

من يتولى ماذا؟ تكامل خدمات إدارة البيانات البحثية في زيورخ

ماتياس توفه، كاترينا باريلاري



المخلص:

يعرض المقال تكامل خدمات إدارة البيانات البحثية في المعهد التقني الاتحادي بزيورخ، وتوزيع المسؤوليات بين المكتبة وخدمات تقنية المعلومات العلمية والوحدات الاستشارية الأخرى. وينظم العرض وفق ثلاث مراحل: التخطيط، والإدارة النشطة، والنشر والحفظ. ويشرح التدريب والاستشارات وخطط إدارة البيانات، ودور (openBIS) في توثيق التجارب وتنظيم البيانات خلال العمل، وتكامل المستودع المؤسسي وأدوات الوصف مع أرشيف الحفظ الطويل. ويناقش التواصل المشترك لمساعدة الباحثين على اختيار الخدمة المناسبة، مع الحفاظ على مرونة الاختيار بين التخصصات. كما يتناول تحديات السياسات والتخزين ونقل الملفات الكبيرة واستدامة الموارد، والتعاون الوطني لتطوير الخدمات. ويقدم التجربة بوصفها إطارًا متطورًا يحتاج التكيف مع احتياجات الباحثين ومتطلبات الممولين وتغيير أدوات البحث الرقمي.

الكلمات المفتاحية: إدارة البيانات البحثية، تكامل الخدمات، (openBIS)، الحفظ الرقمي، المكتبات الجامعية

مقدمة:

يحتاج المعهد التقني الاتحادي بزيورخ، مثل جميع الجامعات، إلى الاستجابة لمتطلبات متزايدة في إدارة البيانات البحثية. وينشأ بعضها من تطور مناهج العلم والتقنية، وبعضها من ضرورة نقل معايير الممارسة العلمية الجيدة الراسخة إلى إجراءات رقمية في معظمها. ويأتي بعضها من رغبة الممولين في زيادة البيانات المفتوحة لتيسير إعادة استخدامها وتعظيم قيمة البيانات الممولة تمويلًا عامًا في إنتاج المعرفة.

يركز المعهد، بوصفه جامعة تقنية، على العلوم والتقنية. ورغم وجود خدمات دعم منذ سنوات، فإنها لم تكن متكاملة أصلًا عبر دورة حياة البيانات كلها. وأصبحت الحاجة إلى دعم الدورة كاملة أكثر إلحاحًا مؤخرًا، خصوصًا مع تغيير سياسات الممولين.

وبأرقام 2019، تجاوز عدد الطلاب 22 ألفاً، منهم أكثر من أربعة آلاف طالب دكتوراه. وتخدم الخدمات الباحثين، بمن فيهم طلبة الدكتوراه وباحثو ما بعد الدكتوراه والموظفون العلميون الأعلى خبرة، بما يعادل أكثر من ستة آلاف وظيفة بدوام كامل. وجميعهم مستخدمون محتملون لأنهم ينتجون بيانات، مع اختلاف مسؤولياتهم وأنشطتهم واحتياجاتهم. كما ينبغي خدمة التخصصات العلمية في أقسام المعهد الستة عشر.

يعرض المقال الخدمات وتكاملها وتعاون الوحدات المعنية والأدوات المستخدمة. ولأن الوضع ليس ثابتاً أو نهائياً، يناقش نقاط قوته وضعفه وتطويره مستقبلاً.

الإطار التنظيمي:

تحكم إدارة البيانات لوائح الجامعة وما يتجاوزها. وتوجد منذ مدة إرشادات النزاهة البحثية لعام 2011، التي تحدد متطلبات عامة، وتتوقع من المجتمعات العلمية معرفة المعايير المعترف بها دولياً أو إنشاؤها عند غيابها، وتوضح مسؤوليات الباحثين الرئيسيين والموظفين. ويجب كذلك احترام الالتزامات المتعلقة بحماية البيانات والرفق بالحيوان والأخلاق والملكية الفكرية وغيرها.

وفي 2017، دفع الصندوق الوطني السويسري للبحث العلمي، وهو الممول العام الرئيس، باتجاه البيانات المفتوحة، باشتراط خطة إدارة بيانات مع معظم طلبات المنح تقريباً.

يحكم الإطار الدورة كلها، لكن تأثيره يختلف بين المراحل، وتزيد المتطلبات التخصصية تنوع الأنشطة. ولذلك يواجه الباحثون وخدمات الدعم مطالب متنوعة، وتوجد جهات اتصال بحسب السؤال، منها نقل التقنية ومكتب البحث والخدمات القانونية ولجنة الأخلاق.

