الشفافية وتوثيق المنشأ والمجموعات بوصفها بيانات: منصة (Data Foundry) في المكتبة الوطنية الإسكتلندية

سارة إيمز

الملخص:

يعرض هذا المقال دراسة حالة لعمل المكتبة الوطنية الإسكتلندية في إتاحة مجموعاتها بوصفها بيانات قابلة للقراءة الآلية بطريقة مفتوحة وشفافة. ويبدأ بشرح دوافع إطلاق خدمة البحث العلمي الرقمي في المكتبة عام 2019 وأهدافها الخمسة ومستويات الخدمة الثلاثة التي تقدمها، ثم يصف التغييرات التي أُدخلت على برنامج الرقمنة لإنتاج منصة (Data Foundry) لتقديم البيانات المفتوحة. ويتناول ثلاث طرائق لترسيخ الشفافية: التواصل مع الجمهور عبر وسائل التواصل الاجتماعي لإبراز العمل غير المرئي وراء تحويل المجموعات إلى بيانات، ووضوح عرض مجموعات البيانات على الإنترنت بما يرافقها من نقاط رئيسية وبيانات حقوق وتنزيلات متدرجة ومعرّفات دائمة، وتوثيق منشأ البيانات وأسباب رقمنة المجموعات في ملفات (METS). ويخلص إلى أن المكتبات أصبحت منتجة لمجموعاتها وبياناتها، وأن الشفافية في عملياتها وقراراتها ضرورية للحفاظ على الثقة بها وتعزيز قيمة مجموعاتها بوصفها مادة بحثية.

الكلمات المفتاحية: المجموعات بوصفها بيانات، الشفافية، مصدر البيانات، البحث العلمي الرقمي، المكتبات الوطنية

1. المقدمة:

في عام 2017، كتب توماس باديلا (Thomas Padilla) عن «ضرورة المجموعات بوصفها بيانات» بالنسبة إلى المكتبات ومؤسسات التراث الثقافي، وركّز فيها على ثلاثة مفاهيم رئيسية: القدرة على توليد استخدامات جديدة، والوضوح، والإبداع (Padilla, 2017, p. 2). وفي هذا السياق يوضح أنه «لكي تصبح المجموعات بوصفها بيانات قابلة للاستخدام، يجب إظهار العمليات التي تُنشأ بها وشرحها» (p. 3)؛ فمنشأ البيانات (data provenance) والشفافية أمران أساسيان لإصدار مجموعات بوصفها بيانات مفيدة وقابلة للاستخدام. ويشير باديلا إلى أن عمليات

صنع القرار في المكتبات والتحويلات التي تخضع لها المجموعات والبيانات لا تُتاح للمستخدمين في العادة - وهي معلومات كثيرًا ما تحدد سبب إتاحة مجموعات معينة دون غيرها، فتؤثر في الأجندات البحثية وفي المعلومات المتاحة للجمهور على حد سواء.

وبعد عامين من هذه الورقة، أطلقت المكتبة الوطنية الإسكتلندية خدمتها للبحث العلمي الرقمي (Digital Scholarship Service) وسط اهتمام متنامٍ بمفهوم «المجموعات بوصفها بيانات» (collections as data) النابع من عمل باديلا وزملائه في مشروع (Always Already Computational)، الذي امتد من 2016 إلى 2018 (Padilla et al., 2019a)، ومن امتداده الممول من مؤسسة ميلون، (Collections as Data: Part to Whole) (Padilla et al., 2019b)، وبدأت في إصدار مجموعاتها بصيغة قابلة للقراءة الآلية. وهذا العمل جزء من تحول أوسع يُلاحَظ حاليًا في المعارض الفنية والمكتبات والأرشيفات والمتاحف (GLAMs) نحو إصدار المجموعات إصدارًا مفتوحًا وبصيغة قابلة للقراءة الآلية، ونحو دعم المستخدمين الراغبين في الوصول إلى هذه المجموعات أو تحليلها باستخدام الأساليب الحاسوبية. فمجتمع مختبرات (GLAM) الدولي (International GLAM Labs Community)[1] مثلًا يدعو إلى التجريب الابتكاري على نمط «المختبر» في المجموعات الرقمية، وأصبحت فرق البحث العلمي الرقمي وخدماته ووظائفه مألوفة في المؤسسات الأمريكية، وتتحول على نحو متزايد إلى جزء أساسي من المكتبات البحثية الأوروبية (مثل المكتبة البريطانية) وغيرها.

