพร้อมการประมวลผลแหล่งข้อมูล

การบูรณาการข้อมูลระหว่างแหล่งข้อมูลที่แตกต่างกันเป็นสิ่งที่ท้าทาย เนื่องจากมีปัญหาที่เกิดจากความต้องการตัวระบุที่ไม่ซ้ำกัน การทำงานร่วมกัน และคุณภาพของข้อมูล ดังนั้น การเตรียมพร้อมการประมวลผลจึงถือเป็นขั้นตอนที่สำคัญในการฝึกบูรณาการข้อมูลทุกครั้ง

การเตรียมพร้อมการประมวลผลต้องมีการทำความสะอาดข้อมูล (data cleaning) และการกำหนดมาตรฐานข้อมูล (standardising data) เพื่อให้แน่ใจว่าข้อมูลที่ใช้สำหรับการบูรณาการนั้นถูกต้อง สมบูรณ์ และสอดคล้องกันกับแหล่งข้อมูลต่าง ๆ และพร้อมสำหรับการเชื่อมโยงระเบียบ โดยกล่องที่ 2 จะนำเสนอขั้นตอนหลักของการเตรียมพร้อมการประมวลผล

กล่องที่ 2 – ขั้นตอนหลักของการเตรียมพร้อมการประมวลผล

- **การทำความสะอาดข้อมูล** เกี่ยวข้องกับการระบุและแก้ไขค่าผิดพลาด ค่าที่ไม่สอดคล้องกัน ค่าสูญหาย รูปแบบข้อมูลที่ไม่ถูกต้อง ค่าที่เขียนผิด ข้อมูลซ้ำกัน และอื่น ๆ
- **การแปลงข้อมูล (data transformation)** หมายถึง การแปลงหน่วยการปกครอง (administrative unit) และ/หรือตัวแปรให้เป็นหน่วยสถิติ (statistical unit) และ/หรือตัวแปร นอกจากนี้ยังเกี่ยวข้องกับการแปลงรหัสประจำตัวประชาชนเป็นระเบียบประจำตัวเพื่อปกป้องความเป็นส่วนตัว
- **การกำหนดมาตรฐานข้อมูล** ทำให้เกิดความสอดคล้องกันของแหล่งข้อมูลต่าง ๆ ในแง่ของรายละเอียดตัวแปร (variable label) ค่าข้อมูล (value) การแปลงรหัส (coding) รูปแบบ การสะกด ฯลฯ และขั้นตอนนี้ยังเกี่ยวข้องกับการระบุตัวแปรที่เหมือนกันระหว่างแหล่งข้อมูลต่าง ๆ และกำหนดมาตรฐานให้อยู่ในรูปแบบเดียวกัน

บางปัญหาของข้อมูลที่พบในการดำเนินการของเนปาล ได้แก่ ก) การขาดหมายเลขประจำตัวที่ไม่ซ้ำกัน ข) ความแตกต่างในการถอดความชื่อ (ชื่อ ชื่อกลาง และนามสกุล) ค) ความแตกต่างในคาบเวลาที่ข้อมูลอ้างอิง (reference period) และ ง) ความแตกต่างในการจำแนกประเภทและหมวดหมู่ทางภูมิศาสตร์ ตลอดจนตัวแปรสำคัญบางตัวแปร เช่น กลุ่มทางสังคม (caste group) และกลุ่มชาติพันธุ์ (ethnic group)

การเตรียมพร้อมการประมวลผลดำเนินการโดยใช้แนวทางกึ่งอัตโนมัติ (semi-automated approach) เนื่องจากบางปัญหายังต้องใช้การแก้ไขด้วยมือ (manual) ปัญหาที่เกี่ยวข้องกับการแปลงชื่อจากภาษาเนปาลเป็นภาษาอังกฤษได้รับการแก้ไขด้วยการตรวจสอบข้อมูลอย่างละเอียดโดย