والجهازان الرئيسيتان للخدمات العملية هما قسم خدمات تقنية المعلومات العلمية في الإدارة المركزية للتقنية، ومجموعة إدارة البيانات والتنظيم الرقمي في مكتبة المعهد. وتعتمدان على البنية المركزية. وللتبسيط والتواصل تُجمع الدورة في ثلاث مراحل: التخطيط، والإدارة النشطة، والنشر والحفظ.

المرحلة الأولى: التخطيط:

تشمل هذه المرحلة الاستفسارات والتخطيط لإعداد مقترح بحث سليم: أنواع البيانات وحجمها وطريقة استخدامها ومشاركتها ونشرها وحفظها. وتقدم المكتبة وخدمات التقنية مقاً تدريباً وورشاً تمهيدية مرتين سنوياً. ومنذ 2019 تنظمان مدرسة صيفية سنوية للدكتوراه وما بعد الدكتوراه؛ وعُقدت نسخة سنة المقال إلكترونياً خلال أسبوع وحققت نجاحاً كبيراً.

وتُنظم أنشطة حسب الطلب للأقسام والمعاهد والمجموعات، مع تكييفها محلياً، واستشارات فردية يشارك فيها موظفو الجهتين بخبراتهم وخدماتهم. وتقدمها المكتبة ضمن «احجز اختصاصي مكتبات».

وفي 2019 قُدمت ست ورش، مدة كل منها نصف يوم، وتسعة تدريبات مخصصة حسب الطلب، و37 استشارة تبعثها أحياناً طلبات أخرى. ويهدف التدريب إلى تمكين الباحث من طرح الأسئلة المناسبة وتقييم الحلول واختيارها. ومن الأفضل أن ينقل النقاش إلى مجموعته ومجموعته ليتكون توافق أوسع على الممارسات الجيدة التي لا تزال متاحة في بعض المجالات فقط.

المرحلة الثانية: الإدارة النشطة:

تشمل تقريبًا كل ما يحتاج إلى الدعم أثناء المشروع، حين تُستخدم البيانات وتُعالج باستمرار، وتقودها خدمات تقنية المعلومات العلمية، التي تدعم الأنشطة كثيفة الحوسبة والبيانات، ومنها تشغيل عناقيد الحوسبة عالية الأداء والبيانات الضخمة، ودعم برامج المجموعات وتطوير برامج لحسابها، وخطوط المعالجة والتحليل المتخصصة، وتمثيل البيانات الضخمة بصريًا، والتدريب في هذه المجالات.

وتُطوّر الجهة منصة (openBIS) المفتوحة لإدارة بيانات البحث الكمي النشطة وتدعمها. وهي تطبيق ويب للتخزين والوصف والنسخ الاحتياطي خلال المشروع. واستُخدمت أولًا لإدارة كميات كبيرة من بيانات علوم الحياة، مثل المجهر والفحص عالي المحتوى والبروتيوميات والأيض والتسلسل، والبيانات المشتقة من المعالجة والتحليل.

واتجه التطوير في السنوات الأخيرة إلى دفتر مختبر إلكتروني للمختبرات المتوسطة، لتوثيق التجارب ووصف البيانات وجرد المواد والإجراءات، وبفضل البنية العامة امتد الاستخدام إلى علوم البيئة والمواد والفيزياء والكيمياء.

ولكل عضو مجلد للمشروعات والتجارب في دفتر المختبر، وتُنظّم العينات والمواد والطرائق في الجرد وتُربط بوصف التجربة. ويمكن رفع البيانات عبر الويب أو من أجهزة القياس مباشرة، وتحليلها باستخدام دفاثر (Jupyter).

وتقدم خدمات التقنية المنصة للمجموعات، بما يشمل تشغيلها وصيانتها على البنية المركزية والتدريب والدعم. وتستخدمها أكثر من أربعين مجموعة في مجالات مختلفة وبعض المرافق وقت المقال. لكن لا إلزام باستخدام دفتر إلكتروني أو هذه المنصة، ولا يُتوقع تغير ذلك قريبًا.

ولبعض الاستخدامات، تقدم المكتبة المحرر والعارض البسيط (docuteam packer) لتنظيم الملفات ووصفها محليًا استعدادًا للأرشفة. وليس حل إدارة بيانات متكاملًا، بل وسيلة لتنظيم البيانات المحفوظة في ملفات مع معلومات سياقية على مستوى الملف والمجلد، وهو مفتوح المصدر، طورته شركة (Docuteam) أساسًا للأرشفات.

