كيف تحفظ قيمة بياناتك العلمية؟ عشر قواعد للإدارة والمشاركة

Ten Simple Rules for the Care and Feeding of Scientific Data

أليسا غودمان، ألبرتو بيبي، ألكسندر و. بلوكر، كريستين ل. بورغمان، كايل كرانمر، ميرسيه كروساس، روزان دي ستيفانو، يولاندا غيل، بول غروث، مارغريت هيدستروم، ديفيد و. هوغ، فيناي كاشياب، آشيش ماهابال، أنيتا شيميغينوفسكا، ألكسندرا سلافكوفيتش

الملخص:

يقدم خمسة عشر باحثًا عشر قواعد للحفاظ على قيمة البيانات العلمية وإتاحة استخدامها والاعتراف بجهود أصحابها. وتشمل القواعد توثيق البيانات ونشرها بمعرّفات دائمة، والتخطيط لإعادة استخدامها، وإتاحة خطوات التحليل والشيفرات وربطها بالمنشورات. كما تناقش اختيار المستودعات وطرق الاستشهاد بالبيانات ودور الباحثين والمؤسسات في تشجيع المشاركة وتطوير مهارات إدارة البيانات. ويعرض المقال موارد عملية للمستودعات وأنظمة سير العمل واستضافة الشيفرات والترخيص وحماية الخصوصية.

الكلمات المفتاحية: إدارة بيانات البحث، مشاركة البيانات، المعرّفات الدائمة، مستودعات البيانات، الاستشهاد بالبيانات

مقدمة:

في مطلع القرن السابع عشر، وجّه غاليليو غاليلي (Galileo Galilei) مقرابه نحو كوكب المشتري، وراح يرسم كل ليلة مواضع الكوكب والنقاط الضوئية القريبة منه بمقياس دقيق، مع تأريخ كل رسم. وقادته أرصاده إلى التعرف إلى أربعة أقمار تدور حول المشتري، بما دعم تصوره عن النظام الشمسي. ولم تقتصر قيمة عمله على النتائج الفلكية؛ فقد أتاحت دقة سجلاته ومنشوراته للآخرين فهم الطريق الذي قاده إليها. جمعت ملاحظاته بين البيانات، أي رسوم المشتري وأقماره، والبيانات الوصفية الأساسية، مثل توقيت الرصد والطقس وخصائص المقراب، والنص

الذي يشرح المناهج والتحليل والاستنتاجات. وحافظ كتابه «الرسول النجمي» (Sidereus Nuncius) [1] على هذا الترابط بين النص والبيانات ووصفها. وأسهمت أعمال غاليليو ونهجه في الرصد والتحليل في دفع الثورة العلمية وتشكيل المنهج العلمي الحديث [2]، [3].

تُعدّ معظم المشروعات البحثية اليوم مكتملة حين يُكتب مقال في دورية مبني على التحليل ويُنشر. والمشكلة أن كمية البيانات الفعلية ووصف البيانات في المنشورات الحديثة، على خلاف تقرير غاليليو في «الرسول النجمي»، لا تكاد تكفي أبدًا لإعادة الدراسة المعروضة أو حتى للتحقق منها إحصائيًا. والأسوأ من ذلك أن الباحثين الراغبين في البناء على الأعمال المعروضة في الأدبيات وتوسيعها كثيرًا ما يجدون صعوبة في استعادة البيانات المرتبطة بمقال ما بعد نشره. بل إن العلماء، في أحيان أكثر مما يودّون الاعتراف به، لا يستطيعون حتى استعادة البيانات المرتبطة بأعمالهم المنشورة هم أنفسهم.

ومما يزيد الوضع الحديث تعقيدًا أن لكلمتي «البيانات» و«التحليل» اليوم تعريفات أكثر تنوعًا مما كانت عليه في زمن غاليليو. فبإمكان البحوث النظرية أن تُنشئ مجموعات «بيانات» كبيرة عن طريق المحاكاة (مثل مشروع محاكاة الألفية [(The Millennium Simulation Project)]: رابط المصدر). وكثيرًا ما يجري جمع البيانات على نطاق واسع بوصفه جهدًا يشارك فيه مجتمع علمي بأكمله (مثل مشروع الجينوم البشري [(The Human Genome project)]: رابط المصدر)، وهو ما يؤدي إلى «قواعد بيانات» هائلة على الإنترنت (أي مجموعات منظمة من البيانات). كما أن الحواسيب أساسية في المحاكاة وفي معالجة البيانات التجريبية والرصدية إلى حد أنه كثيرًا ما يصعب أيضًا رسم خط فاصل بين «البيانات» و«التحليل» (أو «الشيفرة البرمجية») عند الحديث عن إدارة «البيانات» والعناية بها. ففي بعض الأحيان تكون نسخة من الشيفرة المستخدمة في إنشاء البيانات أو معالجتها أساسية لاستخدام تلك البيانات إلى حد أنه ينبغي النظر إلى الشيفرة تقريبًا بوصفها جزءًا من الوصف الذي تقدّمه «البيانات الوصفية» للبيانات. وفي أحيان أخرى تكون الشيفرة المستخدمة في دراسة علمية أكثر قابلية للفصل عن البيانات، ولكن حتى في هذه الحالة تنطبق كثير من مبادئ الحفظ والمشاركة على الشيفرة كما تنطبق على البيانات.