ใช้หลักการตั้งชื่อตัวแปรภาษาเนปาล (Nepali naming convention) นอกจากนี้ยังมีปัญหาเกี่ยวกับการสะกดชื่อซึ่งต้องแก้ไขด้วยมือเช่นกัน มาตรฐานการให้รหัสได้รับการพัฒนาเพื่อให้แน่ใจว่าป้ายและหมวดหมู่ของแอตทริบิวต์ (attribute) สอดคล้องกันในทุกฐานข้อมูล ส่วนความแตกต่างของคาบเวลาที่ข้อมูลอ้างอิงได้รับการจัดการโดยการปรับตามความแตกต่างของอายุ

เนื่องจากการขาดตัวระบุที่ไม่ซ้ำกันระหว่างแหล่งข้อมูล จึงตัดสินใจที่จะใช้การจับคู่ตามความน่าจะเป็น (probabilistic matching) โดยอ้างอิงจากชื่อ สถานที่ และตัวแปรด้านประชากร ตามความเหมาะสม

การเชื่อมโยงแหล่งข้อมูลและผลลัพธ์ที่อาจเกิดขึ้น

การเชื่อมโยงระเบียบ หมายถึง การระบุและการรวมกันของระเบียบที่สัมพันธ์กับบุคคลเดียวกัน ระหว่างแหล่งข้อมูลตั้งแต่สองแหล่งขึ้นไป ซึ่งประกอบด้วยสองแนวทางหลักในการเชื่อมโยงระเบียบ คือ การจับคู่เชิงกำหนด (deterministic matching) และการจับคู่ตามความน่าจะเป็น

การจับคู่เชิงกำหนดเกี่ยวข้องกับการเปรียบเทียบระเบียบในฐานข้อมูลสองชุดเพื่อค้นหาการจับคู่ที่เหมาะสมตามตัวระบุที่ไม่ซ้ำกัน ในทางตรงกันข้าม การจับคู่ตามความน่าจะเป็นใช้วิธีการทางสถิติเพื่อพิจารณาความเป็นไปได้ที่ระเบียบสองระเบียบบ่งบอกถึงความเหมือนกัน แม้ว่าการจับคู่เชิงกำหนดจะเป็นวิธีตรงไปตรงมา แต่อาจส่งผลให้การจับคู่ผิดพลาดหากมีเกณฑ์ที่เข้มงวดเกินไป หรือเป็นการจับคู่เท็จหากมีเกณฑ์ที่ผ่อนปรนเกินไป การจับคู่ตามความน่าจะเป็นแม้ว่าจะซับซ้อนกว่า แต่สามารถให้ผลการจับคู่ที่แม่นยำมากกว่า โดยกำหนดน้ำหนักให้กับเกณฑ์การจับคู่ที่แตกต่างกันตามความสำคัญและค่านวนคะแนนความน่าจะเป็นสำหรับในแต่ละการจับคู่ที่เป็นไปได้ โดยรวมแล้ว การเลือกแนวทางใดขึ้นอยู่กับข้อกำหนดเฉพาะของแอปพลิเคชัน (application) รวมถึงคุณภาพของข้อมูลระดับความถูกต้องที่ต้องการ และทรัพยากรที่มีอยู่

ในการฝึกปฏิบัติของเนปาล เนื่องจากขาดหมายเลขประจำตัวที่ไม่ซ้ำกันในแหล่งข้อมูลต่าง ๆ จึงไม่สามารถจับคู่เชิงกำหนดได้ ดังนั้นจึงต้องใช้วิธีการจับคู่ตามความน่าจะเป็นโดยใช้ตัวแปรร่วมกัน และเพื่อความสะดวกในการเชื่อมโยงระเบียบ จึงได้พัฒนากล่องเครื่องมือไอทีขึ้นมาซึ่งมีคุณสมบัติที่จำเป็นแก่ผู้ใช้ในการสำรวจแหล่งข้อมูลสร้างหลักเกณฑ์ และดำเนินงานในกระบวนการเชื่อมโยง