وطلبت المكتبة تحسينات تلائم البيانات البحثية وتتكامل مع أرشيف البيانات ونظام معلومات الأرشيف الجامعي. وتستخدمه مجموعات قليلة ذات احتياجات خاصة، ويستخدمه الأرشيف الجامعي لوصف صفحات الويب المحصودة وتسليمها إلى أرشيف البيانات.

المرحلة الثالثة: النشر والحفظ:

تقود المكتبة مشاركة البيانات ونشرها وحفظها. ولأن نشرها وحفظها يتطلبان معلومات سياقية، ينبغي توثيقها مبكرًا عند إنشائها؛ ولذلك تدعم المكتبة الباحث منذ البداية كما سبق.

وتعتمد الخدمات تقنيًا، إلى جانب (docuteam packer)، على «مجموعة البحوث» (Research Collection) وأرشيف بيانات المعهد، وتشغلها المكتبة على البنية التقنية المركزية.

أطلقت مجموعة البحوث منتصف 2017 مستودعًا للمخرجات باستخدام تهيئة مخصصة من (DSpace) المفتوح. وتجمع مستودع الوصول المفتوح والبليوغرافيا الجامعية، اللذين كانا منفصلين، وتفذي التقارير الأكاديمية وقوائم المنشورات في مواقع الباحثين، وأضافت دور مستودع البيانات.

ويمكن إيداع مجموعات البيانات مواد مكملة لمنشورات، أو وحدات مستقلة، فتعمل المنصة بوابة موحدة لمهام نشر متعددة. ويمكن النشر فورًا بترخيص المشاع الإبداعي، أو بعد حظر زمني، أو بوصول مقيد. ويخدم الأخير البيانات غير المناسبة للتوزيع الواسع، التي تُشارك انتقائيًا أو تحفظ بأمان للإجابة عن اعتراضات محتملة على بيانات أو مقالات منشورة.

وأدمجت مجموعة البحوث مؤخرًا مع (openBIS)، فيختار المستخدم البيانات من واجهته لإرسالها مباشرة. وما زالت الملفات الكبيرة تحديدًا، وينبغي تقليل نقلها. ومن غير المفيد عرض ملفات بمئات الغيغابايت للتنزيل المباشر بالمتصفح، فيلزم توفير سبل أخرى. وبنهاية 2019 ضمت مجموعة البحوث 544 كيان بيانات بحثية.

ويُصدّر محتواها كله إلى أرشيف البيانات للحفظ الطويل، لمدد محددة قدرها عشر أو خمس عشرة سنة أو للحفظ الدائم. ويستخدم الأرشيف برنامج (Rosetta) التجاري من إكس ليبريس، ويمثل العمود الفقري للحفظ الرقمي بالمعهد.

وُثِرسل بيانات (docuteam packer) حاليًا مباشرة إلى الأرشيف، لكن قد يُعاد توجيهها مستقبلًا إلى مجموعة البحوث لتقليل تعدد الإجراءات التي ينفذها المستخدم ومقدم الخدمة.

وتتكون الغالبية الكبرى من محتوى الأرشيف حتى الآن من الملفات الأصلية لمشروعات رقمنة المكتبة؛ فلا إلزام للباحثين بالإيداع في مجموعة البحوث، التي لم تتوافر أصلًا إلا منتصف 2017. كما تُعالج بيانات الرقمنة وتودع آليًا وبكميات كبيرة، بينما كان رفع البيانات البحثية يدويًا. ويتقدم التكامل مع (openBIS) لتجاوز ذلك.

وتقدم المكتبة كذلك استشارات النشر المفتوح وإدارة البيانات وخطتها وحفظها عند استخدام المستودع، وتحيل الباحث حسب الحاجة إلى التقنية العلمية أو نقل التقنية أو غيرهما. وتتاح الاستشارات أيضًا لمن لا يعتزم استخدام المستودع.

التواصل:

أثبت تقسيم الدورة إلى ثلاث مراحل فائدته في توضيح أن إدارة البيانات مهام متعددة. لكن تنوع الخدمات والأدوات والوحدات قد يُصعب اختيار المناسب. ولذلك أنشأت الجهتان موقعًا مشتركًا عام 2017، يقدم أساسيات وروابط إلى المختصين والخدمات، دون ادعاء الإرشاد الشامل. ويوجد أيضًا صندوق بريد مشترك.

خاتمة:

رغم قصر الخبرة بالإطار الكلي للخدمات، ظهرت نقاط ضعف وتطورات مطلوبة. فالمعهد ما زال يُعدّ سياسة مخصصة للبيانات، لا تضيق اختيار المناهج العلمية، وتصلح لجميع التخصصات، وتوضح الالتزامات وما يُتوقع من الأعضاء. ويجب أن يلتزم هو أيضًا بتوفير موارد الخدمات الحالية والمقبلة.