فكيف نتولى إدارة البيانات والعناية بها إذن؟ لا شك أن العناية ببياناتك تقترن بعمل إضافي، لكن العناية المبكرة توفّر الوقت وتزيد من عمق الفهم لاحقًا. ومع أن عددًا متزايدًا من الباحثين، ولا سيما في المشروعات التعاونية الكبيرة، يدركون أن إجراء البحث مع وضع المشاركة وإعادة الاستخدام في الحسبان أمر أساسي، فإن ذلك لا يزال يتطلب تحوّلًا في النموذج الفكري السائد. فما زال معظم الناس مدفوعين بمراكمة المنشورات وبالانتقال إلى المنشور التالي في أسرع وقت ممكن. ولكن كلما وجدنا نحن العلماء أنفسنا نتمنى لو أتيح لنا الوصول إلى بيانات موجودة لكن يتعذر العثور عليها الآن [4]، أدركنا أكثر لماذا تضر الإدارة السيئة للبيانات بالعلم. فكيف يمكننا التحسّن؟

يقدّم هذا المقال دليلًا موجزًا للخطوات التي يستطيع العلماء اتخاذها لضمان أن تظل بياناتهم وما يرتبط بها من تحليلات ذات قيمة وأن تحظى بالاعتراف. ففي السنوات القليلة الماضية وحدها كُتبت مئات الأوراق العلمية والتقارير عن مسائل مشاركة البيانات، ومنشأ البيانات ومسارها (data provenance) ، وقابلية البحوث لإعادة الإنتاج (reproducibility)، والترخيص، والنسب، والخصوصية، وغير ذلك، غير أن هدفنا هنا ليس مراجعة تلك الأدبيات. بل نقدّم بدلًا من ذلك دليلًا موجزًا موجّهًا إلى الباحثين الذين يريدون أن يعرفوا لماذا من المهم «إدارة البيانات والعناية بها»، مع بعض النصائح العملية حول كيفية القيام بذلك. ويقدّم القسم الأخير في ختام هذا العمل (روابط إلى مصادر مفيدة) روابط إلى أنواع الخدمات المشار إليها في ثنايا النص.

القاعدة 1: اعتنِ ببياناتك وشجّع الآخرين على ذلك:

إدارة البيانات عملية تفاعلية تتكرر مع الوقت. فإذا حرصت على إتاحة بياناتك للآخرين بسهولة، فمن المرجح أن يفعل الآخرون الشيء نفسه، ولو بعد حين. وريثما يتحقق هذا التوازن الجديد في المشاركة، يمكنك اتخاذ إجراءين مهمين. أولًا، اعتنِ ببياناتك ووثّقها وانشر بياناتك، ويُفضَّل أن يكون ذلك بالطرائق المتينة الموصوفة في القاعدة 2. وابدأ الآن، فالأدوات والموارد الأفضل لإدارة البيانات تتكاثر، والجامعات والمجتمعات البحثية تتجه نحو استثمارات أكبر في مستودعات البيانات (القاعدة 8)، وعدد متزايد من أمناء المكتبات والعلماء يتعلمون مهارات إدارة البيانات (القاعدة 10). وفي الحد الأدنى، فإن حبك لبياناتك المتاحة سيفيدك أنت: إذ ستتمكن من العثور على بياناتك وإعادة استخدامها إن أحسنت معاملتها. ثانيًا، مكّن الآخرين وشجّعهم على العناية ببياناتهم وتوثيقها ونشرها. فإذا كنت عالمًا باحثًا، فالأرجح أنك لست مؤلفًا فحسب، بل محكّم أيضًا لدورية متخصصة أو لمؤتمر. وبوصفك محكّمًا، اطلب من مؤلفي الأوراق التي تحكّمها تقديم توثيق لبياناتهم وإتاحة الوصول إليها وفقًا للقواعد المبيّنة في بقية هذا المقال. ومع أن المقاربات المؤسسية أساسية بلا ريب (القاعدتان 8 و10)، فإن تغيير القناعات عالمًا بعد عالم فعّال أيضًا.