กล่องเครื่องมือยังสร้างสถิติที่เฉพาะเจาะจงได้ ในการฝึกนี้กล่องเครื่องมือจะสร้างตารางไขว้ (cross-tabulation) ของตัวแปรจากแหล่งข้อมูลสองแหล่งที่แตกต่างกัน เช่นการจัดเตรียมตารางประเภทของเบี้ยเลี้ยงประกันสังคมจาก BSSA เทียบกับสถานะความยากจนจาก PHIS ซึ่งทำให้สามารถคำนวณสัดส่วนของประชากรที่ครอบคลุมระบบคุ้มครองทางสังคม (social protection system) ตามสถานะความยากจนได้ นอกจากนี้ ยังจัดทำตารางระดับความมั่งคั่ง (wealth quintile) จาก NPHC เทียบกับสถานะความยากจนจาก PHIS โดยตารางนี้ช่วยประมาณค่าร้อยละของครัวเรือนในกลุ่มล่างของดัชนีความมั่งคั่ง (wealth index) ของ PHIS ที่ถือว่ายากจน ทำให้สามารถตรวจสอบสถานะความยากจนที่คำนวณใน PHIS ได้

บทสรุป ข้อจำกัด และแนวทางต่อไป

การฝึกปฏิบัตินี้สร้างต้นแบบสำหรับความพยายามในการบูรณาการข้อมูลในอนาคตของเนปาล ซึ่งแสดงให้เห็นถึงการบูรณาการแหล่งข้อมูลหลายแหล่ง รวมถึงสำมะโนประชากร การสำรวจตัวอย่าง และข้อมูลเพื่อการบริหาร โดยการดำเนินการครอบคลุมขั้นตอนทั้งหมดที่เกี่ยวข้องกับกิจกรรมการบูรณาการข้อมูล ตั้งแต่การกำหนดวัตถุประสงค์สำหรับการบูรณาการข้อมูลไปจนถึงการเข้าถึงแหล่งข้อมูล การเตรียมพร้อมการประมวลผลข้อมูล การเชื่อมโยงระเบียบ และสุดท้าย สร้างสถิติแบบใหม่ภายในหนึ่งจังหวัด

คือ คณิตศาสตร์ แม้ว่าการดำเนินการนี้จะประสบความสำเร็จในหลาย ๆ ด้าน แต่ควรรับทราบข้อจำกัดบางอย่างด้วยเช่นกัน เช่น ข้อมูลล่าสุดจะมีประโยชน์และเกี่ยวข้องกับนโยบายมากกว่า

การสร้างขีดความสามารถขึ้นใน NSO และผู้มีส่วนได้ส่วนเสียในโครงการช่วยให้สามารถขยายขอบเขตของการฝึกปฏิบัติ นำร่องไปทั่วทั้งประเทศ รวมทั้งริเริ่มโครงการบูรณาการข้อมูลอื่น ๆ ที่คล้ายคลึงกัน แม้ว่าการฝึกปฏิบัติ

ในปัจจุบันจะมุ่งเน้นไปที่การบูรณาการข้อมูลจากสำมะโน การสำรวจ และแหล่งข้อมูลเพื่อการบริหาร แต่ NSO ยังอาจต้องการสำรวจการใช้แหล่งข้อมูลใหม่ เช่น ข้อมูลภูมิสารสนเทศและ/หรือข้อมูลขนาดใหญ่

NSO ที่กำลังสนใจในการบูรณาการข้อมูลจากหลายแหล่ง ควรพิจารณาการพัฒนาหรือเสริมสร้างกลไกในการทำงานร่วมกันและการประสานงานกับหน่วยงานของรัฐ องค์กรพัฒนาเอกชน (non-governmental organisations: NGOs) ภาคเอกชน และสถาบันการศึกษา กลไกเหล่านี้อาจอยู่ในรูปแบบของข้อตกลงอย่างเป็นทางการกับฝ่ายที่เกี่ยวข้อง เช่น บันทึกความเข้าใจ (Memorandum of Understanding: MoU)⁹ หรืออาจหมายถึงกิจกรรมที่ส่งเสริมความสัมพันธ์ในระดับผู้บริหารระดับสูงและระดับผู้เชี่ยวชาญ เช่น การประชุม หรือความริเริ่มร่วมกัน