وفي هذا السياق الأوسع لنشاط «المجموعات بوصفها بيانات»، تستكشف دراسة الحالة هذه، المبنية على عرض قُدِّم في مؤتمر (LIBER2020) (Ames, 2020a)، كيف صُمِّمت منصة المكتبة لتقديم البيانات المفتوحة، (Data Foundry)، لتتضمن منشأ البيانات؛ وكيف تعمل خدمة البحث العلمي الرقمي في المكتبة على ترسيخ الشفافية في عمليات المكتبة؛ وما الآثار العملية لهذا النشاط وفوائده وتحدياته. ومع أن مصدر المواد (material provenance) كان دائمًا موضوعًا مهمًا للمكتبات ولحقل تاريخ الكتاب، فإن هذه الورقة تستكشف سبب أهمية هذا التوثيق وهذه الشفافية للممارسات الحالية في المكتبات، ولا سيما في ما يتعلق بالرقمنة وإصدار البيانات، وكيف يمكن اتخاذ خطوات لإتاحة هذه المعلومات للمستخدمين. وهي تقرّ بأن المكتبات أصبحت على نحو متزايد منتجة للبيانات - ومنتجة لمجموعاتها الخاصة - وبأن ذلك أمر إشكالي، وتستكشف الطرق التي تتناول بها المكتبة الوطنية الإسكتلندية هذه المسألة.

2. إطلاق خدمة للبحث العلمي الرقمي:

دعا كثير من الحوار الدائر حول البحث العلمي الرقمي (digital scholarship) والإنسانيات الرقمية (digital humanities) في المكتبات إلى النظر إلى هذا العمل بوصفه مجالًا بحثيًا، لا مجرد خدمة: فقد قالها مونيوث (Muñoz) (2012) صراحة: «الإنسانيات الرقمية في المكتبة ليست خدمة»، ونادت تقارير أحدث بدور المكتبات شريكًا في مشروعات الإنسانيات الرقمية: فقد دافعت رابطة (LIBER) بقوة عن المكتبة بوصفها شريكًا (Wilms et al., 2019)، ويدور عمل رابطة المكتبات البحثية في المملكة المتحدة (RLUK) حول المكتبات بوصفها «مزوّدًا وشريكًا ورائدًا» (Greenhall, 2019, p. 5). وفوائد التعاون والشراكة في هذا المجال للباحثين والمكتبات واضحة: فكلاهما يستفيد من خبرة الآخر. وفي هذا السياق، لماذا إذن إطلاق «خدمة» للبحث العلمي الرقمي؟

أُنشئت خدمة البحث العلمي الرقمي في المكتبة الوطنية الإسكتلندية في سبتمبر 2019، ولها خمسة أهداف رئيسية:

- تشجيع استخدام أساليب البحث الحاسوبية مع المجموعات وتمكينه ودعمه؛
- ضمان استخدام المجموعات بكامل إمكاناتها؛
- إرساء ثقافة في المكتبة تدعم البحث العلمي الرقمي؛
- ممارسة الشفافية في عمليات إنشاء البيانات لدينا وتعزيزها؛
- استشراف مستقبل البحث. (Ames, 2020b)

تركز دراسة الحالة هذه إلى حد كبير على رابع هذه الأهداف، غير أن الغايات تتمحور عمومًا حول ثلاثة مسارات من النشاط: إتاحة مجموعات المكتبة بوصفها بيانات؛ والتواصل والتعاون الخارجيين؛ والتواصل الداخلي داخل المكتبة. ونتيجة لذلك، تستطيع المكتبة أن تقدم لمستخدميها ثلاثة مستويات من خدمة البحث العلمي الرقمي:

- الخدمة الذاتية (الاستفادة من مجموعات البيانات أو الأدوات المنشورة)؛
- الخدمة الذاتية للمجموعات أو الأدوات مع قدر من وقت الاستشارة من الموظفين (مثل المساهمة في محاضرة دراسية أو تقديم خبرة في المجموعات)؛
- الشراكة في المشروعات الممولة.

وتغطي هذه المستويات مجتمعةً المستخدمين الذين يحتاجون إلى قليل من التفاعل مع المكتبة لإنجاز مهمتهم، وأولئك الذين يرغبون في التعاون في مشروعات أكبر.

