Meta-Index | INSTITUTE FOR SURVEY RESEARCH

# البيانات الإثنوغرافية في عصر البيانات الضخمة: كيفيات المقارنة والتوليف

'أندرياس بيير-نيلسن'/ 'كريستوفر لينـد غلافيند'

ترجمة: محمد جليد

دراسات وأبحاث | 17 مارس 2023

## ملخص:

تمكّن البيانات الضخمة (Big Data) الباحثين من متابعة سلوك جماعات الأفراد الواسعة عن كثب، باستخدام الآثار الرقمية للتردد العالي. غير أن هذه الآثار الرقمية غالبا ما تفتقر إلى السياق، ولا يتّضح دائما ما يتم قياسه. في المقابل، تتابع المعلومات المستقاة من العمل الإثنوغرافي الميداني عددا محدودا من الأفراد، لكنها توفر السياقَ الذي غالبا ما تفتقر إليه البيانات الضخمة. ومع ذلك، ثمة إمكانات لم تكتشف بعد بما يكفي للجمع بين البيانات الإثنوغرافية ومصادر البيانات الرقمية الأخرى؛ إذ تقدم هذه الورقة طرقا يمكن أن تدمج بها تصاميم البحث الكمي البيانات الضخمة مع البيانات الإثنواغرافية، وتعلل أشكال التعاضد الذي قد يتيحه مثل هذا الدمج. نسلط الضوء على الاختلافات والتشابهات بين البيانات الإثنوغرافية والبيانات الضخمة، مع التركيز على الأبعاد الثلاثة: الأفراد وعمق المعلومة والوقت. ونحدد كيفية الموافقة على البيانات الضخمة عبر إتاحة 'حقيقة ميدانية' واستكمالها بتقديم 'وصف مكثف'. فضلا عن ذلك، نضع طرقا تجعل التحليل الذي تم إجراؤه باستخدام البيانات الضخمة يستفيد من التعاون مع الإثنوغرافيين، ونناقش احتمالات ذلك في حقول التعلم الآلي والاستدلال السببي.

[1] هذا المقال جزء من التيمة الأساسية حول أنثروبولوجيا الآلة. للاطلاع على اللائحة الكاملة لجميع المقالات حول هذه التيمة الأساسية، يرجى الولوج إلى الرابط التالي:
https://journals.sagepub.com/page/bds/collections/machineanthropology

[2] Andreas Bjerre-Nielsen، شعبة الاقتصاد، كلية العلوم الاجتماعية، جامعة كوبنهاغن، كوبنهاغن، الدنمرك.
[3] Kristoffer Lind Glavind، مركز كوبنهاغن لعلم البيانات الاجتماعية (SODAS)، كوبنهاغن، الدنمرك.
[4] باحث ومترجم وأستاذ مساعد بالنواة الجامعية بأزيلال، جامعة السلطان مولاي سليمان- بني ملال.

# مقدمة:

أنتجت الرقمنة، خلال العقود الأخيرة، كميات غير مسبوقة من 'البيانات الضخمة'، غالبا ما تشتمل على إشارات دالة على السلوك الإنساني والتفاعلات الاجتماعية التي يرجح أن تفيد العلوم الاجتماعية. وبما أن هذه البيانات تُجمع بوصفها نتاجات ثانوية لعمليات إدارية أو تجارية أخرى، فإنها لا تلتقط سوى عدد محدود من تفاصيل وضع ما. في المقابل، يطرح مجال البيانات المجمعة بفضل العمل الإثنوغرافي الميداني بعض القيود، في جوانب سلوكية واجتماعية، على وجه الخصوص. ورغم أن مصادر البيانات الرقمية أثمرت رؤى داخل العلوم الاجتماعية الكمية والكيفية ('بويالسكايا' (Buyalskaya) وآخرون، 2021)، إلا أنها قلّما تدمج إدماجا مباشرا. يتباين هذا الأمر مع التطور التجاري السريع للنماذج المبتكرة حول البيانات التي تجمع بين البيانات الضخمة والبيانات التي يشرحها الإنسان. ويمكن لهذه النماذج في كثير من الأحيان أن تستشف بدقة تعليق الإنسان أو حكمه انطلاقا من البيانات الضخمة، ومن ذلك، على سبيل المثال، ما إذا كان منشور ما على مواقع التواصل الاجتماعي ينطوي على العنف، كما صارت مؤخرا تلتقط الأوصاف الإثنوغرافية للوضعيات ('كوري' (Cury) وآخرون، 2019). غير أنه، وبعيدا عن مجرد تنشيط شرح الإنسان هذا، كُرِّست أبحاث قليلة لدراسة كيفية الجمع الفعال بين العمل الإثنوغرافي الميداني و'البيانات الضخمة' باستخدام المناهج الكمية.