القاعدة 2: شارك بياناتك على الإنترنت بمعرّف دائم:

لا شيء يدوم إلى الأبد حقًا، ولذا فإن «دائم» تعني في الواقع «طويل الأمد» فحسب. فعلى سبيل المثال، من غير المرجح أن يكون موقعك الشخصي على الويب خيارًا جيدًا لتخزين البيانات على المدى الطويل (ومع ذلك، فإن وضع بياناتك على موقعك على المدى القصير جدًا أفضل من عدم فعل أي شيء على الإطلاق!). وبوجه عام، ومع أن كثيرًا من الأوراق تتضمن محددات موارد موحدة (URLs) لإتاحة الوصول إلى مجموعات البيانات، فإن معظمها يصبح متعذر الوصول خلال بضع سنوات [5]. وأفضل خيار لإصدار بياناتك مع ضمان طويل الأمد هو إيداعها في أي أرشيف بيانات يكون الوجهة المعتمدة في مجالك. وسيقوم الأرشيف الملائم الجدير بالثقة بما يلي: (1) إسناد معرّف مثل «هاندل» ((Handle)، (hdl)) أو «معرّف الكائن الرقمي» ((digital object identifier)، (doi))؛ (2) واشتراط أن تقدّم توثيقًا وبيانات وصفية كافية؛ (3) وتولّي «حفظ بياناتك والعناية بها» باتباع ممارسات جيدة في الإشراف على البيانات (curation). وإذا لم يوجد مثل هذا الأرشيف في مجالك، فهناك أيضًا خدمات عامة على الإنترنت (غير مخصصة لمجال بعينه) يمكنها استضافة بياناتك وإصدار معرّفات دائمة (persistent identifiers) لها (انظر القاعدة 8). وترد إشارات إلى بضعة مستودعات عامة في «روابط إلى مصادر مفيدة» (القسم أ)، وترد قوائم أطول لمثل هذه الخدمات في «روابط إلى مصادر مفيدة» (ب).

القاعدة 3: خطط مسبقًا لإعادة استخدام البيانات:

يصعب استخدام بيانات الآخرين من دون معرفة طبيعتها وكيفية الحصول عليها. ويعرّف فريق (W3C) المعني بمنشأ البيانات ومسارها (W3C Provenance Group) هذا المفهوم بأنه مجموع العمليات والأشخاص والمؤسسات والوثائق التي أسهمت في توليد معلومة أو أثّرت فيها أو نقلتها (رابط المصدر). وكلما كان توثيق هذا المسار أدق، زادت فرص إعادة استخدام البيانات. ويتطلب ذلك إتاحة ثلاثة عناصر: البيانات، والبيانات الوصفية التي تشرحها، والمعلومات المتعلقة بإنتاجها، ومنها الشيفرة البرمجية. ومن الأفضل التخطيط لهذه العناصر منذ البداية وفق المستوى المنشود من إعادة الاستخدام. فإذا أردت تمكين الآخرين من إعادة إنتاج العمل بالكامل، فستحتاج إلى شيفرة تحليل قابلة للتشغيل، وبيئة مناسبة لها، ونسخ من البيانات يمكن التحقق منها. وإذا كان الهدف فحص العمل

فحسب، فقد تكفي المخرجات الوسيطة ووصف شبه برمجي للخطوات. أما إذا أردت استخدام البيانات في تطبيقات متنوعة، فاعتمد مبكرًا صيغًا ومعايير وصفية قياسية. والحد الأدنى في جميع الحالات هو حفظ سجل مؤرخ لإصدارات البيانات والشيفرة. وستفيدك هذه الخطوات أنت والباحثين الآخرين حين تعودون إلى العمل لاحقًا. ويضم القسم (هـ) من «روابط إلى مصادر مفيدة» أدوات لجمع المواد البحثية وتنظيمها لإعادة الاستخدام.

