تجنب الإفشاء والتعداد السكاني

موضوعات منتقاة حول التعدادات السكانية الدولية[1]

بقلم Sam Dupre

صدر في أكتوبر 2020

مقدمة

يتمثل أحد أهم الأدوار التي تلعبها المكاتب الإحصائية الوطنية (NSOs، بحسب الاختصار في اللغة الإنجليزية) في تنفيذ التعداد الوطني للسكان والمساكن. للقيام بهذا، تلتزم المكاتب الإحصائية الوطنية بمهمتين للإشراف على البيانات واللتين قد تتعارضان بشكل مباشر. الإشراف الجيد على البيانات وينطوي على كل من حماية خصوصية المشاركين الذين عهدوا إلى المكاتب الإحصائية الوطنية بالمعلومات الخاصة بهم، إضافة إلى نشر بيانات تعداد سكاني دقيقة ومفيدة للعامة. حال رغب المكتب الإحصائي الوطني في مشاركة العامة وقيامهم بتوفير إجابات دقيقة في أثناء التعداد السكاني، يجب أن يؤمن العامة بأنه سيتم الاحتفاظ بإجاباتهم بشكل آمن.

تناقش هذه المذكرة الفنية ثلاثة أنواع من أحداث الإفشاء، وأشكال الهجمات التي قد تحدث، ومراحل تجنب الإفشاء الأربعة—متتالية بوجه عام—التي قد يتبعها المكتب الإحصائي الوطني في أثناء التعداد السكاني لضمان الإشراف الجيد على البيانات.

المفاهيم الأساسية

أنواع البيانات

هناك نوعان من البيانات التي قد يصدرها المكتب الإحصائي الوطني للعامة: البيانات الجزئية والبيانات الإجمالية. البيانات الجزئية هي مجموعة من الإجابات لوحدة الملاحظة (البيت في حالة تعداد السكان والمساكن). عادة ما يتم إصدار البيانات الجزئية للعامة في صورة عينة صغيرة لكل بيانات المشارك. البيانات الإجمالية هي ملخص المعلومات لمجموعات كاملة من الأفراد في صورة أعداد التكرارات أو أرقام الحجم (مثال، المتوسطات، أو النطاقات، أو غيرها من الإحصاءات الموجزة).

للحفاظ على ثقة العامة، يجب ألا يقوم المكتب الإحصائي الوطني بإصدار مجموعة بيانات التعداد السكاني الكاملة غير المعدلة قبل مرور وقتٍ كافٍ بحيث لا تصبح الخصوصية ذات صلة بالموضوع (UNECE-CES, 2015). كمثال على هذا الأمر، تقوم الولايات المتحدة بإصدار ملفات بيانات التعديد السكاني فقط بعد مرور 72 سنة (USCB, n.d.).

أشكال الإفشاء

هناك ثلاثة أشكال رئيسية للإفشاء. يمثل كل منها مستوى مختلفًا من الخطر على المشاركين (Buron and Fontaine, 2018; UNSD, 2015). *إفشاء الهوية* يحدث عندما تكون هوية المشارك مرتبطة بشكل مباشر بسجل البيانات المنشورة (مثال، عبر الاسم، أو العنوان، أو رقم تحديد الهوية، أو البصمة، أو عنوان البريد الإلكتروني، أو رقم الهاتف). *إفشاء السمات* يحدث عندما تتسبب القيم الموجودة في البيانات المنشورة في إفشاء سمات أخرى للفرد. *إفشاء استنتاجي* (إضافة إلى ما يلحق ذلك من إعادة التشكيل وإعادة تحديد الهوية) يحدث عند استخدام البيانات المنشورة لاستنتاج قيم خاصة بالمشاركين استنادًا إلى القيم الإحصائية للبيانات الصادرة. تختلف احتمالية حدوث الإفشاء وتكراره استنادًا إلى مادة النشر (McKenna and Haubach, 2019). على سبيل المثال، التقرير الذي يوضح وجود كل أعضاء مجموعة عرقية ما في منطقة جغرافية واحدة يكون عرضة بشكل أكبر لخطر إفشاء السمات، لكن أقل عرضة لخطر إفشاء الهوية. وهذا لأن منفذ الهجوم يمكنه تحديد المنطقة التي يعيش فيها الشخص إذا علم بأن هذا الشخص أحد أعضاء تلك المجموعة العرقية. تختلف أساليب الهجوم، لكن تم تضمين بعض من أكثرها شيوعًا في المربع 1.