นอกจากนี้ NSO ควรดำเนินการเพิ่มเติมสำหรับการพัฒนาการใช้ข้อมูลเพื่อการบริหารสำหรับการผลิตสถิติทางการ ด้วยหลายประเทศได้เลือกที่จะใช้ประโยชน์จากข้อมูลเพื่อการบริหารที่มีอยู่เพิ่มขึ้น จึงทำให้ตอนนี้มีทรัพยากรมากมายที่ใช้งานได้¹⁰

การเสริมสร้างความสามารถเพิ่มเติมจำเป็นอย่างมากสำหรับทั้ง NSO และผู้มีส่วนได้ส่วนเสียของประเทศ นอกจากนี้ กิจกรรมการสร้างควมตระหนักรู้ยังมีความ

จำเป็นในการสื่อสารให้ผู้กำหนดนโยบาย สื่อ และภาคประชาสังคมทราบถึงประโยชน์ของการใช้และการบูรณาการ แหล่งข้อมูลที่มีอยู่เพื่อผลิตสถิติทางการ การสร้างวัฒนธรรมการใช้ข้อมูลจะทำให้มั่นใจได้ว่าจะนำข้อมูลที่บูรณาการมาใช้อย่างมีประสิทธิภาพสำหรับการตัดสินใจ ซึ่งนำไปสู่ผลลัพธ์ของนโยบายที่มีข้อมูลและมีประสิทธิภาพมากขึ้น

ในระดับภูมิภาค แนะนำให้สำรวจกลไกสำหรับการขยาย การสนับสนุนทางเทคนิคเกี่ยวกับกิจกรรมการบูรณาการข้อมูล ซึ่งอาจรวมถึงการพัฒนาหลักสูตร e-Learning และการจัดการฝึกอบรมเชิงปฏิบัติการระดับภูมิภาคเกี่ยวกับการบูรณาการข้อมูล การจำลองโครงการในประเทศอื่น การขยายกลองเครื่องมือที่มีอยู่ให้เป็นกลองเครื่องมือทั่วไปที่ใช้กับโครงสร้างข้อมูลต่าง ๆ และ/หรือสนับสนุน NSO ที่พัฒนาแล้วในการให้ความช่วยเหลือทางเทคนิคแก่ NSO ที่ต้องการความช่วยเหลือในภูมิภาค

ไม่สามารถปฏิเสธได้ว่า NSO ไม่สามารถตอบสนองความต้องการข้อมูลทั้งหมดด้วยตนเองอีกต่อไป ภูมิทัศน์ข้อมูล (data landscape) กำลังเปลี่ยนแปลง และเพื่อให้ทันกับการเปลี่ยนแปลงในโลกที่เปลี่ยนไปนี้ NSO ควรก้าวออกจากเขตความสะดวกสบาย (comfort zone) ของตนและสำรวจแนวทางใหม่ การใช้ประโยชน์จากความรู้และประสบการณ์ที่มีอยู่ และการเพิ่มขีดความสามารถทางสถิติ จะทำให้การเดินทางเป็นไปได้ด้วยความราบรื่นมากขึ้น

เอกสารอ้างอิง: <http://www.unescap.org/resource-series/stats-brief>

⁹ คู่มือแบบฉบับร่างที่พัฒนาโดย United Nations Statistics Division ที่ <https://unstats.un.org/capacity-development/admin-data/docs/mou-guide-and-template.pdf>.

¹⁰ คู่มือการความร่วมมือเกี่ยวกับการใช้ข้อมูลเพื่อการบริหารสำหรับสถิติที่ <https://unstats.un.org/capacity-development/admin-data/#>.