القاعدة 4: انشر خطوات العمل لتوضيح سياق البيانات:

يوفر نشر خطوات المعالجة سياقًا ضروريًا لفهم البيانات وإعادة استخدامها. ولهذا يشرح الباحثون عادة جمع البيانات ومعالجتها وتحليلها في أقسام المنهج والتحليل. ويُطلق على تسلسل هذه الخطوات اسم «سير العمل» (workflow)، ويُعدّ توثيقه جزءًا من سجل منشأ البيانات ومسارها. وفي بعض المجالات، ولا سيما علم الجينوم، تساعد برمجيات مثل (GenePattern) (رابط المصدر) على تنفيذ التجارب وتسجيلها، ويمكن إتاحة البرنامج وإصداره وإعداداته مع البيانات. لكن إدارة العملية كاملة بحزمة واحدة لا تزال نادرة في مجالات أخرى. ومن المأمول أن يربط التوثيق الآلي مستقبلًا بين البيانات والشيفرات وخطوات العمل والمنشور العلمي. أما الآن، فالحد الأدنى هو إرفاق مخطط بسيط يوضح انتقال البيانات عبر البرامج وكيفية إنتاج المخرجات الوسيطة والنهائية. ويمكن أيضًا استخدام خدمات لتنظيم سير العمل وإتاحته، كما في القسم (ج) من «روابط إلى مصادر مفيدة». وحتى عند استخدام بيانات مأخوذة من أرشيف موثق، ينبغي تسجيل الاستعلام الذي استُخدم لاسترجاعها، وجميع المعالجات اللاحقة. ويعود هذا التوثيق بفوائد مباشرة عليك وعلى زملائك، فضلًا عن فائدته للمجتمع العلمي.

القاعدة 5: اربط بياناتك بمنشوراتك كلما أمكن:

سواء أكانت «بياناتك» تتضمن جداول، أم جداول بيانات إلكترونية، أم صورًا، أم رسومًا بيانية، أم قواعد بيانات، و/أو شيفرة برمجية، فعليك أن تتيح أكبر قدر ممكن منها مع أي ورقة تعرضها. وإذا كان ذلك عمليًا ومفيدًا، فشارك بياناتك في أبكر مرحلة ممكنة من سير عملك البحثي: بمجرد انتهائك من التحليل، حتى قبل أن تكتب أي مقالات عنه. بل يمكن الاستشهاد ببياناتك قبل إدراجها في ورقة (أو من دون إدراجها) (انظر القاعدة 7). وتقدّم كثير من الدوريات الآن طرائق قياسية للإسهام بالبيانات في أرشيفاتها وربطها بورقتك، وغالبًا ما يكون ذلك بمعرّف دائم. وكلما أمكن، ضمّن في متن ورقتك مباشرة استشهادات (روابط) ببياناتك وشيفرتك، لكل منها معرّفها الدائم الخاص، تمامًا كما تحيل إلى الأدبيات الأخرى. وإذا لم توفّر الدورية التي تستضيف ورقتك مكانًا لبياناتك و/أو معرّفًا لها، فاستخدم مستودعًا (القاعدة 8) واحصل على معرّفك الخاص (القاعدة 2). وفي الحد الأدنى، يمكنك أن تنشر حزمة من الملفات (البيانات، والشيفرات، وتوثيق المعاملات، والبيانات الوصفية، ومعلومات الترخيص، و/أو قوائم روابط إلى مثل هذه المواد) بمعرّف دائم على الإنترنت (القاعدة 2)، وأن تحيل إليها. وإذا كانت سياسات الدوريات في مجالك لا تسمح بربط جيد بين البيانات والأدبيات، فحاول إحداث التغيير (انظر القاعدتين 1 و10).

القاعدة 6: انشر شيفرتك البرمجية (حتى الأجزاء الصغيرة منها):

هل كتبت أي شيفرة برمجية لإجراء تحليلك؟ مهما وجدتها مليئة بالأخطاء وتافهة، انشرها. فهناك كثير من مستودعات الشيفرة المصدرية سهلة الاستخدام، التي لا تتيح استضافة البرمجيات فحسب، بل تيسّر أيضًا التعاون وتتبع النسخ (انظر «روابط إلى مصادر مفيدة» [د]). فقد تكون شيفرتك، حتى أقصر نص برمجي (سواء أكنت فخورًا بجودته أم لا)، مكوّنًا مهمًا لفهم بياناتك وكيفية توصلك إلى نتائجك [6]. وتؤدي البرمجيات أدوارًا عدة فيما يتصل بالبيانات والبحث العلمي، ولا تزال الأعراف المتعلقة بنشرها في طور التشكل وتختلف من تخصص إلى آخر [7].