توجز هذه الورقة التوليفات المحتملة بين مصادر البيانات (الضخمة) والبيانات الإثنوغرافية الميدانية، وتوضح كيف يمكن أن تقود هذه التوليفات إلى تحقيق التقدم داخل العلوم الاجتماعية عبر أشكال تعاضد قوية. لنفهم دور البيانات الإثنوغرافية الميدانية، نقارنها في البداية مع أنماط بيانات شائعة أخرى في العلوم الاجتماعية، بما في ذلك 'البيانات الضخمة'. تظهر مقارنتنا أن 'البيانات الضخمة' والبيانات الإثنوغرافية تتشابه (في الغالب) بطرق جوهرية من حيث وضوحها الزمني العالي ومستوى تفاصيلها. غير أن 'البيانات الضخمة' تكون في العادة نتاجا ثانويا لخدمات أخرى، تعرف باسم 'بيانات العادم' (exhaust data). ومن ثمة، غالبا ما توثَّق بشكل سيء، ولا تلتقط بالضرورة ما يسعى الباحثون بشكل مثالي إلى قياسه. تدل طبيعة 'البيانات الضخمة' هذه على دور جديد للبيانات الإثنوغرافية قصد التحقق من صحتها واستكمالها.

نجادل أن العمل الإثنوغرافي الميداني قد يعزز البيانات الكمية الضخمة بطرق أساسية ثلاث: (1) بترسيخ 'حقيقة ميدانية' (أي احتمال قياس الفرضية بالبيانات)، (2) بتقديم 'وصف مكثف' للوضعيات التي تمثلها البيانات (عن طريق فهم السياق الاجتماعي والآليات موضوع الدراسة)، و(3) بقياس الأبعاد المستترة بطريقة أخرى. من هنا، يمكِّن الجمع بين البيانات الإثنوغرافية والبيانات الأخرى الباحثين من معالجة أسئلة البحث الجديدة بمناهج كمية.

تساهم هذه الورقة في الأدبيات الناشئة التي تحدد كيفية وسبب استكمال أدوات علم البيانات الجديدة بالبيانات التي تُجمَع وتُحلَّل بمناهج إثنوغرافية (على سبيل المثال، 'وانغ' (Wang، 2013)؛ 'بلوك' (Blok) و'بيدرسن' (Pederson، 2014)؛ 'بلوك' وآخرون، 2017؛ 'بورناك' (Bornakke) و'ديو' (Due، 2018). توسع هذه الأدبيات حقل المناهج المختلطة (التي تستخدم أساسا بيانات المسح عند الجمع بين البيانات الكيفية والكمية) بتحديد كيفية وسبب استكمال أدوات علم البيانات الجديدة والبيانات الضخمة بالبيانات التي تُجمع وتُحلَّل بالمناهج الإثنوغرافية ('بلوك' و'بيدرسن'، 2014). ففي الوقت الذي تنطلق فيه هذه الأدبيات من وجهة نظر الباحثين ذوي الخلفيات الكيفية (جزئيا على الأقل)، نصدر نحن عن العمل القائم باستهداف باحثي العلوم الاجتماعية ذوي الخلفيات الكمية أساسا. ثمة في حقول دراسات العلوم والتكنولوجيا والمناهج الرقمية قدر كبير من الأدبيات التي تتعلق بكيفية استخدام البيانات المستمدة من السلوك الافتراضي ودراستها في العلوم الاجتماعية (انظر على سبيل المثال 'روجرز' (Rogers، 2009)؛ 'مارس' (Marres، 2012)؛ 'مارس' و'غيرليتز' (Gerlitz، 2016)؛ 'فينتوريني' (Venturini) وآخرين، 2019). هكذا، نقصر تركيزنا على كيفية دمج المناهج وتحليل الظواهر في العالم المادي (أي العالم غير الافتراضي) حيث تكمن إمكانات كبيرة غير مكتشفة.

كما تساهم هذه الورقة في الأدبيات الناشئة التي تستخدم البيانات الإثنوغرافية التاريخية، مثل الأطلس الإثنوغرافي ('موردوك' (Murdock، 1967)، بوصفها مدخلات للنماذج الكمية في حقل التاريخ الاقتصادي، مثلا (انظر 'لاوس' (Lowes، 2020، من أجل المراجعة). نكمّل هذه الأدبيات باقتراح دور أكثر فعالية بالنسبة للعمل الإثنوغرافي الميداني المدمج مع المناهج الكمية، وبتوسيع مجال استخدام البيانات الإثنوغرافية المستمدة من مصادر البيانات الحديثة.

# مقارنة أصناف البيانات المختلفة

قد تصنف المناهج المعتمدة لجمع البيانات بعدة طرق مفيدة. يقسم 'كيتشن' (Kitchin- 2014) البيانات الأولية إلى بيانات ملتقطة وبيانات عادم، بناء على كيفية توليدها. تُجمَع البيانات الملتقطة قصد إنتاج بيانات خاصة. بدل ذلك، تُنتج بيانات العادم بواسطة آليات أو نُظم إلكترونية ، بوصفها نتاجا ثانويا لأنشطة أخرى. وخلال العقد الأخير، انتهى الصناعيون والباحثون على حد سواء إلى اعتبار بيانات العادم، لا بما هي نتاج ثانوي، وإنما بوصفها مدخلا قيّما لعمليات التجارة والبحث.

لاستخدام بيانات العادم مزايا وسلبيات في الآن ذاته. تكمن الميزة الرئيسة في انخفاض تكلفة إمكانية جمع البيانات الواسعة النطاق ذات الدقة العالية من الأفراد. أما العيب الأساس، فيتجلى في أن البيانات كانت مخصصة لأغراض غير بحثية؛ من ثمة، ربما تصير مقياسا ضعيفا للموضوع (المواضيع) قيد الدرس، مما قد يقود إلى استنتاجات خاطئة. ثمة قلق إضافي يتمثل في أن بيانات العادم قد تشتمل على معلومات خاصة وحساسة، حيث تثير معالجتها انشغالات أخلاقية.