เอกสารอ้างอิง

Exploring new horizons: Nepal's pilot in implementing data integration approaches



Disclaimer: The designations employed and the presentation of the material in this Stats Brief do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. Where the designation “country or area” appears, it covers countries, territories, cities or areas. Bibliographical and other references have, wherever possible, been verified. The United Nations bears no responsibility for the availability or functioning of URLs. The opinions, figures and estimates set forth in this publication should not necessarily be considered as reflecting the views or carrying the endorsement of the United Nations. The mention of firm names and commercial products does not imply the endorsement of the United Nations.

This Stats brief is prepared by Jose Ramon Albert, Hossein Hassani, and Ram Hari Gaihre, the consultancy team of the project, under the supervision of Afsaneh Yazdani, Statistician, ESCAP Statistics Division, and Petra Nahmias, Chief of Population and Social Statistics Section, ESCAP Statistics Division. The authors prepared this Stats Brief based on the report titled “Strengthening the National Statistical Capacity to Implement Data Integration Approaches”.¹ The authors extend their gratitude to the National Statistics Office of Nepal, particularly Munni Kumari Chaudhary, Deputy Chief Statistician of the Planning and Human Resource Management Division, and Rishi Ram Sigdel, Director of the Training Section, for their generous support throughout the project. The authors are also grateful to Sovannaroth Tey for his invaluable input and contribution to the project. For further information on this policy brief, please address your enquiries to:

Rachael Beaven
Director, Statistics Division
Economic and Social Commission for Asia and the Pacific (ESCAP)
Email: stat.unescap@un.org

Tracking number: May 2023 | Issue No. 33

Photo credit: iStock-849234128.jpg

¹ <https://www.unescap.org/kp/2023/strengthening-national-statistical-capacity-implement-data-integration-approaches-pilot>.

Table of Contents

Table of Contents	4
Summary	5
Introduction	6
Nepal project	7
Enabling access to data	8
Pre-processing of data sources	8
Linkage of data sources and the potential output	9
Conclusion, limitations, and the way forward	10

List of boxes

Box 1 – Data sources involved in the pilot exercise	7
Box 2 – Data pre-processing main steps	8

Summary

Data integration approaches are becoming increasingly popular among official statisticians and data producers; however, more institutional and technical capacity is required to fully realise their potential. To strengthen the capacity of statistical systems in Asia and the Pacific, ESCAP has released the **Asia-Pacific Guidelines to Data Integration for Official Statistics**² and launched the **Data Integration Community of Practice**,³ a forum for sharing relevant knowledge and experience and learning. Furthermore, in 2022, ESCAP, in collaboration with the National Statistics Office (NSO) of Nepal, initiated a project to implement a tailored approach in Nepal to improve the capacity of the National Statistical System (NSS) to integrate data and produce better data for measuring Sustainable Development Goal (SDG) indicators. This Stats Brief provides an overview of the project and the technical advisory and capacity-building support provided to the NSO of Nepal and its data partners.

² <https://www.unescap.org/kp/2021/asia-pacific-guidelines-data-integration-official-statistics>.

³ <https://www.unescap.org/our-work/statistics/communities>.

Introduction

National Statistical Offices (NSOs) around the world are extending their boundaries in the production of statistics in response to the changing context in which they operate. Global commitments, such as the 2030 Agenda for Sustainable Development, and national development plans have significantly raised the demand for quality, timely, and granular statistics. NSOs are responding to budget constraints and declining response rates by moving away from traditional data collection methods, including censuses and sample surveys, towards making better use of administrative data as well as new data sources. With technological advancements and the availability of new data sources, NSOs now have more options at their disposal.

In Asia and the Pacific, according to the recent SDG Progress Report, 128 out of 231 SDG indicators have sufficient data availability. The report indicates that *the lack of sufficient data for 51 out of 169 targets calls for statistical systems in the region to redouble efforts to fill data gaps.*⁴ In their efforts to strengthen the production of official statistics, NSOs should explore new approaches. Data integration, which involves combining disparate data sources such as censuses, sample surveys, administrative records, and big data, is one pathway to bridging data gaps. Integration of data can shed light on obscure areas and generate insights that cannot be derived from single data sources alone.

