لماذا يحتاج العلم إلى بيانات مفتوحة؟

جيفري بولتون

الملخص:

يؤكد المقال أن إتاحة بيانات البحث شرط أساسي لفحص الأدلة وإعادة إنتاج النتائج والتصحيح الذاتي للعلم، ويربط هذه المبادئ بتاريخ النشر العلمي وثورة تقنيات المعلومات. ويناقش كيف أوجد تضخم البيانات فرصًا لاكتشاف العلاقات والأنماط، لكنه أضعف اقتران الأفكار المنشورة بأدلتها القابلة للتحقق. ويدعو إلى «انفتاح ذكي» يجعل البيانات وبياناتها الوصفية والشيفرة قابلة للاكتشاف والوصول والفهم والتقييم وإعادة الاستخدام. كما يتناول تمويل الإتاحة ضمن البحث، والاستشهاد بالبيانات وحوافز المشاركة، والقيود المبررة لحماية الخصوصية والسلامة والمصلحة العامة. ويعرض أدوار الممولين والجامعات والناشرين والشبكات الوطنية والدولية في تطبيق هذه المبادئ، ويخلص إلى أن البيانات المفتوحة أساس علم أكثر نزاهة وكفاءة واتصالًا بالمجتمع واحتياجاته المعرفية والديمقراطية.

الكلمات المفتاحية: بيانات البحث المفتوحة، العلم المفتوح، إعادة إنتاج النتائج، مشاركة البيانات، الانفتاح الذكي

ميلاد العلم الحديث:

العلم المفتوح ليس جديدًا؛ فقد استقرت مبادئه في بدايات عصر التنوير الأوروبي، وثبت أنها أساسية للعلم الحديث بوصفه أكثر وسائل اكتساب المعرفة موثوقية. وتجسدت في الدوريات العلمية الأولى المنشورة علنًا، التي يراها مؤرخو العلم بذورًا للثورات العلمية في القرنين الثامن عشر والتاسع عشر [1].

كانت (Philosophical Transactions of the Royal Society)، التي بدأ صدورها عام 1665 وما تزال تصدر، أول دورية علمية وفق عرض المؤلف. وقد ابتكر فكرتها هنري أولدنبورغ، أول أمين للجمعية الملكية الحديثة التأسيس، وكان كثير المراسلات في الشؤون العلمية. ورأى نشر مراسلاته بدل إبقائها خاصة، وأقنع الجمعية بإنشاء الدورية لهذا الغرض. واشترط على مراسليه أن تصحب الفكرة الأدلةُ، أي البيانات التي تستند إليها، لكي تُنشر.

أتاح ذلك للآخرين فحص العلاقة المنطقية بين الفكرة والدليل، وتكرار التجارب والملاحظات، وإعادة استخدام البيانات. وثبت أن الانفتاح على الفحص أقوى أشكال مراجعة الأقران، وأهم في النهاية بكثير من التحكيم السابق للنشر. وأصبح أساسًا للتصحيح الذاتي العلمي، فيما غدا الاستعداد للدحض لبنة أساسية في تقدم العلم وبناء المعرفة.

يضم الشكل الأول في الأصل صورة أولدنبورغ وصفحة عنوان المجلد الأول. ويوضح تعليقه أنه ابتكر أيضًا مراجعة الأقران بطلب رأي زميلين من الجمعية في الأعمال المقدمة ومدى صلاحيتها للنشر. أُثبتت هذه المعلومة نصيًا، ولم تُنقل الصورة.

ثورة تقنية ونتائجها:

شهدت العقود الأخيرة ثورة تقنية بأبعاد تاريخية أطلقت نموًا غير مسبوق في قدرة الإنسان على اقتناء كميات هائلة من البيانات والمعلومات وتخزينها ومعالجتها وتوصيلها فورًا إلى أنحاء العالم بصرف النظر عن الموقع. وأحدثت تغيرات جوهرية في السلوك البشري والاجتماعي والاقتصادي، وآثارًا في البحث والتعلم أعمق وأوسع انتشارًا، مثلًا، من ثورة الاتصال السابقة المشابهة التي أحدثتها مطبعة غوتنبرغ.

نفذت الثورة إلى العلم والعمل الأكاديمي وطرائق ممارستهما، لكن سلوك أفراد ومؤسسات كثيرين ما يزال، كما في الثورات المشابهة، ملائمًا لعصر تقني سابق. فالكتاب المجلد والدورية ما يزالان أداتي العمل الأساسيتين في كثير من مكتباتنا ولدى العاملين فيها، ويتركزان غالبًا في مبانٍ كبيرة. ومع ذلك أصبح الوصول الإلكتروني في العلم على الأقل متاحًا في كل مكان ووقت، وصارت الحاجة إلى دعم موزع يقدمه مكتبيون إلكترونيون مدربون، وقل العلماء الذين يرتادون المكتبة التقليدية.