ففي بعض الحالات تكون البرمجية هي منتج البيانات الأساسي (مثل الخوارزميات الجديدة). وفي حالات أخرى تكون البيانات هي المنتجات البحثية الأساسية، ومع ذلك فإن أفضل طريقة لتوثيق مصدريتها هي نشر البرمجية التي استُخدمت في توليدها بوصفها «بيانات وصفية». وفي كلتا الحالتين، يُعد نشر الشيفرة المصدرية وتاريخ نسخها أمرًا حاسمًا لتعزيز الشفافية وقابلية إعادة الإنتاج. ويقلل استخدام البرمجيات مفتوحة المصدر، حين يكون ذلك ممكنًا، من العوائق أمام المستخدمين اللاحقين لمنتجات بياناتك المرتبطة بالبرمجيات [8]. وتنطبق أفضل الممارسات نفسها التي نوقشت أعلاه فيما يتعلق بالبيانات وسير العمل على المواد البرمجية أيضًا: استشهد بالبرمجيات التي تستخدمها، وقدّم معرّفات فريدة ودائمة (القاعدة 2) للشيفرة التي تشاركها.

القاعدة 7: حدّد الطريقة التي تريد أن يُنسب إليك الفضل بها:

الأرجح أنك تريد أن يُنسب إليك الفضل فيما تشاركه. غير أن نظام النسب المستخدم للمقالات العلمية، الذي يتحقق عبر الاستشهادات، كثيرًا ما يتعطل في حالة البيانات والبرمجيات. فحين يعيد مؤلفون آخرون استخدام بياناتك أو شيفرتك أو يستشهدون بها، قد تحصل على شكر أو على رابط يشير إلى عملك. وإذا تجشمت أنت وزملاؤك عناء كتابة «ورقة بيانات» (data paper)، غرضها الرئيس وصف بياناتك و/أو شيفرتك، فقد تحصل أيضًا على استشهاد [9]. لكن كتابة «ورقة بيانات» ليست دائمًا مرغوبة أو ذات صلة. فكيف تتصرف إذن للحصول على الفضل الكامل الذي تستحقه عن بياناتك وشيفرتك؟ إن أفضل طريقة هي ببساطة أن تصف توقعاتك بشأن الطريقة التي تود أن يُعترف بك بها. وإذا أردت، يمكنك أيضًا إصدار بياناتك بموجب ترخيص، والإشارة صراحة في الورقة أو في البيانات الوصفية إلى الطريقة التي تريد أن ينسب بها الآخرون الفضل إليك. ولكن، مع أن للآليات القانونية مزاياها، فإنها قد تؤدي أيضًا دون قصد إلى قيود على إعادة استخدام البيانات التي تشاركها. وعلى أي حال، اجعل المعلومات المتعلقة بك (مثل اسمك ومؤسستك)، وبالبيانات و/أو الشيفرة (مثل الأصل، والنسخة، والملفات المرتبطة، والبيانات الوصفية)، وبالطريقة الدقيقة التي تود أن يُنسب إليك الفضل بها، واضحة قدر الإمكان. وتقدّم منظمة المشاع الإبداعي (Creative Commons) (رابط المصدر) تراخيص سهلة التطبيق، يتميز كثير منها بأنه قابل للقراءة آليًا، كما توجد خيارات أخرى مماثلة، مثل تلك التي تقدّمها (Open Data Commons) (رابط المصدر). ويقدّم القسم (ز) من «روابط إلى مصادر مفيدة» مزيدًا من المعلومات.

القاعدة 8: ادعم مستودعات البيانات واستخدمها:

قد تكون أصعب خطوات مشاركة البيانات والشيفرة وأكثرها استهلاكًا للوقت في بعض الأحيان هي البحث عن مكان لوضعها واتخاذ القرار بشأنه. وتتباين ممارسات مشاركة البيانات تباينًا واسعًا بين التخصصات: ففي بعض المجالات تكون مشاركة البيانات وإعادة استخدامها أمرًا أساسيًا وشائعًا، بينما تكون مشاركة البيانات في مجالات أخرى ثقافة «تبادل هدايا» [10]. فإذا كان لدى مجتمعك العلمي مستودع قياسي بالفعل، فاستخدمه. وإذا كنت لا تعرف من أين تبدأ البحث، أو كنت بحاجة إلى مساعدة في الاختيار بين المستودعات ذات الصلة، فاسأل اختصاصي معلومات، مثل عالم بيانات أو أمين مكتبة يعمل في مجالك (وراجع أدلة مستودعات البيانات المدرجة في «روابط إلى مصادر مفيدة» [ب]). وعند الاختيار بين المستودعات، حاول أن تجد المستودع الذي يقدّم أفضل مزيج من سهولة الإيداع، والإقبال عليه داخل المجتمع العلمي، وإمكانية الوصول، وقابلية الاكتشاف (discoverability)، والإشراف على البيانات بما يضيف إليها قيمة، والبنية التحتية للحفظ، والاستمرارية المؤسسية، ودعم صيغ البيانات ومعاييرها التي تستخدمها. وتذكّر أنه حتى إن لم يكن لمجالك مستودع تخصصي، فقد يكون

لدى مؤسستك مستودع، ويمكن لأمين المكتبة أو الأرشيفي في مؤسستك أن يرشدك إلى كيفية استخدام ذلك المورد المحلي. وإذا لم يكن لدى مجتمعك العلمي ولا لدى مؤسستك مستودع ذو صلة، فجرّب مستودعًا عامًا أو فكّر في إنشاء مستودعك الخاص (انظر القاعدة 2، و«روابط إلى مصادر مفيدة» [و]).