وقد أدت أسباب عدة إلى تطوير «خدمة» على هذا النحو، منها اعتبارات عملية: فـ«الخدمة» مصطلح مفهوم داخل المكتبات، يتيح تطوير الخدمات على نحو منهجي ومنسّق وإبرام اتفاقيات مستوى الخدمة بين الفرق. و«الخدمة»، وإن كانت مصطلحًا إشكاليًا، نشاطٌ مستمر أيضًا: فربط هذا العمل بمصطلحات محدودة زمنيًا مثل «مشروع» أو «برنامج» لن ينقل بدقة ضرورة أن يصبح هذا النشاط جزءًا من العمل المعتاد. يضاف إلى ذلك أن المكتبات تحتاج إلى قدر من تقديم الخدمة لتمكين التعاون أصلًا: فتقديم المجموعات بصيغة قابلة للقراءة الآلية، أو تقديم الدعم التقني، يستلزم إنشاء خدمة وإشراك فرق متعددة في تقديمها. وأخيرًا، فإن بعض المستخدمين لا يريدون التعاون أو لا يحتاجون إليه، بل يريدون الاستفادة من بيانات المكتبة المرخصة ترخيصًا مفتوحًا ثم العودة إلى العمل على مشروعاتهم الخاصة. وفي الوقت نفسه، تمكّن هذه اللبنات الأساسية من التعاون في إطار خدمة ممولة، مما يتيح للمكتبة أن تكون شريكًا في مشروعات البحث العلمي الرقمي أيضًا.

وقد تضمّن إطلاق هذه الخدمة نشاطًا على مستوى المكتبة كلها لإتاحة المجموعات بوصفها بيانات. وركّز النشاط الأولي على بحوث المستخدمين: فبناءً على محادثات مع باحثين في الإنسانيات الرقمية وتقييم لاحتياجات المستخدمين، فضلًا عن اختبارات قابلية الاستخدام المبكرة لمنصة البيانات المفتوحة الجديدة في المكتبة، Data) (Foundry، حُدِّدت ثلاث فئات عامة من المستخدمين تتراوح بين «المبتدئ» و«المتقدم» - ففي أحد طرفي الطيف من يريد استخدام أدوات إلكترونية مع ملفات نصية، وفي الطرف الآخر من يملك مهارات تقنية متقدمة ولا يبالي بالصيغة التي تُقدَّم بها المواد المرقمنة (لكنه على دراية بالمعايير وأفضل الممارسات). أما في الوسط فكان هناك

مستخدم ذو احتياجات أكثر تعقيدًا، محدود المهارات التقنية، لكنه يدرك قيمة الصيغ والمناهج المختلفة بالنسبة إلى الأسئلة البحثية. وكان هذا المستخدم أميل إلى توظيف مساعد بحث لتنفيذ أي عمل تقني، لكنه أراد الوصول إلى البيانات بسهولة وسرعة لمعرفة ماهيتها ومدى ملاءمتها لعمله. وقد مكّننا قرار تلبية احتياجات هذا المستخدم الوسيط من تلبية كثير من احتياجات المستخدمين المبتدئين والمتقدمين أيضًا.