ليس الاعتماد على البيانات التي لم تُولَّد قصد الإجابة عن سؤال البحث الخاص بالأمر الجديد في العلوم الاجتماعية. يتم هذا الأمر دوما باستخدام سجل بيانات الإيداعات الضريبية وسجل الناخبين وغير ذلك، مثلا. يطلق 'كيتشن' (2014) على أنماط هذه البيانات اسم البيانات الثانوية. وهو يميز بين البيانات الأولية المجمعة للإجابة عن سؤال البحث الخاص والبيانات الثانوية التي تجمع لغرض آخر وتتاح لآخرين لاحقا بغية إعادة استخدامها وتحليلها. وهذا البعد مستقل عن بعد بيانات العادم/ البيانات الملتقطة. غير أن هذا الصنف من بيانات السجل ليست بيانات عادم لأنها جُمعت وخُزنت عن قصد، مما يعني أن قرارا ما اتخذ حول كيفية جمع البيانات بأكثر الطرق إفادة. في المقابل، عاد ما تولد بيانات العادم بوصفها نتاجا ثانويا لأنشطة تجري على الإنترنت أو الأجهزة الذكية، منها ملفات التسجيل أو سجل الموقع، مثلا. هكذا، ثمة حاجة أكبر إلى فهم ما تقيسه بيانات العادم فعلا- إنها الحاجة إلى 'حقيقة ميدانية.'

أما البيانات الإثنوغرافية، فتجمع في السياقات اليومية، وغالبا ما يشار إليها باسم 'البيانات الميدانية'، باستخدام ملاحظة المشاركين و/أو المقابلات المتعمقة أساسا ('هامرسلي' (Hammersley)، 2007). وبخلاف البيانات الكمية، غالبا ما يكون جمع البيانات الإثنوغرافية "غير مُبَنْيَنٍ بالمعنى الذي لا يتضمن المتابعة من خلال تصميم بحث ثابت ومكثف يُعَدّ في البداية،" مما يسمح بتسجيل ملاحظات وأفكار تلقائية وسديدة ('هامرسلي'، 2007). ورغم أننا واعون أن الإثنوغرافيا يمكن أن تشتمل على جماع من المناهج، فإننا نحيل، في هذه الورقة، على ملاحظات التقطت من الميدان عندما نشير إلى البيانات الإثنوغرافية.

## الحقيقة الميدانية والوصف المكثف

تشير 'الحقيقة الميدانية' إلى الملاحظة المباشرة التي قد تؤكد (أو ترفض) صحة البيانات الحسية (وتأويل الباحثين). نشأ هذا المصطلح من الاستشعار عن بعد باستعمال الأقمار الصناعية ('هوفر' (Hoffer)، 1972) عند علماء الأرصاد الجوية، مثلا، للتحقق من أن إعصارا مسجلا في بيانات الأقمار الصناعية كان يظهر أيضا في الموقع. نجد أن الحقيقة الميدانية تمثل استعارة جيدة تعبر عن كيفية مساعدة الملاحظات البشرية في التحقق من بيانات العادم، لأن جودة البيانات قد لا تكون مثالية أو قد تقيس شيئا غير مقصود.

يرتبط المفهوم الإثنوغرافي 'للوصف المكثف' بالحقيقة الميدانية، لكنه متميز عنها، كما يتضح من المثال الكلاسيكي الذي يقدمه 'غيرتز' (1973) حول الاختلاف بين رفّة العين وغمزتها. قد نلاحظ صبيين يقومان برفة الجفن الأيمن السريعة نفسها. تكون الرفة عند أحدهما نتاجا لإراديا لاعتلال جسدي، بينما تنم عن تواطؤ مع صديق عند الثاني. ورغم أن هاتين الوضعيتين تبدوان متشابهتين تماما، إلا أنهما مختلفتان اختلافا تاما، إذ يمكن للوصف المكثف أن يميز بينهما[1]

[1] انظر 'بروكر' (Brooker) (2021) في هذا العدد الخاص لمزيد من المناقشة حول دور الوصف المكثف وكيفية الحصول عليه.

يكون الحصول على حقيقة ميدانية ووصف مكثف من الوضعيات حيث تولَّد البيانات مفيدا بصورة خاصة عند استخدام بيانات العادم. يمكن أن تساعد الحقيقة الميدانية على معالجة ما تسجله بيانات العادم بالفعل، وهو أمر مهم لأن البيانات قد لا تستقر عبر الزمن أو قد لا تتناسب مع إدراك المحلل لها. ويمكن أن يساعد الوصفُ المكثف الباحثَ على فهم الظواهر الاجتماعية المؤثرة.