[1] تُمثل هذه المذكرة الفنية جزءًا من سلسلة *موضوعات منتقاة حول التعدادات السكانية الدولية* (STIC، بحسب الاختصار في اللغة الإنجليزية)، والتي تتناول بعمق المسائل التي تثير اهتمام المجتمع الإحصائي على الصعيد الدولي. ويقوم مكتب الإحصاء الأمريكي بدور فاعل في دعم الدول لتطوير أنظمتها الإحصائية القومية، مستثمرًا في تنمية القدرات والارتقاء بالخبرات الإحصائية على نحو يحقق الاستدامة والتطور المستمرين. أي وجهات نظر واردة خاصة بالمؤلف (المؤلفين) ولا تخص بالضرورة مكتب الإحصاء الأمريكي.

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

United States Agency for International Development
www.usaid.gov

United Nations Population Fund
www.unfpa.org

المربع رقم 1.

أشكال الهجوم

خطوات الهجوم

إعادة تحديد الهوية: تتم مطابقة سجل بمصدر بشري.

إعادة التشكيل: يتم إلغاء إخفاء هوية القيم التي تم إخفاء هويتها لكل سجل.

لا تكون هذه الخطوات متتالية بالضرورة، حيث يمكن استخدام القيم التي تمت إعادة تشكيلها لإعادة تحديد المصادر أو العكس.

الهجمات الشائعة

هجوم إعادة تشكيل قاعدة بيانات: مطابقة حقل ("المفتاح") في مجموعة بيانات تم إخفاء هويتها بحقل عام في مجموعة بيانات عامة.

هجوم التتبع: محاولة العثور على البيانات الهدف ضمن مجموعة بيانات. استنادًا إلى مدى شمولية مجموعة البيانات أو حساسية الموضوع، حتى مجرد تحديد ما إذا كان الشخص مضمنًا، قد يمثل انتهاكًا خطيرًا للخصوصية أو يتسبب في ظهور فرصة لمزيد من إعادة التشكيل.

هجوم عبر الاختلاف: أحد أكبر المخاطر المتعلقة بالبيانات المكانية. يعمل عبر استخدام الاختلافات في الاستعلامات المتكررة لمعرفة معلومات حول السجلات من خلال مقارنة مجموعات البيانات.

ملاحظة: تم تجميع المعلومات من Dwork et al., 2017؛ و McKenna and Haubach, 2019: وUNSD, 2015.

عملية تجنب الإفشاء

تتكون عملية تجنب الإفشاء عادة من أربع مراحل: (1) تقييم المخاطر، و(2) التشاور مع العامة، و(3) ضوابط الإفشاء، و(4) الأرشفة والوصول/الإصدار. تبدأ المراحل قبل إجراء التعداد السكاني وتستمر حتى بعد اكتمال الجدول الزمني للتعداد السكاني الرئيسي. الجدول 1 يشرح بشكل تفصيلي المراحل الأربع ومراحلها الفرعية.

تستخدم المكاتب الإحصائية الوطنية ثلاث إستراتيجيات لحماية خصوصية المشارك عبر كل مرحلة من تلك المراحل. حيث تقوم بالآتي: *تقييد الجمع*، عبر الحد من جمع البيانات الخاصة بالموضوعات الحساسة حيثما أمكن (NASEM, 2017; UNSD, 2015).

و*تقييد البيانات*، عبر التحكم في مكونات البيانات التي يتم إصدارها وشكل الإصدار (مثال، بيانات إجمالية أو بيانات جزئية) والضوابط الإحصائية التي يتم تطبيقها. وأخيرًا، *تقييد الوصول*، عبر التحكم في المستخدمين الذين يمكنهم الوصول إلى البيانات إضافة إلى مستوى الوصول المعين لكل منهم.

الجدول 1 يمثل نظرة عامة حول هذه المراحل. لا يُقصد بذلك الشرح التفصيلي لكل التفاصيل الفنية الدقيقة، بل توفير نظرة تمهيدية حول الاعتبارات الفنية للعملية. نظرًا إلى تعقيد تلك الأنشطة وأهميتها الكبرى، سنقوم بعد هذه النظرة العامة بتوفير معلومات إضافية حول تقييم المخاطر ما بعد العد، ومراحل التحكم الإحصائي، والوصول إلى البيانات المؤرشفة أو إصدارها.