يتيح انفجار البيانات وقدرتنا على جمعها وتكاملها وتحليلها وسائل قوية جديدة لفك التعقيد، وتحسين توقع سلوك النظم، وكشف أنماط ظواهر عجزنا عن تفسيرها سابقًا. كما يتيح إعادة استخدام البيانات وجمعها وإعادة تنظيمها وربطها لتعزيز تلك القدرات. ويتوقف استثمار ذلك على الوصول إلى مجموعات بيانات كثيرة والربط بينها، مما يقتضي إتاحة بيانات البحث وانفتاحها بصورة اعتيادية سهلة.

ويتطلب أيضًا أخلاقًا لمشاركة البيانات وأنماط تعاون جديدة تزيد إبداع العمل العلمي بتفاعل عقول ومجتمعات كثيرة لا تحدها أسوار المؤسسات. وقد تعزز هذه التغييرات نشر الأفكار العلمية وأدلتها في المجتمع والتعليم، وتغير الديناميات الاجتماعية للعلم ليصبح عملًا عامًا بدل إجرائه خلف أبواب المختبرات المغلقة.

لكن لانفجار البيانات جانب سلبي لم ننتبه إليه إلا حديثًا. فقد تضخمت أحجام البيانات التي تمثل أدلة الأفكار العلمية إلى حد تراجعت معه في العقود الأخيرة العادة الصارمة لإدراجها، وإدراج بياناتها الوصفية التي تبين نشأتها، في العمل المنشور تقليديًا. وربما كان العلم يسير بذلك دون وعي نحو أزمة في المصداقية.

ومثال ذلك ورقة نُشرت قبل عامين من كتابة الأصل، أبلغ مؤلفوها عن محاولات تكرار نتائج 50 ورقة مرجعية في أبحاث الأورام قبل السريرية [2]، فلم ينجحوا إلا في 11% من الحالات. وعكس الإخفاق في 89% خللًا في المنطق العلمي جزئيًا، لكنه عكس في حالات كثيرة قصور البيانات أو بياناتها الوصفية؛ فحتى لو كانت النتائج سليمة منطقيًا، لم تكن هناك وسيلة للتحقق منها.

ثمة الآن انتشار لتعذر إعادة إنتاج النتائج [3] لا ينشأ من الأسباب التقليدية كتجربة أو ملاحظة معيبة أو خلل في الاستدلال الإحصائي، بل من نقص البيانات أو بياناتها الوصفية، الذي يحجب الأدلة فلا يمكن إثبات الحجة أو نفيها. وهذا يقوض مبدأ التصحيح الذاتي العلمي الحيوي.

إعادة ابتكار العلم المفتوح للقرن الحادي والعشرين:

نشر البيانات وإطلاقها ومشاركتها علنًا عادات حيوية ينبغي لمجتمع البحث إعادة تعريفها وتطويرها للعصر الحديث، حتى يستثمر التقنية ويحافظ على التصحيح الذاتي ويعظم إسهام البحث في الفهم والرفاه. ولتكييف قواعد البحث مع عصر غني بالبيانات، واستثمار فرصه وتصحيح الانفصال بين البيانات والفكرة، ينبغي أن يفهم المجتمع ضرورة البيانات المفتوحة ويتبناها مبدأً معياريًا.

لكن الانفتاح وحده لا قيمة له ما لم يكن، بتعبير تقرير الجمعية الملكية عن العلم المفتوح [4]، «انفتاحًا ذكيًا». أي أن تتاح البيانات والبيانات الوصفية ذات الصلة والشيفرة الحاسوبية التي تسند الورقة المنشورة للفحص بالتزامن مع نشرها، وأن تستوفي الشروط الآتية:

- قابلة للاكتشاف: يسهل العثور على وجودها بالبحث الإلكتروني.
- قابلة للوصول: يمكن فحصها والاستعلام عنها عند اكتشافها.
- مفهومة: يمكن استيعاب معناها.
- قابلة للتقييم: يمكن تقدير منشئها وموثوقيتها.
- قابلة لإعادة الاستخدام: يمكن استخدامها من جديد ودمجها مع بيانات أخرى.

وينبغي أيضًا إتاحة بيانات البحث الممول عموميًا أو خيريًا، التي لا تُستخدم أدلة لفكرة علمية منشورة، إتاحة ذكية بعد فترة محددة مسبقًا ينفرد فيها منشئوها بالوصول إليها. وعلى من يعيد استخدامها من غير منشئيها أن ينسب إليهم فضلها رسميًا.