Data integration appears to be promising; however, there are often challenges that need to be overcome to fully realise its potential, particularly in terms of institutional and technical capacities. Institutional capacity mainly refers to the availability of a) relevant data sources, b) a strong legal basis supporting the use of and

access to data and metadata for statistical purposes and legislation ensuring data privacy, c) appropriate mechanisms for collaboration with data holders (either from the public or private sector), d) ICT Infrastructure, e) human resources including expertise and skills.

Technical capacities are also a challenge which includes a) skills to deal with the lack of interoperability, in terms of lack of unique identifiers, differences in concepts, classifications, coverage, data formats, reference periods, etc., b) skills to handle data quality issues, such as missing data, erroneous values, and inconsistencies within and across data sources, and c) expertise in record linkage and statistical matching. NSO staff will also require to have an in-depth understanding and knowledge of the relevant data sources, which can be obtained from appropriate metadata as well as regular interactions with data holders at the expert level. Some of the other skills that NSOs will need include soft skills, like negotiation, relationship building, and communication.

Although the capacities required for data integration appear extensive, NSOs may already have these skills and experience in data integration at the macro-level, such as in the compilation of the National Accounts, and they may be able to build on these for micro-level data integration, where capacity is limited. A regional Data Integration Capacity Assessment Survey conducted in 2020 by ESCAP revealed a significant capacity gap, with only 45 per cent of responding countries holding the required skills, as well as a huge demand for capacity building expressed by 77 per cent of responding countries, including Nepal.

⁴ <https://www.unescap.org/kp/2023/asia-and-pacific-sdg-progress-report-2023>.

Nepal project

In June 2022, ESCAP initiated a project titled “Strengthening the National Statistical Capacity to Implement Data Integration Approaches to Produce and Improve Data for Measuring SDG Indicators”. This project aimed to collaborate closely with the NSO of Nepal on a practical exercise to pilot integrating data from multiple sources, including census, surveys and administrative data. The project intended to develop the capacity and provide the technical assistance and IT toolbox to implement the full process of a data integration exercise.

The first step was to identify a data integration exercise that could be completed within the project’s time and resource constraints, taking into

consideration the country’s own priorities and available and accessible data sources. In this regard, a desk study was conducted, and views of national stakeholders were collected through an inception workshop and multiple discussions.

The defined exercise aimed to integrate relevant data sources to enable calculation of the a) proportion of the population covered by social protection floors/systems by poverty status, which will improve data for monitoring SDG Indicator 1.3.1⁵ and b) produce relevant statistics to be used for validating the poverty status of households. The exercise involved the integration of four specific data sources described in [Box 1](#).

BOX 1 – DATA SOURCES INVOLVED IN THE PILOT EXERCISE

- *Beneficiaries of Social Security Allowance 2022 (BSSA)* is a registry developed and maintained by the Department of National ID and Civil Registration, including information on program beneficiaries and social security allowances for vulnerable segments of the population.
- *Poor Households Identification Survey 2013 (PHIS)* was conducted as a large-scale survey by the Ministry of Land Management, Cooperatives and Poverty Alleviation to identify the poor for providing targeting services.
- *National Population and Housing Census 2011 (NPHC)* collected detailed information on demography, housing and asset ownership.
- *Nepal Living Standards Survey 2010/11 (NLSS)* is a multi-topic survey conducted by the NSO to produce statistics on consumption and other information on correlates of poverty.

Technical and ICT assistance was provided to facilitate accessing, preparing and linking the data sources as well as generating the specified statistics. These were followed by a training

workshop in Kathmandu in February 2023⁶ to enhance the technical capacity in data integration approaches, in general, and in relation to the defined exercise.