الجدول 1

مراحل تجنب الإفشاء

المرحلة	المرحلة الفرعية	ما يستلزم ذلك
تقييم المخاطر	تقييم داخلي	في مرحلة مبكرة من التخطيط للتعداد السكاني، مراجعة المخاطر استنادًا إلى نوع وحساسية البيانات التي يتم جمعها وإصدارها.
	تقييم خارجي	بعد إجراء التقييم الداخلي—لكن قبل العد—تعيين استشاريين خارجيين لإجراء تقييم مخاطر مستقل.
	تقييم داخلي ثانٍ	بعد تطبيق ضوابط العد والإفشاء، تكرار تقييم المخاطر الداخلي بما في ذلك مراجعة كمية للبيانات المجمعة.
	لجنة مراجعة الإفشاء	مطالبة لجنة مراجعة، قبل إجراء التعداد السكاني، وفي أثنائه، وبعد إكماله بإجراء تقييمات مخاطر لخطط نشر البيانات الجديدة. ينبغي لهذه اللجنة مراجعة المواد المنشورة، حيث تتحسن التقنيات.
التشاور مع العامة	N	قبل إجراء التعداد السكاني، يجب التشاور مع أصحاب المصالح بشأن مخاوف الخصوصية لديهم واحتياجات البيانات الخاصة بهم. يجب أن تغطي البيانات نوع البيانات التي يرغبون في إصدارها والتنسيق المرغوب. استخدام هذه المعلومات لتوجيه تقييم المخاطر الخاص بالمكتب الإحصائي الوطني منذ بدء عملية التخطيط واستهداف مجموعات السكان الذين يصعب تعدادهم تاريخيًا (انظر دليل مكتب الإحصاء الأمريكي حول عد السكان الذين يصعب عدهم في التعداد السكاني [2019a]). يمثل المربع 2 دراسة حالة حول كيفية قيام مكتب الإحصاءات الوطنية التابع للمملكة المتحدة بالتعامل مع مرحلة التشاور مع العامة.
ضوابط الإفشاء	ضوابط قانونية	قبل إجراء التعداد السكاني، فرض تشريعات قانونية تلزم المكتب الإحصائي الوطني بمسؤولية حماية بيانات المشارك، وخصوصًا تخطيط كيف يمكن إصدار البيانات. يوفر القيام بذلك، سندًا قانونيًا يُقر عملية اتخاذ القرار الخاصة بالمكتب الإحصائي الوطني.
	ضوابط مادية	قبل العد، فرض سياسات للتخلص من المواد، ودخول المنشآت، وكيفية التعامل مع العينة الممثلة المحفوظة. تشمل عملية التخلص النماذج الورقية ومسح بيانات جهاز إجراء المقابلة الشخصية بمساعدة الكمبيوتر.
	ضوابط فنية	قبل العد، فرض سياسات تمنع مقاطعة عملية الإجابة عن التعداد السكاني على الإنترنت، وتؤمن بيانات جهاز إجراء المقابلة الشخصية بمساعدة الكمبيوتر المفقودة، وتفرض إجراءات الأمان على شبكة المكتب الإحصائي الوطني، وتتحكم في وصول الموظفين إلى بيانات المشارك.
	ضوابط إحصائية	بعد العد، تطبيق تدابير إحصائية على البيانات الجزئية للمشارك (قبل التنسيق الجدولي) أو على البيانات الإجمالية (بعد التنسيق الجدولي). تعتمد التدابير المحددة المستخدمة على شكل الإصدار المخطط.
الوصول إلى/إصدار البيانات المؤرشفة	N	بعد التعداد السكاني، أرشفة البيانات الجزئية (الملفات الأولية والملفات المعدلة بعد تطبيق الضوابط الإحصائية)، والبيانات التعريفية، والبيانات الوصفية وإتاحتها لأصحاب المصالح. ارجع إلى دليل أرشفة بيانات التعداد السكاني والحفاظ عليها الخاص بمكتب الإحصاء الأمريكي للحصول على معلومات تفصيلية حول أرشفة البيانات الآمنة (2019b).

N لا ينطبق.

ملاحظة: تم تجميع المعلومات من Lauger et al., 2014؛ وMcKenna and Haubach, 2019؛ وNASEM, 2017؛ وUNECE-CES, 2015؛ وUNSD, 2015.