ولضمان قدرتنا على تقديم صيغ متسقة من برنامج الرقمنة لدينا، كان علينا بعد ذلك أن نغيّر نهجنا في الرقمنة. فالبحث العلمي الرقمي يقدّم حالة استخدام جديدة للرقمنة. فبعد أن كانت المكتبة ترقمن سابقًا لمعارض الصور الإلكترونية أو لطلبات المستخدمين فقط، أصبح عليها الآن أن تراعي جمهورًا جديدًا. ويتطلب تمكين الاستخدامات الحاسوبية للمجموعات بعض الخطوات الإضافية، منها: تقييم دقيق للحقوق لضمان إمكان إصدار المجموعة وفق تعريف البيانات المفتوحة؛ وإنشاء صيغ ملفات مختلفة (مخرجات التعرف الضوئي على الحروف Optical Character) (Recognition, OCR بصيغة (ALTOXML) ، أي «لغة الترميز القابلة للتوسع لكائنات التخطيط والنص المحللة» (Analysed Layout and Text Object Extensible Markup Language)، التي توفر معلومات عن تخطيط الصفحة)؛ وبيانات وصفية لوصف الكائن الرقمي ((METS) - «معيار ترميز البيانات الوصفية ونقلها» Metadata) (Library of Congress, 2020) Encoding and Transmission Standard))؛ وحلول تخزين ملائمة للتنزيلات الكبيرة (تستخدم منصة (Data Foundry) حاويات (Amazon Web Services))؛ ومعرّفات دائمة لتمكين الاستشهاد (معرّفات الكائنات الرقمية (Digital Object Identifiers, DOIs)). وكان لا بد من اتخاذ قرارات أخرى بشأن المعايير المتسقة، وأحجام الصور، وأي مجموعات البيانات تُختار لتُتاح أولًا، وما المعلومات الإضافية عن هذه العمليات التي يمكن جمعها وإتاحتها - أي كيف نكون شفافين ونعرض مجموعات بيانات المكتبة في سياقها. وفي أعقاب هذا النشاط، نُشرت منصة (Data Foundry)، منصة تقديم البيانات التابعة لخدمة البحث العلمي الرقمي، وأُطلقت على الإنترنت في سبتمبر 2019 (National Library of Scotland, 2019a).

3. ترسيخ الشفافية في ممارسات المكتبة:

وسط هذا العمل، كيف تعزز المكتبة الشفافية تماشيًا مع الهدف الرابع لخدمة البحث العلمي الرقمي، وما تحديات ذلك؟ ثم لماذا نكون شفافين - ما قيمة الشفافية؟ تمكّن الممارسات المفتوحة المكتبة من نقل المعلومات عن العمليات الداخلية، وعن أي تحويلات خضعت لها المجموعات أو البيانات، وعن صنع القرار المتعلق بذلك. وسيمكّن ذلك المكتبات أيضًا من الاعتراف بالتحيزات والحد منها - أي من المصارحة بالمواضع التي قد لا تكون فيها عمليات صنع القرار مثالية، والمواضع التي ربما أخفقت فيها، والمواضع التي لا تزال فيها قرارات أو ممارسات تاريخية تخلّف آثارها حتى اليوم. وهذا يمكّن المكتبات ومستخدميها من تقديم سرديات مضادة: أي استكشاف ما هو غائب، وما لم يُقَل، وما هي الفجوات القائمة، والبدء في تصحيح ذلك.

وحتى الآن، ركزت خدمة البحث العلمي الرقمي على ثلاث طرائق رئيسية لتعزيز الشفافية وترسيخها في ممارسات المكتبة وعملياتها.

3.1. التواصل بشأن الشفافية:

لما كان نطاق عمل المكتبة يشمل جمهورًا واسعًا - لا مجتمع الباحثين وحده - فقد كانت وسائل التواصل الاجتماعي إحدى الطرق المباشرة للتواصل مع عدد كبير من المستخدمين. ومن أمثلة ذلك نشاط المكتبة على تويتر في يوم الإنسانيات الرقمية 2020 (#DayOfDH2020)، حين سلّطت خدمة البحث العلمي الرقمي الضوء على «العمل غير

المرئي» والبحث العلمي الرقمي. وكما يوضح بوستر (Poster) وزملاؤه (2016)، يمكن أن يشمل العمل غير المرئي «العمل المرئي الذي يؤديه أشخاص غير مرئيين (عمال المنازل، أمناء المكتبات)» (p. 7)، ويكتب غرابان (Graban) وزملاؤه (2019) أن «كثيرًا من العمل الذي يجري في مشروعات الإنسانيات الرقمية غير مرئي لمستويات متعددة من السلطة». والمنتج النهائي لعمليات الرقمنة - أي إتاحة مواد مرقمنة مقيّمة الحقوق على الإنترنت - مثال بارز على هذا العمل، إذ يمتد الجهد المبذول لرقمنة مادة واحدة عبر فرق وعمليات متعددة في المكتبة.