القاعدة 9: كافئ الزملاء الذين يشاركون بياناتهم على النحو الصحيح:

سواء أفعلت ذلك شخصيًا في الاجتماعات والمؤتمرات العلمية أم عبر التواصل المكتوب عند تحكيم الأوراق وطلبات المنح، كافئ زملاءك الذين يشاركون البيانات والشيفرة. واحشد زملاءك وأشرك مجتمعك العلمي بتقديم تغذية راجعة عن جودة أصول البيانات في مجالك. وأثنِ على من يتبعون أفضل الممارسات. فكلما كانت البيانات التي ينشئها زملاؤك متاحة بوصفها مجموعة منظمة من نوع ما، تحسّنت القدرة البحثية لمجتمعك. وكلما زادت مشاركة البيانات واستخدامها والاستشهاد بها، تحسّنت. وإلى جانب المشاركة الشخصية والتشجيع، فإن أفضل طريقة لمكافأة مشاركة البيانات هي النسب: استشهد دائمًا بمصادر البيانات التي تستخدمها. واتبع الممارسة العلمية الجيدة وانسب الفضل إلى أصحاب البيانات التي تستخدمها، متبعًا صيغة الإحالة التي يفضلونها ووفقًا لأفضل الممارسات الحالية. ويجري حاليًا تطوير معايير وممارسات للاستشهاد بمصادر البيانات ونسبتها بنشاط عبر شراكات دولية [11]، [12].

القاعدة 10: كن مناصرًا لعلم البيانات:

كما تقول القاعدة 1، ليس المهم أن تحب بياناتك فحسب، بل أن يحب الآخرون البيانات أيضًا. ولا يزال الموقف الذي يعدّ البيانات والشيفرة «مخرجات ثانوية» تأتي بعد المنشورات العلمية التقليدية سائدًا. لكن قيمة تنظيم شبكة مفتوحة من البيانات والشيفرات القابلة لإعادة الاستخدام تزداد وضوحًا يومًا بعد يوم لعدد متزايد من الناس، في الوقت الذي يحاول فيه العلماء استخدام الخليط المحبط والمغري في آن واحد من بيانات البحث المتاحة عبر الشبكة الحالية المرتجلة من الأنظمة على الإنترنت. وعليك، بوصفك عالمًا، أن تساعد على تنظيم تخصصك ومؤسستك للتحرك بسرعة أكبر نحو عالم من البيانات والبحوث المفتوحة والقابلة للاكتشاف والقابلة لإعادة الإنتاج. ومن الخطوات المهمة في هذا السبيل أن تدعو إلى توظيف اختصاصيي البيانات وإلى الدعم الشامل للبرامج المؤسسية التي تحسّن مشاركة البيانات. واحرص على ألا يقتصر الاستمتاع بإجراء البحوث باستخدام بيانات وأدوات متاحة مجانًا على الباحثين المتقدمين (مثل باحثي ما بعد الدكتوراه): اشرح وبيّن قيمة البيانات التي حظيت بالعناية للباحثين من طلبة الدراسات العليا والمرحلة الجامعية الأولى. ودرّس مقررات كاملة، أو مقررات قصيرة، تتعلق بالعناية بالبيانات والبرمجيات، أو أدمج هذه الأفكار في المقررات القائمة. وشكّل مجموعات خاصة بتخصصك لتعزيز مشاركة البيانات والشيفرة. ونظّم جلسات نقاش غير رسمية لذوي الاهتمامات المشتركة (birds-of-a-feather) أو جلسات خاصة في أثناء الاجتماعات الكبيرة، تعرض أمثلة أدت فيها ممارسات المشاركة الجيدة إلى نتائج وتعاونات أفضل. وكن قدوة بأن تطبّق ما تدعو إليه.

روابط إلى مصادر مفيدة:

أ: مستودعات البيانات العامة:

- (Dataverse) (رابط المصدر): مستودع لبيانات البحث يتولى الحفظ طويل الأمد والممارسات الأرشيفية الجيدة، بينما يستطيع الباحثون مشاركة بياناتهم والاحتفاظ بالتحكم فيها ونيل الاعتراف بها.