معلومات إضافية حول تقييم المخاطر ما بعد العد، والضوابط الإحصائية

يعد تقييم المخاطر والضوابط الإحصائية من أحد الجوانب الفنية الأكثر تعقيدًا للتحكم في الإفشاء (الجدول 1). تستند التدابير المحددة التي يجب استخدامها إلى:

- شكل الإصدار (مثال: عينات من البيانات الجزئية لاستخدام العامة [PUMS] في مقابل البيانات الإجمالية) (McKenna and Haubach, 2019).
- مستوى التفاصيل المخطط (ملف PUMS الذي يشتمل على مجموعات بيانات عامة قد يتطلب حدًّا أدنى من السكان يصل إلى 100,000، بينما الملف الذي يشتمل على بيانات شديدة التفصيل قد يتطلب حدًّا أدنى من السكان يصل إلى 400,000) (Buron and Fontaine, 2018).

بعد العد، قد تقوم المكاتب الإحصائية الوطنية بتكرار تقييم المخاطر استنادًا إلى البيانات المجمعة—مع أخذ خطط الإصدار المحدثة في الحسبان—وتطبيق تدابير التحكم الإحصائي. تكون هذه العملية المتكاملة المكونة من خمس خطوات كالآتي:

المربع رقم 2.

دراسة حالة: مكتب الإحصاءات الوطنية (ONS، بحسب الاختصار في اللغة الإنجليزية) التابع للمملكة المتحدة

عند الإعداد لإجراء التعداد السكاني لسنة 2021، بين 2015 و2018، سعى مكتب الإحصاءات الوطنية إلى القيام بجولات متعددة من التشاور مع العامة حول الموضوعات والمطالبات من التعداد السكاني لسنة 2021. بعد كل جولة نشر مكتب الإحصاءات الوطنية معلومات تفصيلية حول (1) الخطط الأولية، و(2) إجابات العامة، و(3) خطط مكتب الإحصاءات الوطنية استنادًا إلى تلك الإجابات، و(4) استنادًا إلى نتائج تغيير الخطط لتمثيل متسالٍ في التعداد السكاني لسنة 2021.

المصدر: ONS, 2018.

الخطوة 1: استبعاد المعلومات الشخصية الحساسة (PII، بحسب الاختصار في اللغة الإنجليزية)

إزالة محددات الهوية المباشرة مثل الاسم، والعنوان، وأي أرقام تحديد هوية حكومية من السجلات لمنع إفشاء الهوية المباشر (UNSD, 2015).

الخطوة 2: تحديد السجلات، والخلايا، والفئات الحساسة

بينما يتطلب تقييم مخاطر الإفشاء الإحصائي خبرة نوعية في المجال لتحديد الموضوعات والمجموعات الحساسة/سريعة التأثر محليًا (UNSD, 2015)، توجد تدابير كمية لتقييم خطر الإفشاء. يتيح استخدام تدابير كمية إجراء مقارنة واضحة بين خيارات النشر المختلفة وتوفير سند قانوني مبرر لعملية اتخاذ القرار في المكتب الإحصائي الوطني (NASEM, 2017).

يمثل الجدول 2 مجموعة من التحديات الشائعة التي قد يواجهها المكتب الإحصائي الوطني عند تحديد السجلات، والخلايا، والفئات الحساسة، إضافة إلى إرشادات حول أساليب التقييم الكمي إذا كان ينبغي تمييز ذلك الموضوع الحساس للخضوع لتدابير التحكم الإحصائي

الخطوة 3: مواجهة الخطر

يمكن أن تكون الضوابط الإحصائية *اضطرابية* أو *غير اضطرابية* (Antal et al., 2017). تبدل التدابير الاضطرابية البيانات بشكل طفيف وبطرق متحكم بها، وتغير بنية البيانات بأقل درجة ممكنة. تعمل التدابير غير الاضطرابية على إزالة (أو تجميع) خلايا الجدول، أو المناطق الجغرافية، أو سجلات البيانات التي تمثل مستويات معينة من الخطورة. تميل الأساليب الاضطرابية إلى الحفاظ على بنية البيانات بشكل أكثر موثوقية وتتسبب في فقد معلومات أقل مقارنة بالأساليب غير الاضطرابية (Antal et al., 2017).