ومن المهم تجنب الفصل الزائف بين ممارسة البحث العلمي ونشر ورقة من جهة، وجعل البيانات مفتوحة بذكاء من جهة أخرى. فتكلفة إنتاج بيانات مفتوحة بهذه الصورة جزء أصيل من العملية العلمية، وليست إضافة اختيارية، وينبغي تمويلها ضمن البحث نفسه.

ومع أن الأصل في بيانات البحث العام أو الخيري هو الانفتاح الذكي، توجد حدود مبررة له: حين يخدم الاستغلال التجاري المصلحة العامة ويتطلب نموذج أعمال القطاع تقييد الانفتاح؛ ولحماية خصوصية الأفراد الذين تتضمن القواعد معلوماتهم الشخصية؛ وحين تهدد الإتاحة السلامة بحوادث غير مقصودة أو الأمن بهجوم متعمد. لكنها لا تبرر استثناءات شاملة من الأصل، بل ينبغي تسويغ كل حالة على حدة.

تنفيذ مبادئ بيانات البحث المفتوحة:

ينبغي أن يتبنى الممولون هذه المبادئ شروطًا لاستمرار الدعم، وأن تتبناها الجامعات والمعاهد مبادئ للبحث الحديث، والناشرون شروطًا لقبول الأعمال. وعلى الجمعيات العلمية التي تعبر لمجتمعات تخصصاتها عن مبادئها وأولوياتها أن تدافع عنها أيضًا.

ومع وجوب اعتبارها قواعد معيارية للعلماء، يتوقف التغيير الكبير على إدراك الباحثين قيمة الالتزام بها. ويرجح اتباعها مع وجود حوافز للباحث ومؤسسته وللمستخدمين. وقد يكون الاستشهاد بالبيانات المودعة حافزًا بسيطًا قويًا، إذ يمنح الباحث منتجًا قابلًا للاستشهاد إلى جانب الورقة التقليدية.

وتوجد أدلة [5] على أن الاستشهاد بمجموعة بيانات قيّمة قد يفوق كثيرًا الاستشهاد بأول ورقة تقليدية تفسرها في المجالات التي أصبح فيها الاستشهاد بالبيانات أمرًا معتادًا. وهذا مفهوم، لأن الإبداع في تصميم الملاحظات والتجارب التي تكشف علاقات أو ظواهر مهمة أو تؤكدها قد يساوي على الأقل الإبداع في تفسير البيانات وكتابة الورقة، ويستحق تقديرًا مكافئًا. ويمكن النظر إلى الورقة المنشورة بوصفها إعلانًا عن العلم الكامن في البيانات.

لا يقتصر أثر التقنيات الرقمية على العلوم، بل يتيح فرصًا لجميع مجالات البحث والعمل الأكاديمي. ففي الإنسانيات الرقمية مثلًا، تتضمن البحوث غالبًا مناهج واستراتيجيات فكرية جديدة تظل راسخة في الاهتمامات الإنسانية التقليدية. ولا يقتصر التحدي على البيانات المولودة رقمية، بل يشمل مجموعات نصية كبيرة وأشكال المعلومات البصرية والسمعية والسمعية البصرية والحسية والعصبية، بل والحركية.

أولويات واتجاهات وطنية ودولية:

ثمة مبررات قوية لتبني إجراء (DataCite) الراسخ [6] بصورة حاسمة في نظام البحث البريطاني؛ لأهميته الذاتية ولقدرته على استثارة حماسة المجتمع. وهو شبكة عالمية تمثل المكتبة البريطانية عقدتها الوطنية، وتعمل على الاعتراف بالبيانات إسهامًا مشروعًا قابلًا للاستشهاد في السجل العلمي. وتوفر معرّفات (DOI) لمجموعات البيانات والمخرجات غير التقليدية، فتيسر تعريفها المستديم والاستشهاد بها.

وينبغي، حيث أمكن، إصلاح الإجراءات وهياكل المكافأة والأعراف السلوكية التي تعوق مشاركة البيانات أو التعاون المفتوح الجديد أو تمنعهما، لتشجيعهما وتيسيرهما ومكافأتهما.

دعت الجمعية الملكية حديثًا إلى منتدى بريطاني لبيانات البحث المفتوحة، متابعةً لتقريرها الصادر عام 2012 «العلم بوصفه عملًا عامًا». وقرر المنتدى أن يواصل دوره جماعة ضغط تحفّز العمل تضم جميع العناصر الأساسية في المجتمع البحثي البريطاني [7]، لترويج مفاهيم مثل تلك الواردة هنا وتنفيذها.