رغم أن البيانات الإثنوغرافية التي يجمعها البشر توفر إمكانية الحصول على تفاصيل متزايدة بدرجة كبيرة، إلا أن هذه البيانات قد تواجه مشكلات مماثلة لطبيعة البيانات الضخمة العادمة. قد يرتكب علماء الإثنوغرافيا، والملاحظون البشر بصورة عامة، أخطاء عند تسجيل البيانات، مثلا، عن طريق إساءة فهم ما يلاحظونه أو العجز عن تذكر حدث أو جزء من المعلومات بصورة صحيحة عند تحرير ملاحظاتهم. فضلا عن ذلك، قد لا يتناسب تأويلهم اللاحق للملاحظات الميدانية مع الواقع أو مع طريقة نظر أولئك الأشخاص موضوع الملاحظات إليها. وأخيرا، قد يكون حضور الآخرين وحده عاملا مزعجا، وقد يغير الأشخاص موضوع الملاحظات سلوكهم. غير أن جميع هذه الاعتبارات، إلى جانب المقاربات الرامية إلى تخفيف هذه الأخطاء والفهوم الخاطئة والتفكير فيها، متضمنة بالفعل في المناهج والتحليلات الإثنوغرافية ('برنارد'، 2017؛ 'أوريلي' (O'Reilly)، 2012). في الواقع، يرجح أن يستفيد الباحثون الكميون الذين يعملون ببيانات العادم من رؤى التقليد الإثنوغرافي في العمل بالبيانات غير المهيكلة وولوج ما تمثله البيانات من الناحية النقدية 'تشارلز' (Charles) و'غيرمان' (Gherman)، 2019).

يعتمد احتمال جمع البيانات حول الحقيقة الميدانية والأوصاف المكثفة من عدمه على القيمة التي تضيفها إلى التحليل. نلاحظ أن الحصول على الحقيقة الميدانية لا يحتاج إلى إثنوغرافيين من ذوي الدربة، بل يمكن أن يحصل عليها ملاحظون يتمتعون بدربة أقلّ أو لا يتمتعون بها نهائيا.[1] غير أننا نحتاج، لفهم السياق الاجتماعي فهما أعمق، إلى الإثنوغرافيين لتقديم وصف مكثف. يقتضي المنطق، في الوضعيات التي تتطلب الحقيقة الميدانية والوصف المكثف، أن يحصل الإثنوغرافي على الاثنين. يمكن أن يكون جمع

[1] طور الباحثون، في المجال الرقمي، خوارزميات رؤية بالحاسوب أو معالجة اللغة الطبيعية، ولم تصبح هذه النماذج ممكنة إلا من خلال الاستفادة من التعليقات التوضيحية الواسعة النطاق التي يسميها/ يصنفها الإنسان بأنها نصوص أو صور ('كريزيفسكي' (Krizhevsky) وآخرون، 2009؛ 'راجبوركار' (Rajpurkar) وآخرون، 2016).

الحقيقة الميدانية جزءا طبيعيا في عمل الباحث الإثنوغرافي الميداني، مثلا، في مراحل العمل الميداني الأولى، التي غالبا ما تكون موجهة نحو المهام ('برنارد' (Brenard)، 2017). وليس من الممكن، ولا من الضروري، في معظم الدراسات التي توظف البيانات الضخمة، أن يحصل الباحث على الحقائق الميدانية و/ أو الأوصاف المكثفة لكل حالة يتم فيها توليد البيانات. ما إن يتم جمع عدد معين من الملاحظات والحالات عبر العمل العمل الإثنوغرافي الميداني، اعتمادا على تطبيق الدراسة وسياقها، حتى يصبح المرء قادرا إما على تقدير المؤشرات المرغوبة بدقة كافية، وإما على استقراء الحقائق الميدانية و/ أو الأوصاف المكثفة للحالات حيث تكون هذه الدقة مفقودة.

## الأبعاد الأساسية للبيانات

تشتمل معظم البيانات الكمية الموجهة للتحليل في العلوم الاجتماعية على ثلاثة أبعاد أساسية تسمى 'ن' و'م' و'ت'.[1] هنا، يمثل 'نون' عدد الأفراد (أو وحدة الملاحظة الخاصة الأخرى) الذين تمت ملاحظتهم، ويمثل 'م' عدد المتغيرات (مثل الصحة أو الدخل) الملحوظة عند هؤلاء الأفراد. أما 'ت'، فهو بعد زمني يخبرنا بعدد النقاط الزمنية التي لوحظ فيها عدد المتغيرات بالنسبة لعدد الأفراد. كما سنناقش فيما يلي، عادة ما تكون البيانات الكمية الكلاسيكية في العلوم الاجتماعية عالية من حيث 'ن' ومتوسطة من حيث 'م' ومنخفضة نسبيا من حيث 'ت'.

يشير 'بورناك' (Bornakke) و'ديو' (Due) (2018)، فيما يتعلق بهذا الأمر، إلى البيانات الضخمة بوصفها ضخمة وضعيفة، وإلى بيانات الملاحظة الإثنوغرافية بوصفها قليلة ومكثفة، مع تركيزهما أساسا على مستوى التعقيد السياقي (مكثف/ ضعيف) وعدد الملاحظات (قليلة/ ضخمة). رغم أن هذا الأمر يمكن أن يشكل تمايزا ملائما، يتأتى من خلفية كمية، إلا أننا نجد أبعاد 'ن' و'م' و'ت' ملائمة بما أنها تلتقط بوضوح الجوانب الزمنية المشتركة في البيانات الضخمة والبيانات الإثنوغرافية معا.

[1] الأبعاد المشتركة الأخرى للبيانات المستعملة في العلوم الاجتماعية هي المعلومات الجغرافية والبيانات العلائقية، الخ.