الجدول رقم (2)
ميزات شائعة تؤدي إلى سجلات، وخلايا، وفئات حساسة

التحدي	سبب عده من التحديات	تقييم هذا الخطر كميًا
وجود خلايا تحتوي على أعداد قليلة.	يتزايد خطر إفشاء الهوية عند وجود عدد صغير جدًا من السجلات داخل مجموعة.	تمييز جميع الوحدات الأقل من الحد القياسي. بالنسبة إلى استبيان المجتمعات المحلية في الولايات المتحدة الأمريكية (ACS) التابع لمكتب الإحصاء الأمريكي وملف PUMS للتعداد السكاني لسنة 2010: • يجب أن تشتمل كل فئة في المتغير الصريح على 10000 من الأشخاص أو البيوت غير المرجحة على الأقل. • يجب أن تشتمل جميع المناطق الجغرافية (بما في ذلك الحضرية/الريفية) على 50 من الأشخاص أو البيوت غير المرجحة على الأقل للمتغير الواحد. • تتطلب الجدولة متوسط حجم خلية يسع ثلاث حالات غير مرجحة على الأقل.
وجود أعداد غير صفرية لمجموعات حساسة.	حتى العلم بوجود أشخاص ذوي خصائص معينة قد يؤدي إلى إفشاء الخصوصية.	تمييز جميع الخلايا للخصائص الحساسة أو مجموعات الخصائص المحددة سابقًا.
مجموعات فرعية مختلفة من النتائج تتضمن فئات السكان نفسها.	يمكن مقارنة هذه المجموعات الفرعية في الهجوم عبر الاختلاف لاستنتاج بيانات المشارك.	تمييز المناطق الجغرافية غير المتداخلة ومجموعات المشاركين للمراجعة الإضافية، وإيلاء اهتمام خاص بأي حالات حيث توجد فقط اختلافات طفيفة بين المجموعات الفرعية المتكررة للسكان.
يتم تمييز الأفراد داخل البيت بالحالة "عرضة للخطر".	إذا كان أحد الأفراد يمثل خطر الإفشاء، فيمكن أن يكون البيت بالكامل عرضة للخطر.	تقييم الخطر على المستوى الفردي لكل متغير ومستوى جغرافي. تجميع الأفراد لتكوين بيوت وتمييز أي بيت بالحالة "فرد عرضة للخطر".
وجود قيم خارجة ضمن الإجابات لأي متغير.	يكون هؤلاء الموجودون أعلى أو أسفل توزيع الإجابات أسهل في التحديد مقارنة بهؤلاء الذين تكون إجاباتهم قريبة من المتوسط.	للمتغيرات المستمرة، تمييز السجلات المشتملة على قيم قريبة من الحد الأقصى والحد الأدنى من التوزيع. بشكل نموذجي، قد يشمل هذا النسبة المئوية الـ 0.5 العلوية/السفلية من كافة القيم (أو نسبة مئوية 3 من كل القيم غير الصفرية إذا كان هذا سيشمل مزيدًا من السجلات).
	من المرجح أن يكون المشارك (المشاركون) الذي يتميز بأكبر قدر مقيد من القيم في إحدى المجموعات هو الأكثر عرضة للخطر من المشاركين الآخرين في طرفي تلك المجموعة.	تقوم القاعدة (n, k)، والقاعدة p%، والقاعدة pq بتمييز الحالات حيث يكون بإمكان المشاركين ذوي القيم الخارجة تحديد مشاركين آخرين ذوي قيم خارجة استنادًا إلى إجاباتهم. تقوم القاعدة (n, k) بتمييز متغير إذا كانت القيم الخاصة بالمشاركين ذوي n الأكبر تشكل نسبة k على الأقل من إجمالي القيم. وتقوم القاعدتان p%، وpq بتمييز الحالات حيث يكون بإمكان مشاركين آخرين تقدير القيم الخاصة بالمشارك ذي القيمة الأكبر ضمن نطاق النسبة المئوية p للقيمة الحقيقية. عادة ما تكون القيم المحددة التي يستخدمها المكتب الإحصائي الوطني للإشارة إلى n، أو k، أو p، أو q سرية نظرًا إلى أنه حتى هذه المعلومات عرضة لخطر هجوم الإفشاء الاستنتاجي.

ملاحظة: تم تجميع المعلومات من Antal et al., 2017؛ وBuron and Fontaine, 2018؛ وLauger et al., 2014؛ وMcKenna and Haubach, 2019؛ وOECD, 2005.

غير الاضطرابية

إخفاء أولي وثانوي/إضافي. يعمل الإخفاء الأولي على الحماية من إفشاء الهوية/السمات عبر استبدال الخلايا أو السجلات بعلامة تشير إلى إخفائها أو إظهارها في صورة "لا توجد بيانات" (Antal et al., 2017). ينطوي الإخفاء الثانوي على إخفاء خلايا إضافية غير مميزة بحيث يتعذر استخراج القيم المخفية عبر الإفشاء الاستنتاجي. عوضًا عن ذلك، يمكن إخفاء كل المتغيرات التي تنطوي على مشكلات أو المجموعات أو المناطق الجغرافية المميزة بشكل كامل من النشر (UNECE-CES, 2015).