5 SDG Indicator 1.3.1: Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable.

6 <https://www.unescap.org/events/2023/national-training-workshop-implementing-data-integration-approaches-official-statistics>.

Enabling access to data

A key requirement for an NSO to carry out data integration is to obtain access to the required data and metadata. In Nepal, the legal basis supporting the use of and access to data and metadata for statistical purposes as well as the legislation respecting confidentiality and privacy, do exist. The NSO also had an arrangement, through official letters, with the Department of National ID and Civil Registration and the Ministry of Land Management, Cooperatives and Poverty Alleviation to access their data for statistical purposes.

However, the relevant data holders were reluctant to share their 'micro-level data' due to privacy concerns, as well as uncertainty about the feasibility of the project. To resolve the concerns and encourage collaboration, multiple meetings were organised among the NSO, data holders and the project team, in addition to the project's inception workshop. The NSO's strong

commitment to protecting privacy and confidentiality in accordance with the Statistics Act⁷ and the Fundamental Principles of Official Statistics⁸ was underscored during these meetings. Furthermore, the project's purpose, process, methodology, and potential outputs were communicated with data holders and mutual benefits were explored.

To make the data integration exercise more manageable and to encourage data sharing from data holders, the scope of the exercise was decided to be limited to one province, Gandaki, which compared to other provinces, contained more households/persons across all data sources. Furthermore, because microdata from the most recent population census and Nepal's Living Standard Survey were unavailable, it was decided to utilise older databases. All these efforts enhanced the willingness of the data holders to share data.

Pre-processing of data sources

Integrating data across disparate sources is challenging due to the issues stemming from the need for unique identifiers, interoperability, and data quality. Hence, pre-processing is regarded as a critical step in every data integration exercise.

Pre-processing entails cleaning and standardising data to ensure that the data used for integration is accurate, complete, and consistent across different sources and it is ready for record linkage. [Box 2](#) presents the main steps of pre-processing.

BOX 2 – DATA PRE-PROCESSING MAIN STEPS

- *Data cleaning* involves identifying and correcting errors, inconsistent values, missing values, incorrect data formats, misspelling, duplications, etc.
- *Data transformation* refers to the mapping of administrative units and/or variables into statistical units and/or variables. It also involves converting identification numbers into record identity numbers to protect privacy.
- *Data standardisation* entails ensuring consistency across data sources in terms of variable labels, values, coding, formats, spelling, etc. This step also involves identifying common variables across different sources and standardising them to a common format.

⁷ <http://rajpatra.dop.gov.np/welcome/book/?ref=25085>.

⁸ <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

Some of the data issues encountered in Nepal's exercise included: a) the lack of a unique identification number; b) differences in transcribing names (first, middle, and last names); c) differences in reference periods; and d) differences in classifications and categories for geographical as well as some substantive variables, such as caste and ethnic groups.

Pre-processing was conducted using a semi-automated approach, as some issues required manual resolution. The issues related to the translation of names from Nepali to English were resolved by thoroughly examining the data using

Nepali naming conventions. There were also problems with the spelling of names, which were corrected manually. Standardised coding was developed to ensure consistent attribute labels and categories across all databases. The difference in reference periods was addressed by adjusting for the difference in ages.

Due to the lack of a unique identifier across data sources, it was decided to use probabilistic matching based on names, location, and demographic variables, as appropriate.

Linkage of data sources and the potential output

Record linkage refers to the identification and combination of records corresponding to the same individuals across two or more data sources. There are two main approaches to record linkage, deterministic and probabilistic matching.

Deterministic matching involves comparing records in two datasets to find an exact match based on a unique identifier. In contrast, probabilistic matching uses statistical methods to determine the likelihood that two records refer to the same entity. While deterministic matching is straightforward, it can result in missed matches if the criteria are too strict or false matches if the criteria are too lenient. Probabilistic matching, although more complex, can result in more accurate matches by assigning weights to different matching criteria based on their importance and calculating a probability score for each potential match. Overall, the choice of approach depends on the specific requirements of the application, including the quality of the data, the level of accuracy required, and the resources available.