غالبا ما يشار إلى البيانات الرقمية الجديدة بوصفها 'بيانات ضخمة'. وغالبا ما يوصف ما يجعل 'البياناتِ الضخمةِ' ضخمةً من منظور علوم الحاسوب الذي يركز على الحجم والتنوع والسرعة ('غاندومي' (Gandomi) وحيدر (Haider)، 2015). غير أن نعت 'ضخم' قد يشير إلى العديد من الأفراد أو العديد من المتغيرات الملحوظة ('ن' أو 'م' عالي). ففي نظرنا، تتمثل الجدة من حيث 'البيانات الضخمة' في بحوث العلوم الاجتماعية في السرعة؛ أي في البعد الزمني ('تـ' عالٍ).[1] ورغم أن 'البيانات الضخمة' لا ترصد جميعها الأفراد عبر الزمن، فإن هذا هو الحال. من هنا، نعتبر 'البيانات الضخمة' بأنها 'تـ' عالٍ (أي تجمع باستمرار وبدقة عالية)، لكن بحجمي 'م' و'ن' ملتبسين.

## البيانات الإثنوغرافية في مقابل 'البيانات الضخمة'

عادة ما تكون بيانات العمل الإثنوغرافي الميداني بمستوى 'ن' منخفض، بالمعنى الذي يفيد أنه يُخضِع قلة من الأفراد فقط للملاحظة ('هامرسلي'، 2007).[2] غير أن البيانات الإثنوغرافية تتميز بعمق كبير ('م' عالٍ) بما أن الإثنوغرافي قد يسرد، بالنسبة للفرد أو السياق، مئات، وربما آلاف، التفاصيل خلال وصف وضعية، منها المظهر والنبرة والسلوك والتصرف، مثلا. على النحو ذاته، عادة ما تكتسي البيانات الإثنوغرافية بعدا زمنيا يجعلها بمستوى 'ت' عالٍ. تتراوح مقاييس الزمن هذه من الأفراد موضوع الملاحظة في سياق محدد على امتداد بضع ساعات إلى متابعة الأفراد ذاتهم عبر السياقات طيلة أشهر أو سنوات.[3] يتعلق الأمر بتوليف المعلومات العميقة ('م' عال) والملاحظات المستمرة ('ت' عال)، توليف يسمح للإثنوغرافيين بإنجاز أوصاف مكثفة.[4]

## مثال: الهواتف الذكية المعدية

لنقدم مثالا ملموسا حول كيفية وسبب أهمية التوليف بين البيانات الضخمة والعمل

[1] من النادر، في البحث الكمي في العلوم الاجتماعية التي تستعمل البيانات الكلاسيكية (الملتقطة)، أن يكون 'ن' عالٍ و'م' عالٍ و'تـ' عالٍ. غالبا ما تكون البيانات عبارة إما عن 'ن' عال، أو 'م' متوسط/ عال و'تـ' منخفض، أو 'ن' عال و'تـ' متوسط/ عال و'م' منخفض، رغم أن هذا أخذ يتغير مع بروز 'البيانات الضخمة'.

[2] نحيل إلى وحدة الملاحظة بوصفها أفرادا، إلا أنها قد تكون شركات أو جماعات أو وحدات ملاحظة أخرى.

[3] ثمة طريقة أخرى للتفكير في 'م' و'ت'، تتمثل تموضع الإثنوغرافيين في المكان والزمن مثلا، انظر 'أبوت' (Abbot) (1997).

[4] للاطلاع على اعتبارات ذات صلة، انظر 'بيدرسن' (Pedersen) (2019).

الإثنوغرافي الميداني، سنراجع ورقتنا وسنعطي أمثلة نستدل بها عن كيفية استفادة التحليل من الجمع بين البيانات الضخمة والبيانات الإثنوغرافية. يعالج 'غلافيند' وآخرون (2021) ما إذا كان استعمال الهواتف الذكية يصير معديا عندما يقترب الأشخاص بعضهم من بعض (جسديا) في سياقات اجتماعية. تعدّ معرفة مثل هذه التداعيات السلوكية مهمة لأنها تضخّم العواقب العكسية على الصحة والتعلم ('فرغيسون' (Ferguson, 2017؛ 'بيير-نيلسون' وآخرون، 2020)، تلك الناتجة عن استعمال الهواتف الذكية المتزايد. فضلا عن ذلك، تغير هذه التداعيات استعمال الهواتف الذكية في السياقات الاجتماعية من اختيار فردي إلى اختيار جماعي (جزئيا)، مع اقتراح أن القيود يمكن أن تكون مفيدة.

يعتمد تحليلنا على بيانات ضخمة مأخوذة من دراسة شبكات 'كوبنهاغن' ('ستوبزينسكي (Stopczynski) وآخرون، 2014) التي تتيح لنا تتبع الوجود المادي المشترك (عبر مستشعرات البلوتوث)، وكذا زمن المشاركة على الشاشة والرسائل النصية الواردة/ المرسلة. نحن نركز عما إذا كان استخدام الأفراد الحاضرون هواتفهم الذكية يتزايد عند وصول الرسائل النصية، مما يتيح تقديرا سببيا للعدوى، بالنظر إلى زمنها شبه العشوائي. نجد أنه عندما يتلقى فرد ما رسالة نصية، تزداد مدة استعمال الأفراد القريبين منه لشاشاتهم خلال الدقيقتين المواليتين. لا يظهر هذا الأثر إلا في حالة وجود علاقة اجتماعية بين شخصين موجودين (يتواصلان عبر الهواتف أو صديقين على الفايسبوك).