التسجيل. عند وجود عدد قليل جدًا من السجلات لقيمة ما أو مجموعة من القيم، يمكن دمجها مع مجموعات، أو سجلات، أو أعمدة، أو صفوف أخرى حتى يتم الوصول إلى الحد المعين. عندما يكون من الممكن ربط البيانات المتاحة للعامة ببيانات التعداد السكاني، ربما يكون التسجيل ضروريًا لمنع إفشاء السمات أو الإفشاء الاستنتاجي حتى عند الإيفاء بالحد القياسي بالفعل. تشمل خيارات تسجيل البيانات الكمية التقريب، أو الاستيفاء داخل نطاق/توزيع محدد سابقًا، أو تقليل الكمية لتقليل تحديد البيانات (Dajani et al., 2017). يعد الترميز العلوي، والترميز السفلي من أشكال التسجيل المستخدم لإخفاء القيم الخارجة للمتغيرات المستمرة. يتم استبدال القيم الخارجة في الحد العلوي أو السفلي من الشريحة المئوية بقيمة مقتطعة، أو بالوسط أو الوسيط لكل القيم العلوية/السفلية التي تم ترميزها.

الاضطرابية

إضافة تشويش. تتم إضافة تشويش عشوائي إلى الخلايا المحفوفة بالمخاطر عبر إجراء تغييرات طفيفة على القيم الأصلية. يحافظ التشويش المضاف على بنية البيانات لهذا المتغير عبر التحكم في التحيز، والتباين، والترددات العددية، وضبط الخلايا ذات القيمة الصفرية (Antal et al., 2017).

تبديل السجلات، وتبديل التصنيف، والخلط. ينطوي تبديل السجلات على مطابقة مجموعات السجلات حول بعض المعايير ثم استبدال القيم غير المتماثلة بين تلك المجموعات (Antal et al., 2017). يجب أن تكون المنطقة الجغرافية الأصل للمجموعات التي تم تبديلها هي نفسها كلما أمكن، لتقليل الإخلال بالبيانات والحد من التحول الجغرافي (على سبيل المثال، التبديل داخل المناطق، لكن ليس بين المناطق (Buron and Fontaine, 2018)، على الرغم من أن هذا ليس هو الحال دائمًا عند وجود خطر إفشاء ضخم (Lauger et al., 2014). في تبديل التصنيف والخلط يتم تبديل القيم لبعض المتغيرات في السجلات التي تشتمل على قيم متشابهة لهذه المتغيرات.

بيانات اصطناعية. إنشاء نموذج إحصائي يصف مجموعة البيانات ثم استبدال السجلات الفريدة باستخدام قيمة نموذجية (Dajani et al., 2017). تتطلب مجموعات البيانات هذه أيضًا تقييم المخاطر حيث لا تزال حوادث الإفشاء عرضة للحدوث؛ بالرغم من ذلك، فهي تتيح للباحثين الوصول إلى البيانات التي قد تمثل خلافًا لذلك خطرًا مانعًا.

الخطوة 4: التحقق من النتائج

التحقق من الخطر المتكبد باستخدام التدابير الموضحة في الخطوة 2، والتحقق من مستوى فقدان المعلومات (مثال: تزايد التباين في تقدير المعلمة أو وجود تحيز). لتقييم فقدان المعلومات، يجب التحقق من الآتي:

- إذا تسبب أي اضطراب في تغيير ضخم في المتغيرات الصغرى/القصوى، أو الوسط/الوسيط/المنوال، أو الشرائح المئوية (الفروق المطلقة والنسبية) (Antal et al., 2017).
- نسبة الخلايا التي يتجاوز فيها الاضطراب مستوى التغيير المحدد سابقًا، حيث إن التغييرات الصغيرة في المناطق منخفضة الكثافة قد يكون تأثيرها أكبر من التغييرات الأكبر في المناطق مرتفعة الكثافة (Buron and Fontaine, 2018).
- إذا أضاف الاضطراب قدرًا ضخمًا من "النتائج الإيجابية الخاطئة" (اضطراب القيم الصفرية مع غير الصفرية) و"نتائج سلبية خاطئة" (اضطراب القيم غير الصفرية مع الصفرية) (Buron and Fontaine, 2018).
- إذا ما كانت العلاقات بين البيانات لا تزال موجودة بعد الاضطراب (مثال: التكافؤ أو عدم التكافؤ المتوقع—أو علاقة إحصائية محددة—بين اثنين من المتغيرات) (Antal et al., 2017).