- (FigShare) (رابط المصدر): مستودع يستطيع المستخدمون فيه إتاحة جميع مخرجاتهم البحثية على نحو قابل للاستشهاد والمشاركة والاكتشاف.
- (Zenodo) (رابط المصدر): خدمة مستودع تمكّن الباحثين والعلماء والمشروعات والمؤسسات من مشاركة نتائج البحوث متعددة التخصصات (البيانات والمنشورات) التي ليست جزءًا من المستودعات المؤسسية (institutional repositories) أو المستودعات الموضوعية القائمة، وعرضها.
- (Dryad) (رابط المصدر): مستودع يهدف إلى جعل أرشفة البيانات بسيطة ومجزية قدر الإمكان، عبر مجموعة من الخدمات لا يقدّمها بالضرورة الناشرون أو المواقع المؤسسية.

ب: أدلة مستودعات بيانات البحث:

- (DataBib) (رابط المصدر): أداة تساعد الناس على تحديد مستودعات بيانات البحث على الإنترنت والعثور عليها. ويُنشئ المستخدمون والببليوغرافيون سجلات تصف مستودعات البيانات ويشرفون عليها، ويستطيع المستخدمون البحث فيها.
- (Re3data.org) (رابط المصدر): سجل عالمي لمستودعات بيانات البحث من تخصصات أكاديمية مختلفة، موجّه إلى الباحثين وجهات التمويل والناشرين والمؤسسات العلمية.
- دليل الوصول المفتوح (Open Access Directory) (رابط المصدر): قائمة بالمستودعات وقواعد البيانات الخاصة بالبيانات المفتوحة.
- فهرس (Force 11) (Force 11 Catalog) (رابط المصدر): قائمة ديناميكية بالموارد العلمية القائمة على الويب، ومجموعة من أنظمة النشر البديلة، وقواعد البيانات، والمنظمات والمجموعات، والبرمجيات، والخدمات، والمعايير، والصيغ، وأدوات التدريب.

ج: أنظمة إدارة سير العمل:

- (Taverna) (رابط المصدر): نظام مفتوح المصدر ومستقل عن المجالات لإدارة سير العمل، وهو مجموعة من الأدوات المستخدمة لتصميم مسارات سير العمل العلمية وتنفيذها والمساعدة في إجراء التجارب الحاسوبية (in silico).
- (Kepler) (رابط المصدر): برمجية مصممة لمساعدة العلماء والمحللين ومبرمجي الحاسوب على إنشاء النماذج والتحليلات وتنفيذها ومشاركتها عبر طائفة واسعة من التخصصات العلمية والهندسية.
- (Wings) (رابط المصدر): نظام دلالي لسير العمل يساعد العلماء على تصميم التجارب الحاسوبية.
- (VisTrails) (رابط المصدر): نظام مفتوح المصدر لإدارة سير العمل العلمي ومنشأ البيانات ومسارها، يدعم استكشاف البيانات وتمثيلها مرئيًا.
- (Knime) (رابط المصدر): منصة عمل رسومية لعملية التحليل بأكملها: الوصول إلى البيانات، وتحويل البيانات، والاستقصاء الأولي، والتحليلات التنبؤية القوية، والتمثيل المرئي، وإعداد التقارير.

د: مستودعات الشيفرة المصدرية:

- (Github) (رابط المصدر): خدمة استضافة على الويب لمشروعات تطوير البرمجيات التي تستخدم نظام التحكم في الإصدارات (Git)، بما فيها كثير من المشروعات مفتوحة المصدر.
- (Git) (رابط المصدر): نظام موزع للتحكم في النسخ، مجاني ومفتوح المصدر، مصمم للتعامل مع كل شيء من المشروعات الصغيرة إلى المشروعات الكبيرة جدًا بسرعة وكفاءة.
- (Mercurial) (رابط المصدر): أداة مجانية وموزعة لإدارة التحكم في الشيفرة المصدرية. تتعامل بكفاءة مع المشروعات أيًا كان حجمها، وتوفّر واجهة سهلة وبديهية.
- (BitBucket) (رابط المصدر): خدمة استضافة على الويب للمشروعات التي تستخدم أحد نظامي التحكم في الإصدارات (Mercurial) أو (Git).