صار التحليل ممكنا باستعمال الدقة العالية للبيانات الضخمة التي أتاحت لنا تتبع الديناميات الاجتماعية للوضعية في المستوى الثاني. غير أننا واجهنا، في مرات عديد خلال تحليلنا، مانعا ناتجا عن 'ضعف' البيانات الضخمة، حيث ستساهم الإضافتان التاليتان في تحليلنا مساهمة جوهرية. أولا، افتقرنا في تحليلنا إلى الحقيقة الميدانية حول قيام تفاعل اجتماعي بين الشخصين الحاضرين من عدمه. مثل هذه البيانات يمكن أن تتيح لنا أن نأخذ في الحسبان سبب الضآلة النسبية الحاصلة في حجم الأثر الذي قِسْناه، بما أن المرء يتوقع أن تحدث التداعيات أساسا عندما يكون هناك تفاعل اجتماعي. ثانيا، سيتيح لنا الوصف المكثف للوضعيات الاجتماعية أن نفهم كيفية وسبب ظهور تداعٍ ما أثناء الاستخدام. ويمكن للإثنوغرافيين الذي ينجزون ملاحظات بالمشاركة أن يساعدوا على فهم ما إذا كانت عدوى الهواتف الذكية تعتمد، مثلا، على الأعراف الاجتماعية والوضع الاجتماعي النسبي وتشكل المجموعات والنشاط وغير ذلك. ومن المحتمل أن يتيح لنا هذا الأمر فهما أعمق لكيفية حصول العدوى، وعما إذا كانت تطرح مشكلة ما، ومتى تحصل هذه المشكلة، وكذا الطريقة المثلى لمعالجة الآثار العكسية المحتملة، وذلك بمنع الهواتف الذكية في سياقات معينة أو تشجيع بعض المعايير حول استخدام الهاتف الذكي.

# الجمع بين البيانات الإثنوغرافية والأدوات الكمية

عند تطبيق الأدوات الكمية على بيانات العمل الإثنوغرافي الميداني، تصير البيانات الإثنوغرافية في حاجة إلى أن تُبَنْيَن إلى حد ما، ويفضَّل أن تُرقمَن. تقوم أفضل طريقة لإنجاز هذا الأمر على حجم البيانات والتطبيق العملي لذلك. في بعض الأحيان، قد تكون السجلات الأولية ملائمة؛ وفي أحيان أخرى، تكون المدخلات من الملاحظات الميدانية 'المشفرة' هي الملائمة ('دوهان' (Dohan) و'سانشيز جانكوفسكي' (Sanchez-Jankowski، 1998؛ 'أبرامسون' (Abramson) و'دوهان'، 2015). من ثمة، يمكن أن تدمج البيانات الإثنوغرافية المبنينة مع مصادر البيانات الأخرى أو تستعمل كما هي.

يمثل استعمال أدوات معالجة البيانات مسبقا أو وصفها إحدى الطرق الأساسية لدمج البيانات الإثنوغرافية مع المناهج الكمية وتقنيات علم البيانات. ويمكن أن تتراوح مثل هذه الأدوات من تخطيط تطوير عدد المرات التي تذكر فيها الكلمات المنتقاة عبر الزمن في مقابلة مكتوبة أو استعمال التعلم الآلي غير الخاضع للإشراف الذي يتيح مقاربة مرنة، إلى إيجاد أنماط جديدة في البيانات، كمجموعات الملاحظات المتماثلة مثلا. في الواقع، كان علماء الأنثروبولوجيا الكمية روادا في تحليل المجموعات هذه ('درايفر' (Driver) و'كروبر' (Kroeber)، 1932). نادرا ما تثمر هذه الأدوات نتائج نهائية، إلا أنها قد تتيح مدخلات مفيدة تكمل أجزاء أخرى من التحليل الكمي. ويمكن أن تكون مفيدة، على وجه الخصوص، خلال المراحل الأولى من العمل الميداني، بوصفها دليلا لإيجاد الأنماط التي تستحق المزيد من الاستكشاف في الميدان،[1] والتحليل اللاحق للعمل الميداني، كمكمِّل للتحليل الإثنوغرافي الكلاسيكي.

ثمة مقاربة مختلفة جذريا تكمن في دمج البيانات الإثنوغرافية بنماذج من نتائج محددة. وتتجلى مقاربة مثل هذه في النهج السريع التطور للتعلم الآلي الخاضع للإشراف، الذي يستعمل لانتقاء أفضل النماذج قصد التنبؤ بالنتائج أو استنتاجها. تفيد هذه الأدوات في العديد من سياقات العلوم الاجتماعية، في التنبؤ بأداء الطلبة الأكاديمي مثلا ('بيير-نيلسن' وآخرون، 2021)، وقد تدمج في الخوارزميات التي تحدد القرارات بناء على التوقعات.

[1] وفق مراحل الملاحظة بالمشارِكة السبع عند 'برنارد'، يمكن أن يُجرى التحليل الكمي خلال مرحلة 'الاستراحة'، قصد إرشاد مرحلة 'التركيز' اللاحقة.