وقد تضمن إبراز ذلك «سلسلة تغريدات» على تويتر تشرح العمليات التي تمر بها مجموعة ما عند تحويلها إلى بيانات، بهدف الكشف عن صنع القرار والتحويلات، وتشجيع فهم حجم العمل وعدد الفرق المشاركة فيه (National Library of Scotland, 2020). ومع الحرص على صياغة السلسلة لجمهور واسع، أُدرجت خطوات منها عمل أمناء المجموعات (curatorial)، وتقييم الحقوق، والحفظ والصون، وإحضار الكتب، والرقمنة (الالتقاط)، والإيداع (ingest) والحفظ، كما أُدرجت معلومات عن العمل الذي تؤديه فرق البيانات الوصفية والمطورين؛ وتفاصيل عن ماهية «مجموعة البيانات» (dataset)، مع إبراز أهمية الشفافية في هذه العمليات؛ وموقع تخزين الكائنات الرقمية؛ ودور المعرّفات الدائمة.

ومكّن ذلك المكتبة من أن تُطلع المستخدمين على ما يجري «خلف الكواليس» في أحد مجالات نشاط خدمة البحث العلمي الرقمي، وهم في بيوتهم، مما أتاح لنا إبراز الجهد البشري - ومن ثم التحيزات - المبذول في مجموعاتنا. كما أسهم في بدء حوار مع مستخدمي المكتبة حول انعكاسات هذا العمل على المجموعات التي تُصدَر وعلى ما يمكن استخدامها فيه. ونُشرت «سلسلة تغريدات» مماثلة لشرح تقنية التعرف على النص المكتوب بخط اليد (Handwritten Text Recognition, HTR) للجمهور العام، بمناسبة إصدار المكتبة أول مجموعة بيانات مولّدة بالذكاء الاصطناعي (AI) باستخدام منصة (Transkribus) (National Library of Scotland, 2021)، مما مكّن المكتبة من التعريف بالمشاركة البشرية في العمل المولّد بالذكاء الاصطناعي، وبالمشكلات المرتبطة بذلك.

3.2. قيمة الوضوح في عرض البيانات على الإنترنت:

ماذا تعني كل هذه العمليات بالنسبة إلى الطريقة التي ننقل بها البيانات إلى المستخدمين؟ وسط الفرق والتحويلات المتعددة التي تمر بها المادة وهي تصبح «مجموعة بيانات»، يكتسب عرض البيانات نفسها، وطريقة تقديمها للجمهور، أهمية كبيرة. ولدى المكتبة الوطنية الإسكتلندية خطة لنشر البيانات المفتوحة، تحدد مستوى البيانات المفتوحة الذي تعمل المكتبة وفقه (ثلاث نجوم)، والصيغ المستخدمة، ووتيرة نشر البيانات ومكان نشرها (National Library of Scotland, 2019b). ويكمّلها سجلّ للبيانات المفتوحة يضم جميع مجموعات البيانات التي نشرتها المكتبة، وتواريخها، ووتيرة تحديثها، وصيغ ملفاتها؛ ويمكن تنزيل كليهما من منصة (Data Foundry) (National Library of Scotland, 2019c). ويتيح ذلك للمكتبة أن تحدد توقعات واضحة بشأن موعد نشر البيانات وكيفيته.

وتُنشر مجموعات البيانات نفسها على موقع (Data Foundry). وعرض مجموعات بيانات التراث الثقافي بالغ الأهمية: فنظرًا إلى حجمها وضرورة تنزيل البيانات للبدء في استكشافها، قد يبدو مفهوم مجموعة بيانات المجموعات غامضًا إلى حد ما. ويلاحظ باديلا (2017) أن «جودة البيانات [لا] يُشار إليها عادة» (p. 3)، وتصبح هذه الجودة، إلى جانب المعلومات عن حجم البيانات - معبَّرًا عنه بالبايتات والكلمات معًا - مهمة لفهم مجموعات بيانات التراث الثقافي. فمن خلال إصدار عدد من النقاط الرئيسية عن كل مجموعة بيانات، مثل عدد الملفات؛ وعدد الكلمات

والأسطر؛ ومستوى تنقيح مخرجات التعرف الضوئي على الحروف؛ والنطاق الزمني للبيانات - أي التصريح بخصائص المدوّنة في جوهر الأمر - فضلًا عن تقديم مقدمة للمجموعة يكتبها أمينها، تمنح مجموعات البيانات على منصة (Data Foundry) هذا المفهوم المجرد ملامح واضحة.