الخطوة 5: إجراء دراسات داخلية حول الهجوم

بإمكان المكاتب الإحصائية الوطنية إجراء دراسات داخلية حول هجوم الإفشاء للكشف عن الثغرات الأمنية الجديدة مع ظهور تهديدات جديدة (McKenna and Haubach, 2019). يجب أن تستخدم دراسات الهجوم هذه الوسائل والتقنيات نفسها التي قد يستخدمها منفذ الهجوم، وتضمين مجموعات البيانات العامة والخاصة، والتطورات التقنية الجديدة. يجب تطبيق تلك الاختبارات على إصدارات البيانات الجديدة والقديمة للتأكد من أن السجلات التي تم إخفاء هويتها سابقًا لم تصبح عرضة لخطر الإفشاء.

معلومات إضافية حول الأرشفة والوصول/الإصدار

قد تمثل البيانات الجزئية (كل من الملفات الأولية والملفات المعدلة بعد إجراءات تجنب الإفشاء الإحصائي)، والبيانات التعريفية، والبيانات الوصفية مخاطر لإفشاء البيانات سواء بشكل منفرد أو مع مصادر أخرى (UNECE-CES, 2015). يجب على المكتب الإحصائي الوطني الاحتفاظ بالإصدارات الأصلية غير المعدلة من البيانات داخليًا، وإنشاء سجل بكل التعديلات، وتخزينه بشكل منفصل عن ملفات البيانات التي تم إخفاء هويتها (Van den Eynden et al., 2011). في بعض الأحيان، يمكن الاحتفاظ بعينة ممثلة من نماذج التعداد السكاني المكتملة بواسطة المكتب الإحصائي الوطني. في هذه الحالة، يجب تطبيق كل مبادئ خصوصية البيانات قبل أي عملية إصدار.

تشتمل بعض الترتيبات التي يمكن للمكاتب الإحصائية الوطنية استخدامها للحفاظ على نماذج آمنة للوصول على مناطق البيانات المعزولة أو منشآت يمكن الوصول إليها عن بُعد، أو قواعد بيانات على الإنترنت لطلب مجموعات بيانات أو تحليل البيانات، أو ترتيبات الترخيص للمستخدمين المعتمدين، أو إصدار ملف PUMS للعامة (FCoSM, 2005; Hundepool et al., 2012). بغض النظر عن الترتيب المعمول به، تُطبق المبادئ الشائعة. لأسباب أمنية، ينبغي للمكاتب الإحصائية الوطنية عدم السماح لمستخدمين خارجيين بالوصول إلى شبكات الإنترنت. يمكن تشفير البيانات التي لم تتم الموافقة على إصدارها للعامة قبل أي نقل للشبكات الداخلية الآمنة. يجب ألا تشتمل مناطق البيانات المعزولة على وصول إلى الإنترنت، أو شبكات خارجية، أو منافذ USB (UNSD, 2015). يجب تنشيط تدابير تجنب الإفشاء الإضافية تلقائيًا عند طلب جداول مماثلة مرات متعددة، للحماية من الهجمات عبر الاختلاف من خلال إنشاء مجموعات فرعية بشكل متكرر. يجب أن تنطوي الترتيبات على عمليات تدقيق غير معلنة لمنشآت تخزين البيانات، ومراجعة النواتج الإحصائية، وأن تغطي التخلص من البيانات والملفات المستخلصة. لضمان الفعالية، يجب أن تكون كافة الترتيبات ملزمة قانونيًا وتشمل عقوبات للانتهاك (FCoSM, 2005; UNSD, 2015).

خاتمة

تتناقص قيمة التعداد السكاني بشكل كبير من دون نشر بيانات قابلة للاستخدام في الوقت المناسب. ومع ذلك، تؤدي زيادة التفاصيل في إصدارات البيانات إلى زيادة الخطر المتمثل في احتمالية انتهاك خصوصية المشارك. يتزايد هذا الخطر في عصر البيانات الضخمة، حيث تعمل التطورات في أدوات استخراج البيانات، والإسناد الجغرافي للبيانات، وإمكانات معالجة البيانات الإحصائية، على زيادة احتمالية وقوع حوادث إفشاء البيانات. بإمكان المكاتب الإحصائية الوطنية الإيفاء بالتزاماتها العامة عبر إدارة الخطر باستخدام السياسات والإجراءات المزودة في هذه المذكرة.