وكما جاء في المقدمة، يمكن أن تستعمل الآلاتُ البياناتِ التي يشرحها البشر بغية تعلم التعرف على نماذج الوضعية وخصائصها بشكل صحيح ('كوري' وآخرون، 2019). وثمة استعمال آخر يكمن في استخدام البيانات الإثنوغرافية بوصفها مدخلا للتنبؤ، الذي يؤدي إلى توقعات أفضل (بالنظر إلى أن البيانات الإثنوغرافية المضمنة هي بيانات ملائمة). ويكشف 'بيير-نيلسن' وآخرون (2021) أن توقعات استنتجت من المعلومات الخاصة بالمهام (كاستعمال درجة المدرسة الثانوية عند توقع درجات الكليةمثلا) قد تتفوق على التوقعات المستنتجة باستعمال البيانات الفردية الضخمة الدقيقة. ويمكن أن نتصور أن الإثنوغرافيين قد يتيحون مثل هذه البيانات الخاصة بالمهام؛ ومن ثمة، يعززون التوقع.

ويمثل الاستدلال السببي مقاربة مختلفة جذريا عن التوقع، حيث يعتمد على المعرفة النظرية بالبنية السببية و/ أو التدخلات في شكل تجارب أو شبه تجارب. إننا نرى أساسا طريقتين يمكن من خلالهما للعمل الميداني الإثنوغرافي أن يحسّن الاستدلال السببي. أولا، قد يتدخل الإثنوغرافي بوصفه عامل تغيير في الوضعية الملحوظة. بعبارة إحصائية، سيتصرف الإثنوغرافي بوصفه مصدر تباين (خارجي) يمكن أن يحول الدراسة الميدانية إلى سياق شبه تجريبي. على سبيل المثال، إذا أردنا أن نعالج ما إذا كان استخدام الهاتف المحمول يؤثر في الأقران القريبين، يمكن للإثنوغرافي أن يتلاعب بالوضعية عن طريق استعمال الهاتف عشوائيا لمعالجة مدى تأثير ذلك في الوضعية الاجتماعية. ثانيا، يتيح العمل الإثنوغرافي الميداني إمكانية تقديم فهم أفضل للآليات التي تجعل المتغيرات المشتركة تؤثر في النتيجة، وتعزز من ثمة الادعاء السببي. وتتمثل الطريقة الوحيدة لإنجاز هذا في جمع البيانات عبر العمل الميداني الذي يتضمن الأوصاف المكثفة لوضعيات مماثلة لمجموعة البيانات (الضخمة) المستخدمة، أو كامنة بداخلها.

نؤكد أن دمج العمل الإثنوغرافي الميداني بمصادر البيانات لأغراض سببية أو كمدخل للنماذج التوقعية سيقتضي تخطيطا دقيقا، مما قد يتطلب جمع البيانات الإثنواغرافية والضخمة بشكل متزامن. وغالبا ما يكون هذا التعاون مطلبا مكلّفا للباحثين، إذ ينظر 'مواتس' (Moats- 2021) في التحديات والطرق المثلى لتنفيذ ذلك.

# كلمة شكر

نشكر المراجعيْن المجهولي الهوية على تعليقاتهما المفيدة والبناءة. لقد تمت البحوث البحثية التي نناقش في هذا التعليق بمركز 'كوبنهاغن' لعلم البيانات الاجتماعية (SODAS)، وهو مركز ذو تخصصات متعددة للغاية، يعنى بالبحث والتعليم في جامعة 'كوبنهاغن'. إننا ممتنون لمناقشتنا مع الزملاء داخل هذه البيئة المتعددة التخصصات، لا سيما 'مورتن أكسيل بيدرسن' (Morten Axel Pedersen) و'يالمار بانغ كارلسن' (Hjalmar Bang Carlsen). كانت مساهمة 'كارلسن' مهمة في عملية كتابة هذه المقالة، خاصة عند توضيح الاختلاف بين الحقيقة الميدانية والوصف المكثف. كما نشعر بالامتنان لـ'أن صوفي بيك نادسن' (Anne Sofie Beck Knudsen)، لنقاشاتها موضوع استخدام البيانات الإثنوغرافية في الاقتصاد، و'إسبن بروغر ليمينغر' (Esben Brøgger Lemminger) لمساعدته على إيجاد مقالات بحثية ذات صلة بموضوعنا.

# المراجع:


- Abbott A (1997) Of time and space: The contemporary relevance of the chicago school. Social Forces 75(4): 1149–1182.
- Abramson CM and Dohan D (2015) Beyond text: Using arrays to represent and analyze ethnographic data. Sociological methodology 45(1): 272–319.
- Bernard HR (2017) Research Methods in Anthropology: Qualitative and Quantitative Approaches, chapter 13, pages 354–356, 379–381. Rowman & Littlefield.
- Bjerre-Nielsen A, Andersen A, Minor K, et al. (2020) The negative effect of smartphone use on academic performance may be overestimated: Evidence from a 2-year panel study. Psychological Science 31(11): 1351–1362.
- Bjerre-Nielsen A, Kassarnig V, Lassen DD, et al. (2021) Task-specific information outperforms surveillance-style big data in predictive analytics. Proceedings of the National Academy of Sciences 118(14): 1–3.
- Blok A, Carlsen HB, Jørgensen TB, et al. (2017) Stitching together the heterogeneous party: A complementary social data science experiment. Big Data & Society 4(2): 1–15. DOI: 2053951717736337.
- Blok A and Pedersen MA (2014) Complementary social science? quali-quantitative experiments in a big data world. Big Data & Society 1(2): 1–6. DOI: 2053951714543908.
- Bornakke T and Due BL (2018) Big–thick blending: A method for mixing analytical insights from big and thick data sources. Big Data & Society 5(1). DOI: 2053951718765026.
- Brooker P (2021) Machine anthropology: A view from sociology. Submitted to special issue on Machine Anthropology in Big Data & Society.