علاوة على ذلك، تمر جميع المجموعات المرقمنة في المكتبة بعملية مسبقة لتقييم الحقوق، ولذلك فإن لجميع مجموعات البيانات على منصة (Data Foundry) تراخيص أو بيانات حقوق واضحة، ولا تدّعي المكتبة حقوق نشر إضافية على المجموعات. وتُضاف بيانات الحقوق أو التراخيص إلى كل مجموعة بيانات؛ وبعض المجموعات، مثل الموسوعة البريطانية (Encyclopaedia Britannica)، التي تمتد عبر فترات زمنية طويلة، تندرج تحت عدد من بيانات الحقوق.

وتُقدَّم مجموعات بيانات (Data Foundry) في صورة تنزيلات بسيطة للحد من العوائق التقنية أمام الاستخدام، مع خطط مستقبلية لإتاحة الوصول إلى بعض المجموعات عبر واجهة برمجة التطبيقات (Application Programming Interface, API). وتتضمن كل مجموعة بيانات ملف جرد وملف «اقرأني» (readme) يشتمل على معلومات عامة عن مجموعة البيانات، منها أعداد الملفات وصيغها، وتاريخ النشر والمراجعات اللاحقة، ومعلومات الحقوق. وقد أدى تضمين ملفات (METS) و(ALTO) وملفات الصور في هذه الحزم إلى تجاوز حجم بعض مجموعات البيانات 40 غيغابايت بعد فك الضغط، ولذلك فإن تخزينها وإتاحتها في السحابة يضمن سرعة التنزيل. وتوفر منصة (Data Foundry) تنزيلات متدرجة للبيانات، لتمكين المستخدمين من تنزيل عينة قبل الالتزام بتنزيل مجموعة البيانات كاملة، ولتمكين المستخدمين الذين لا يرغبون إلا في الوصول إلى ملفات النص الخالص (كما في حالة المستخدم «المبتدئ» المشار إليه أعلاه).

وعند تجميع البيانات وضغطها ونشرها على الإنترنت، يُضاف إليها بعد ذلك معرّف (DOI) بوصفه معرّفًا دائمًا لتمكين الاستشهاد. وتهدف ميزات منصة (Data Foundry) وملفاتها مجتمعةً إلى تقليل الجهد اللازم لتكوين نظرة عامة على البيانات قبل الالتزام بتنزيلها.

3.3. منشأ البيانات:

وأخيرًا، فإن توضيح منشأ بيانات التراث الثقافي أمر أساسي للحفاظ على الصلة بين مجموعة البيانات والمجموعة المطبوعة. ومن المهم لأسباب عدة ضمان فهم القراء من أين جاءت البيانات التي يستخدمونها، وما العمليات التي خضعت لها، ولماذا وُجدت أصلًا؛ غير أن كيفية نقل هذه المعلومات في مجال لا توجد فيه حاليًا إلا معايير متفق عليها قليلة تمثل تحديًا.

ويمكّن استخدام معيار (METS) المكتبة من تضمين معلومات المصدر، مثل البيانات الوصفية الببليوغرافية المتعلقة بالمعلومات الببليوغرافية ومعلومات المقتنيات؛ والبيانات الإدارية، بما فيها المعلومات عن التقاط الصور وإنشاء الملفات والحقوق؛ وقوائم الملفات؛ والبيانات الوصفية البنيوية. وهذا النوع من المعلومات، ولا سيما معلومات المصدر التقنية التي تفصّل عملية الرقمنة، ضروري لإعادة بناء العملية الفعلية من المجموعة المطبوعة إلى مجموعة البيانات، بما ينسجم مع حركة البحث القابل لإعادة الإنتاج (Ames & Lewis, 2020; Padilla, 2017). غير أن ذلك لا يمتد إلى المعلومات عن سبب رقمنة مادة أو مجموعة ما أصلًا: فكما يشير باديلا (2017)، «لا تكشف المكتبات عادةً عن سبب إتاحة بعض المجموعات دون غيرها» (p. 3).