هـ: أنظمة لتجميع البيانات والشيفرة والوصول إليها وتنفيذها:

- (IPython Notebook) (رابط المصدر): بيئة حوسبة تفاعلية قائمة على الويب، يمكنك فيها الجمع بين تنفيذ الشيفرة والنصوص والرياضيات والرسوم البيانية والوسائط الغنية في وثيقة واحدة.
- (ROpenSci) (رابط المصدر): مجموعة من الحزم البرمجية تتيح الوصول إلى مستودعات البيانات عبر بيئة البرمجة الإحصائية (R).
- (Authorea) (رابط المصدر): معالج نصوص تعاوني على الإنترنت للأوراق العلمية، يتيح كتابة مقالات أصيلة على الويب، حيّة وديناميكية و«قابلة للتنفيذ»، تتضمن النصوص والترميز الرياضي والصور والبيانات. ويدعم حاليًا تضمين الرسوم البيانية المنشأة بـ (d3.js) ودفاتر (IPython) وعرضها.
- (Dexy) (رابط المصدر): أداة متعددة الأغراض لأتمتة المشروعات، للعمل على الوثائق عبر واجهة سطر الأوامر.

و: أدوات برمجية لتشغيل مستودع الوثائق الخاص بك:

- (Invenio) (رابط المصدر): مجموعة برمجيات حرة تمكّنك من تشغيل مكتبتك الرقمية أو مستودع الوثائق الخاص بك على الويب. ويُعد (Invenio) حلًا مثاليًا لتشغيل مستودعات الوثائق المتوسطة إلى الكبيرة الحجم (عدة ملايين من السجلات). ويشترك في تطويره كل من (CERN) و(DESY) و(EPFL) و(FNAL) و(SLAC).
- (Eprints) (رابط المصدر): يُعد (EPrints) من أسهل الطرق وأسرعها لإنشاء مستودعات صغيرة إلى متوسطة الحجم لأدبيات البحث المتاحة وصولًا مفتوحًا، والبيانات العلمية، والرسائل الجامعية، والتقارير، والوسائط المتعددة. طُوِّر في جامعة ساوثهامبتون بالمملكة المتحدة.
- (DSpace) (رابط المصدر): تطبيق جاهز للتشغيل للمستودعات المؤسسية، طوّرته منظمة (Duraspace).

ز: الترخيص والخصوصية:

- مبادرة المصادر المفتوحة (Open Source Initiative) (رابط المصدر): تراخيص المصادر المفتوحة هي التراخيص التي تمتثل لتعريف المصادر المفتوحة: فهي تتيح استخدام البرمجيات وتعديلها ومشاركتها بحرية. ومنها تراخيص (Apache) و(BSD) و(GNU) (GPL) و(MIT) وترخيص (Mozilla) العام.

- أدوات الخصوصية لمشاركة بيانات البحث (Privacy Tools for Sharing Research Data) (رابط المصدر): جهد تعاوني متعدد التخصصات مقره جامعة هارفارد، يهدف إلى المساعدة على تمكين جمع البيانات الشخصية وتحليلها ومشاركتها لأغراض البحث في العلوم الاجتماعية وغيرها من المجالات، مع توفير الخصوصية للأفراد موضوع البحث.

المراجع:

1. Galilei G (1610) Sidereus nuncius. Mapping of the stars series. Thomas Baglioni
2. Galilei G (1618) The Assayer, as translated by Stillman Drake (1957). Available: http://books.google.com/books?id=uSctMwEACAAJ.
3. Drake S (1957) Discoveries and opinions of Galileo: including The Starry Messenger (1610), Letter to the Grand Duchess Christina (1615), and Excerpts from Letters on Sunspots (1613), The Assayer (1623). Anchor Books. Available: http://books.google.com/books?id=P2tmjengyBcC.
4. Holdren J (2013) Increasing public access to the results of scientific research. Memorandum of the US Office of Science and Technology, 22 February 2013.
5. Wren JD (2008) URL decay in MEDLINE–a 4-year follow-up study. Bioinformatics 24: 1381–1385.
6. Barnes N (2010) Publish your computer code: it is good enough. Nature 467: 753–753.
7. Shamir L, Wallin JF, Allen A, Berriman B, Teuben P, et al. (2013) Practices in source code sharing in astrophysics. Astronomy and Computing 1: 54–58.
8. Prlić A, Procter JB (2012) Ten simple rules for the open development of scientific software. PLoS Comput Biol 8: e1002802.
9. Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodi-versity science. BMC Bioinformatics 12: S2.
10. Wallis JC, Rolando E, Borgman CL (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE 8: e67332.
11. Uhlir PE (2012) For attribution – developing data attribution and citation practices and standards: summary of an international workshop. The National Academies Press.
12. FORCE11 (2013) Amsterdam Manifesto on Data Citation Principles. Available: http://www.force11.org/AmsterdamManifesto.