- Buyalskaya A, Gallo M and Camerer CF (2021) The golden age of social science. Proceedings of the National Academy of Sciences 118(5).
- Charles V and Gherman T (2019) Big Data Analytics and Ethnography: Together for the Greater Good. 19–33, Cham: Springer International Publishing.
- Cury M, Whitworth E, Barfort S et al. (2019) Hybrid methodology: Combining ethnography, cognitive science, and machine learning to inform the development of context-aware personal computing and assistive technology. Ethnographic Praxis in Industry Conference Proceedings.
- Dohan D and Sanchez-Jankowski M (1998) Using computers to analyze ethnographic field data: Theoretical and practical considerations. Annual review of Sociology 24(1): 477–498.
- Driver HE and Kroeber AL (1932) Quantitative expression of cultural relationships. Publications in American Archaeology and Ethnology 31(4): 211–256.
- Ferguson CJ (2017) Everything in moderation: Moderate use of screens unassociated with child behavior problems. Psychiatric Quarterly 88(4): 797–805.
- Gandomi A and Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35(2): 137–144.
- Geertz C (1973) Thick description: Towards an interpretive theory of culture. In Geertz, C., editor, The Interpretation of Cultures, volume 3, pages 143–168. Basic Books.
- Glavind KL, Andersen A and Bjerre-Nielsen A (2021) Smartphone use is socially contagious. Unpublished working paper. Available as Chapter 3 at: https://www.econ.ku.dk/ forskning-og-publikationer/Publikationer/ph.d˙se.
- Hammersley M (2007) Ethnography. In The Blackwell Encyclopedia of Sociology. John Wiley & Sons, Ltd.

- Hoffer RM (1972) The importance of ground truth data in remote sensing. Technical Report NASA-CR-130020, National Aeronautics and Space Administration (NASA).
- Kitchin R (2014) The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. London: Sage.
- Krizhevsky A, Hinton G, et al (2009) Learning multiple layers of features from tiny images.
- Lowes S (2020) Ethnographic and field data in historical economics. Working Paper 27918, National Bureau of Economic Research.
- Marres N (2012) The redistribution of methods: On intervention in digital social research, broadly conceived. The Sociological Review 60: 139–165.
- Marres N and Gerlitz C (2016) Interface methods: Renegotiating relations between digital social research, sts and sociology. The Sociological Review 64(1): 21–46.
- Moats D (2021) Rethinking the 'great divide'. Science & Technology Studies 34(1): 19–42.
- Murdock GP (1967) Ethnographic atlas: A summary. Ethnology 6(2): 109–236.
- O'Reilly K (2012) Ethnographic methods, second edition, chapter 8. Routledge.
- Pedersen MA (2019) Anthropology of/as repetition. History and Anthropology 30(2): 226–232.
- Rajpurkar P, Zhang J, Lopyrev K, et al. (2016) Squad: 100, 000+ questions for machine comprehension of text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 2383–2392.
- Rogers R (2009) The End of the Virtual: Digital Methods, volume Inaugural Lecture 339. Amsterdam University Press.
- Stopczynski A, Sekara V, Sapiezynski P, et al. (2014) Measuring large-scale social networks with high resolution. PloS one 9(4): e95978.

- Venturini T, Munk AK and Jacomy M (2019) Actor-network versus network analysis versus digital networks: Are we talking about the same networks? In Digital STS : A Field Guide for Science & Technology Studies, pages 510–524. Princeton University Press.
- Wang T (2013) Big data needs thick data. In: Ethnography Matters. Available at: http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data.

## نبذة عن المؤسسة

مؤسسة بحثية استشارية عابرة للتخصصات. نكرس جهودنا البحثية وخبراتنا المنهجية لتقديم توصيات وتصورات استراتيجية، مستمدة من أساليب العلوم الاجتماعية. كما يساعد عمل المؤسسة على سد الفجوة المتنامية بين احتياجات الناس، والباحثين والخبراء، وصانعي السياسات والأسواق الاستهلاكية. ولتحقيق ذلك، نستعمل في ذلك أفضل الطرق، والممارسات والتقاليد، العلمية والمنهجية، المستخدمة في الاستطلاعات والعمل الميداني، من أجل دراسة أنماط السلوك البشري، وتحقيق فهم أعمق للمعاني الاجتماعية والممارسات والاحتياجات الحقيقية للمجموعات المدروسة.

تنبع رؤية المؤسسة من الفهم العميق، للسلوكات والتفاعلات الإنسانية المركبة في الحياة اليومية، وفي العالم الاجتماعي المتشابك. تعد هذه الرؤية أداة حاسمة في تطوير منظورات متجددة ومبتكرة للديناميات المجتمعية المتسارعة في عالمنا اليوم.


www.mt-index.com

contact@mt.index.com

+212.661.45.56.79


Meta-Index | INSTITUTE FOR SURVEY RESEARCH