لماذا يُعد ذلك مهمًا؟ كما يشير إيمز ولويس (Ames and Lewis, 2020, p. 4)، فإن مجموعة المكتبة في مجملها مفهوم إشكالي - كثيرًا ما تتشكل بفعل تحيزات في ممارسات الاقتناء التاريخية (والحالية). ويضاف إلى ذلك أن بعض مواد هذه المجموعة فقط هي التي تُرقمن، ويتوقف ذلك على عوامل منها حقوق النشر، والحفظ والصون، وعمليات الاختيار الداخلية. ومن بين هذه المجموعات المرقمنة، لا يصلح بعد ذلك إلا بعضها لإتاحته مجموعاتِ بيانات - تبعًا لدقة تقييم الحقوق لإتاحة المجموعات بوصفها بيانات «مفتوحة»، أو جودة التعرف الضوئي على الحروف، أو الممارسات السابقة في إدارة الأصول الرقمية. «إن كيفية عرض هذه المجموعات في سياقها، وكيف يمكن لهذه المجموعات المقلّصة أن تصبح ممثِّلة لفهم أوسع وضمني لـ"الثقافة"، أمر إشكالي» (Ames & Lewis, 2020, p. 4) - وهو يؤثر مباشرة في الموارد المتاحة للحقول البحثية والاستخدامات الإبداعية والتعلم المدرسي. علاوة على ذلك، فإن مدى تمثيل هذه المجموعات للمجتمعات المتنوعة التي تخدمها المكتبات كثيرًا ما يكون إشكاليًا - وكذلك كيفية شرحنا لمستخدمي المكتبة مشكلةَ الإتاحة المنحازة للمواد الرقمية هذه.

ونتيجة لذلك، تتضمن مجموعات البيانات على منصة (Data Foundry) معلومات عن أسباب إنتاج مجموعات البيانات وكيفية إنتاجها، لضمان فهم أكمل للمجموعات: وهذا جانب مهم من وضع المجموعات الرقمية والمرقمنة للمكتبات في سياقها. وتعمل المكتبة على إتاحة المعلومات على نحو منهجي عن سبب رقمنة مجموعات معينة، والجهة التي موّلتها، وما إذا كانت رقمنتها قد أُسندت إلى جهات خارجية، وما قرارات أمناء المجموعات التي اتُّخذت عند اقتراح المجموعة للرقمنة. ويتيح ذلك للمكتبة تقديم معلومات عن سبب نشوء المجموعة في صورتها الرقمية الحالية، كما يمكّننا من دمج حرية المعلومات في عمليات إنشاء البيانات لدينا: فالانفتاح بشأن هذا النوع من المعلومات يمكن أن يحقق فوائد إضافية في كفاءة العمل. ولعدم وجود معايير حالية لكيفية تقديم هذه المعلومات، فإنها تُقدَّم حاليًا في حقل نصي حر غير مهيكل باسم «(Other)» (أخرى) في ملفات (METS) الخاصة بمجموعات البيانات، بوصف ذلك وسيلةً لإبقاء هذه المعلومات مباشرةً إلى جانب مجموعات البيانات.

4. مسؤوليات المكتبات: الاستنتاجات:

يشير كل هذا النشاط إلى تغيّر مسؤوليات المكتبات. فتقليديًا، كانت المكتبات جهات اقتناء: فهي تشتري المواد أو المجموعات، أو تُهدى إليها، ثم تصفها وفق معايير (يُفترض أنها) موضوعية ومحايدة - وهو أمر إشكالي في حد ذاته - ثم تتيحها للمستخدمين للإعارة أو البحث. غير أننا، مع التحول نحو الرقمي، نرى المكتبات تصبح على نحو متزايد منتجة لمجموعاتها الخاصة: ففي المكتبة الوطنية الإسكتلندية مثلًا، لا تُنتج البيانات عند معالجة النصوص المرقمنة بالتعرف الضوئي على الحروف فحسب، بل هناك أيضًا بيانات تنشئها معدات التصوير، وبرمجيات التعرف الضوئي على الحروف نفسها، وأدوات الإنتاج الرقمي المستخدمة في المعالجة اللاحقة. ونحن، مثل كثير من المكتبات، ننتج أيضًا البيانات الوصفية، وننتج بيانات مؤسسية «إدارية».