المراجع

Antal, L., T. Enderle, and S. Giessing, *Harmonised Protection of Census Data in the ESS: Statistical disclosure control methods for harmonised protection of census data*, Eurostat Centre of Excellence on Statistical Disclosure Control, The Hague, 2017.

Buron, M. L., and M. Fontaine, Confidentiality of Spatial Data, in Loonis, V. and Marie-Pierre de Bellefon, *Handbook of Spatial Analysis: Theory and application with R*, chapter 14, Insee Méthodes No. 131, Eurostat, The Hague, 2018.

Council of Europe (CoE), *Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data*, Directorate General of Human Rights and Rule of Law, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Strasbourg, France, 2017.

Dajani, A.N., A.D. Lauger, P.E. Singer, D. Kifer, J.P. Reiter, A. Machanavajjhala, S.L. Garfinkel, S.A. Dahl, M. Graham, V. Karwa, H. Kim, P. Leclerc, I.M. Schmutte, W.N. Sexton, L. Vilhuber, and J.M. Abowd, The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau, in *September 2017 Meeting of the Census Scientific Advisory Committee*, Suitland, MD, 2017.

Dwork, C., A. Smith, T. Steinke, and J. Ullman, *Exposed! A Survey of Attacks on Private Data*, Annual Review of Statistics and Its Applications, 4(12): 1–24, 2017.

Federal Committee on Statistical Methodology (FCoSM), *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology Version 2*, U.S. Office of Management and Budget, Washington, DC, 2005.

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, E. Shulte Nordholt, K. Spicer, and P.P. de Wolf, Statistical Disclosure Control, In: *Wiley Series in Survey Methodology, Wiley*, Chichester, United Kingdom, 2012.

Lauger, A., B. Wisniewski, and L. McKenna, Disclosure Avoidance Techniques at the *U.S. Census Bureau: Current practices and research, research report series (Disclosure Avoidance #2014-02)*, Center for Disclosure Avoidance Research, U.S. Census Bureau, Washington, DC, 2014.

McKenna, L. and M. Haubach, *Legacy Techniques and Current Research in Disclosure Avoidance at the U.S. Census Bureau*, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC, 2019.

National Academies of Sciences, Engineering, and Medicine (NASEM), *Innovations in Federal Statistics: Combining data sources while protecting privacy*, The National Academies Press, Washington, DC, 2017, <https://doi.org/10.17226/24652>.

OECD, Glossary of Statistical Terms, <https://stats.oecd.org/glossary/detail.asp?ID=6943>, 2005, accessed on July 15, 2020.

Office for National Statistics (ONS), Initial View on 2021 Census Output Content Design: Response to consultation, Office for National Statistics, United Kingdom, 2018.

United Nations Economic Commission for Europe—Conference of European Statisticians (UNECE-CES), *Recommendations for the 2020 Censuses of Population and Housing*, United Nations Publications, New York, NY, 2015.

_____ , *Census Data Archiving and Preservation*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Archiving-Brief.pdf>, 2019b.

_____, The "72-Year Rule," <www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html>, n.d., accessed on July 15, 2020.

Van den Eynden, V., L. Corti, M. Woollard, L. Bishop, and L. Horton, *Managing and Sharing Data*, UK Data Archive, UK, 2011.

United Nations Statistics Division (UNSD), *Principles and Recommendations for Population and Housing Censuses, Revision 3*, United Nations Publications, New York, NY, 2015.

United States Census Bureau, *Counting the Hard to Count in a Census*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Hard-to-Count-Populations-Brief.pdf>, 2019a.

يتم نشر سلسلة موضوعات منتقاة حول التعدادات السكانية الدولية (STIC) بواسطة برامج دولية داخل قسم السكان التابع لمكتب الإحصاء الأمريكي. تتولى الوكالة الأمريكية للتنمية الدولية رعاية إنتاج سلسلة موضوعات منتقاة حول التعدادات السكانية الدولية، هذا فضلاً عن الدعم المزدوج الموجه إلى المنظمات الإحصائية التي تعد مصدر خبرات المؤلفين. كذلك يتعاون صندوق الأمم المتحدة للسكان فيما يتعلق بالمحتوى والنشر، بما يضمن وصول سلسلة موضوعات منتقاة حول التعدادات السكانية الدولية إلى نطاق أكثر اتساعًا من الجمهور.