In Nepal's exercise, due to the lack of a unique identification number across data sources, deterministic matching was not possible, so probabilistic matching methods had to be applied, using common variables. To facilitate the record linkage, an IT toolbox was developed, which provided the user with the required features to explore data sources, set the rules and run the linkage process.

The toolbox also generated specified statistics. In this exercise, the toolbox provided cross-tabulations of variables from two different data sources. For instance, a table of the type of social security allowance from BSSA versus poverty status from PHIS was provided, which enabled the calculation of the proportion of the population covered by the social protection system based on poverty status. Also, wealth quintiles from the NPHC were cross-tabulated versus the poverty status from PHIS. This table enabled estimating the percent of those households in the bottom quintile of the wealth index that had been identified as poor in PHIS, allowing validation of the poverty status calculated in PHIS.

Conclusion, limitations, and the way forward

The exercise created a prototype for future data integration efforts in Nepal. It demonstrated the integration of multiple data sources, including the population census, sample surveys and administrative data. The exercise covered all steps involved in a data integration activity, from setting the objective for data integration to accessing data sources, pre-processing data, linking records and finally generating new statistics within a single province, Gandaki. While the exercise was successful in many ways, some limitations should be acknowledged, such as that more recent data would have been more useful and policy-relevant.

The capacity that has been built in NSO and project stakeholders enables them to expand the scope of the pilot exercise to the entire country, as well as to initiate other similar data integration projects. Although the current exercise focused on integrating data from census, survey, and administrative data sources, the NSO may also wish to explore the use of new data sources, such as geospatial information and/or big data.

NSOs interested in integrating data from multiple sources should consider developing or strengthening mechanisms for collaboration and coordination with other government agencies, non-governmental organisations (NGOs), the private sector, and academia. These mechanisms can take the form of formal agreements with relevant parties, such as a Memorandum of Understanding (MoU),⁹ or they can refer to activities that strengthen relationships at the senior management and expert levels, such as forums, regular meetings, or joint initiatives.

The NSOs should also take further actions to advance the use of administrative data in producing official statistics. With many countries opting to further benefit from available administrative data, a wealth of resources is now available.¹⁰

Further capacity-building is required both for NSOs and their national stakeholders. Additionally, awareness-raising activities are necessary to communicate to policymakers, the media, and civil society the benefits of using and integrating available data sources for producing official statistics. Building a culture of data use would ensure that integrated data is effectively utilised in decision-making, leading to more informed and effective policy outcomes.

At the regional level, it is recommended to explore mechanisms for expanding the technical support on data integration activities. This may include developing e-Learning courses and conducting regional training workshops on data integration, replicating the project in other countries, expanding the current toolbox into a general toolbox applicable to various data structures, and/or encouraging more developed NSOs to provide technical assistance to NSOs in need of assistance in the region.

This is unquestionable that NSOs can no longer meet all data demands in isolation. The data landscape is evolving, and to catch up with this changing world, NSOs should step out of their comfort zone and explore new horizons. Benefitting from available knowledge and experiences and enhancing statistical capacity makes the journey smoother.

9 See the draft template developed by United Nations Statistics Division at <https://unstats.un.org/capacity-development/admin-data/docs/mou-guide-and-template.pdf>.

10 See the inventory of the Collaborative on the Use of Administrative Data for Statistics at <https://unstats.un.org/capacity-development/admin-data/#>.

Get connected. Follow us.



www.unescap.org



[instagram.com/unitednationsescap](https://www.instagram.com/unitednationsescap)



[facebook.com/unescap](https://www.facebook.com/unescap)



[youtube.com/unescap](https://www.youtube.com/unescap)



[twitter.com /unescap](https://twitter.com/unescap)



[linkedin.com/company/united-nations-escap](https://www.linkedin.com/company/united-nations-escap)

Previous issues of Stats Brief: <http://www.unescap.org/resource-series/stats-brief>