ما مسؤوليات المكتبات إذن بوصفها منتجة للبيانات؟ كيف تصف المكتبات بياناتها الخاصة وتنقل المعلومات عن كيفية إنتاجها وسببه؟ كيف يمكنها - وهل يمكنها أصلًا - أن تضع عملياتها الخاصة في سياقها بدقة؟ وكيف تظل في موقع الثقة، وسط خلفية من «الأخبار الزائفة» والذكاء الاصطناعي والتزييف العميق (Deepfakes)، وهي التي تنشئ تلك المعلومات بنفسها؟ هنا تصبح الشفافية أمرًا أساسيًا: فكلما استطاعت المكتبات أن تضع مجموعات البيانات في سياقها، وأن تكون واضحة بشأن العمليات ومراحل صنع القرار التي مرت بها، زادت قيمتها بوصفها

مادة بحثية. ويظل هذا النشاط عملًا قيد التطوير، غير أن العمل الذي تؤديه المكتبات حول «ضرورة المجموعات بوصفها بيانات» يزداد أهمية يومًا بعد يوم، ونحن نسعى إلى ضمان بقاء مجموعات المكتبات ذات صلة وقابلة للاستخدام ومفيدة مع تغيّر التقنيات والمنهجيات - والعالم الأوسع من حولنا.

المراجع:

1. International GLAM Labs Community. https://glamlabs.io/.
2. Ames, S. (2020a). Transparency, provenance and collections as data: the National Library of Scotland's Data Foundry. [Conference presentation]. 50th LIBER Annual Conference – LIBER2020, online. http://doi.org/10.5281/zenodo.3921784.
3. Ames, S. (2020b). Digital scholarship and the Data Foundry. 21st Century Curator talk series, British Library. http://doi.org/10.5281/zenodo.3862050.
4. Ames, S., & Lewis, S. (2020). Disrupting the library: Digital scholarship and big data at the National Library of Scotland. Big Data & Society, 7(2), 1–7. https://doi.org/10.1177/2053951720970576.
5. Graban, T. S., Marty, P., Romano, A., & Vandegrift, M. (2019). Questioning collaboration, labor, and visibility in Digital Humanities research. Digital Humanities Quarterly, 13(2), 1–9. http://www.digitalhumanities.org/dhq/vol/13/2/000416/000416.html.
6. Greenhall, M. (2019). Digital Scholarship and the Role of the Research Library: The Results of the RLUK Digital Scholarship Survey. Research Libraries UK report. https://www.rluk.ac.uk/wp-content/uploads/2019/07/RLUK-Digital-Scholarship-report-July-2019.pdf.
7. Library of Congress. (2020). METS - Metadata encoding & transmission standard - Official website. https://www.loc.gov/standards/mets/.
8. Muñoz, T. (2012, August 19). Digital humanities in the library isn't a Service. [blog] http://trevormunoz.com/notebook/2012/08/19/doing-dh-in-the-library.html.
9. National Library of Scotland. (2019a, September 2). Data Foundry launched. https://www.nls.uk/news/archive/2019/09/data-foundry.
10. National Library of Scotland. (2019b, September 2). Open Data publication plan. https://data.nls.uk/download/national-library-of-scotland-open-data-publication-plan.pdf.
11. National Library of Scotland. (2019c, September). Standards-Open Data Plan. https://data.nls.uk/about/standards/.
12. National Library of Scotland. (2020, April 29). Data foundry. Twitter. https://twitter.com/natlibscot/status/1255478042207621126.

13. National Library of Scotland. (2021, January 28). Handwritten Text Recognition (HTR). Twitter thread. https://twitter.com/natlibscot/status/1354751444004597762.

14. Padilla, T. (2017, February 15). On a Collections as data imperative. Library of Congress. https://labs.loc.gov/static/labs/work/reports/tpadilla_OnaCollectionsasDataImperative_final.pdf.

15. Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019a). Final report – Always already computational: Collections as data. Zenodo. https://doi.org/10.5281/zenodo.3152935.

16. Padilla, T., Scates Kettler, H., Varner, S., & Shorish, Y. (2019b). Collections as data: Part to whole. Github. https://collectionsasdata.github.io/part2whole/.

17. Poster, W. R., Crain, M., & Cherry, M. A. (2016). Introduction: Conceptualizing invisible labor. In M. Crain, W. R. Poster, & M. A. Cherry (Eds.), Invisible labor: Hidden work in the contemporary world (pp. 3-27). University of California Press. https://doi.org/10.1525/9780520961630-003.

18. Wilms, L., Derven, C., O'Dwyer, L., Lingstadt, K., Verbeke, D., & Lefferts, M. (2019). Europe's Digital Humanities landscape: A study from LIBER's Digital Humanities & Digital Cultural Heritage Working Group. Zenodo. http://doi.org/10.5281/zenodo.3247286.