واتخذت الحكومة موقفًا قويًا بإنشاء مجلس شفافية قطاع البحث [8] لتعزيز بيانات البحث المفتوحة ومراقبتها ضمن أجندتها للبيانات المفتوحة. لكن البحث دولي بطبيعته، ولذلك تحتاج قواعده إلى الطابع الدولي والتبني الدولي. وكانت خطوة أولى بيان وزراء علوم مجموعة الثماني عام 2013 عن أهمية بيانات البحث المفتوحة واتفاقهم على تعزيزها.

وينبغي أن تتضمن الوسائل الدولية داخل المجتمع العلمي المجلس العالمي للبحوث [9]، بوصفه جهة مناسبة للاتفاق على المبادئ وترويجها، و(CODATA) [10]، وهي لجنة تابعة للمجلس الدولي للعلوم [11] ذات توجه تقني يمكّنها من تحديد المعايير المفتوحة اللازمة لنظام دولي فعال وترويجها.

ولتعزيز الاتساق الدولي ثنائيًا، التقى وفد المنتدى البريطاني نظيره الأمريكي، لجنة الاتساق على نطاق واسع للتعليم العالي [12]، في واشنطن في أبريل من سنة كتابة المقال، لبحث التعاون الفعال بين المجتمعين في قضية البيانات المفتوحة.

لكن أكبر تأثير في إنشاء بيئة دولية مفتوحة سيأتي من قدرة العلماء في مجالات مثل المعلوماتية الحيوية وبلورة المواد الكيميائية، وفي موضوعات واسعة مثل الصحة العامة [13]، على إقناع زملائهم؛ فقد أدرك هؤلاء بالفعل قيمة النظام الدولي المفتوح للعالم الفرد وللجهد العلمي المشترك.

وماذا عن العلم المفتوح:

اهتم المقال بالبيانات المفتوحة، لكن العلم المفتوح يتكون من ثلاثة عناصر: إجراء العلم بانفتاح في تحديد أولويات البحث وجمع البيانات؛ وإطلاق بيانات مفتوحة بذكاء، سواء احتفظت بها السلطات العامة أم أنشأتها معاهد القطاع العام أو الجامعات الممولة بحثيًا من المال العام؛ والنشر بالوصول المفتوح.

وتنبع أهمية العلم المفتوح بهذا المعنى من سببين. أولهما أن أثر العلم في الحياة الحديثة عميق إلى حد يجعل الفحص المفتوح لأدلة استنتاجاته شرطًا لديمقراطيات فعالة تعكس اختيارات مواطنيها وأعرافهم. وثانيهما أن نزاهة العملية العلمية وكفاءتها تعتمدان على الانفتاح الذكي وسرعة فحص الاستدلال العلمي وصرامته، وكشف الأخطاء واختبار النظريات الجديدة.

ينبغي أن نطمح إلى علم يمارس بوصفه عملًا عامًا لا خلف أبواب المختبرات المغلقة. ومبدأ البيانات المفتوحة هو الأساس الذي يقوم عليه هذا الطموح.

المراجع:

1. Shapin S , A social history of truth: civility and science in seventeenth-century England , 1994 , Chicago , University of Chicago Press .
2. Begley C E Ellis L M , Raise standards for preclinical cancer research , Nature , 2012 , 483 , 533 .
3. PubMed Retraction Notices – By Year (2012): http://pmretract.heroku.com/byyear (accessed 15 May 2014).
4. The Royal Society , Science as an Open Enterprise . Science Policy Centre publication 2 December 2012 , London , Royal Society .
5. Piwowar H A Day R S Fridsma D B , Sharing detailed data is associated with increased citation rate , 2007 , PLOS ONE , 2 , 3 , e308 .
6. DataCite : http://www.datacite.org (accessed 15 May 2014).
7. Royal Society UK Open Data Research Forum : https://royalsociety.org/events/2014/01/open-data-forum/ (accessed 15 May 2014).
8. UK Government Research Sector Transparency Board : https://www.gov.uk/government/groups/research-sector-transparency-board (accessed 15 May 2014).
9. Global Research Council: http://www.globalresearchcouncil.org (accessed 15 May 2014).

10. CODATA : http://www.codata.org (accessed 15 May 2014).
11. International Council for Science : http://www.icsu.org (accessed 15 May 2014).
12. US Committee on Coherence at Scale for Higher Education : http://coherence.clir.org (accessed 15 May 2014).
13. Wellcome Trust 'Spotlight issue' : http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues (accessed 15 May 2014).