وعادت تكلفة التخزين إلى الواجهة؛ فكلية التيرا بايت لا تنخفض بسرعة نمو البيانات في بعض المجالات. ولا يوجد حل مستدام لهذه الفجوة، وقد تضرر المجتمعات إلى تقييد ما يُحفظ ومدته. وتساعد إدارة البيانات بتمييز البيانات المؤقتة عن الأساسية وفق معايير مجتمعية مقبولة، لبناء قرارات واعية بالحذف أو الاحتفاظ المؤقت أو الحفظ.

ولا يخدم الإطار جميع التخصصات بالدرجة نفسها، ومن المبالغة توقع ذلك قريباً؛ فم منصة (openBIS) تركز على العلوم الكمية. وتظل المجموعات حرة في اختيار الأدوات الأنسب، مع استحسان التأكد من دعمها مركزياً أو على مستوى القسم. ولا ينبغي توقع فرض نظام بعينه أو فرض استخدام نظام أصلاً.

وتتطور المتطلبات والممارسة والأدوات سريعاً، فتحتاج الخدمات إلى التكيف. وقد يفيد الربط المرن نسبياً بين الخدمات في الاستجابة من زوايا متعددة، لكن الزمن سيبيّن إن كانت له عيوب أكبر.

ولا تكفي خدمات المؤسسة وحدها. فقد شاركت الوحدات في مشروع وطني لإدارة دورة حياة البيانات (DLCM)، بتمويل (swissuniversities)، فسهّل التبادل بين مؤسسات التعليم العالي السويسرية، وتقدمت الحلول القائمة والجديدة نحو الاستدامة.

وفي 2018 مُوّل مشروعان لاحقان: (openRDM.swiss) لإنشاء خدمة وطنية للمؤسسات الأكاديمية مبنية على (openBIS)، و(DLCM2) لمستودع وطني ومنصة تدريب. وقدمت خدمات التقنية مؤخراً مقترحاً إلى مؤسسة البنية الشبكية الأوروبية (EGI) لتوسيع الخدمة أوروبياً.

المراجع:

- Barillari, C, et al. 2016. openBIS ELN-LIMS: an open-source database for academic laboratories. *Bioinformatics*, 32(4): 638-640. DOI: <https://doi.org/10.1093/bioinformatics/btv606>
- Data Life-Cycle Management. 2017. The DLCM Project. Geneva, Switzerland: DLCM. Available at <https://www.dlcm.ch/about-us/dlcm-project> [Last accessed 7 August 2020].
- Docuteam. 2018. Software - Our tools for digital archives. Baden-Dättwil, Switzerland: Docuteam GmbH. Available at <https://www.docuteam.ch/en/products/it-for-archives/software/> [Last accessed 7 August 2020].
- Duraspace. 2018. About DSpace. Beaverton (OR), United States of America: Duraspace. Available at <https://duraspace.org/dspace/about/> [Last accessed 7 August 2020].
- Eidgenössische Technische Hochschule Zürich. 2011. Richtlinien für Integrität in der Forschung - Guidelines for Research Integrity. Zurich, Switzerland: ETH Zürich. DOI: <https://doi.org/10.3929/ethz-b-000179298> [Last accessed 7 August 2020].
- Eidgenössische Technische Hochschule Zürich. 2017. openBIS ELN-LIMS - Features. Available at <https://openbis.ch/> [Last accessed 7 August 2020].
- Eidgenössische Technische Hochschule Zürich. 2018a. IT in Research - Your IT support for research. Zurich, Switzerland: ETH Zürich. Available at <https://ethz.ch/services/en/it-services/it-in-research.html> [Last accessed 7 August 2020].
- Eidgenössische Technische Hochschule Zürich. 2018b. Research Data at ETH Zurich. Zurich, Switzerland: ETH Zürich. Available at <https://www.ethz.ch/researchdata> [Last accessed 7 August 2020].

- ETH Library, Eidgenössische Technische Hochschule Zürich. 2017. Research Collection. Zurich, Switzerland: ETH Library. Available at <https://www.research-collection.ethz.ch/> [Last accessed 7 August 2020].
- Ex Libris Ltd. 2018. Rosetta - Preserve your digital assets for the future. Jerusalem, Israel: Ex Libris. Available at <https://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/> [Last accessed 7 August 2020].
- Kluyver, T, et al. 2016. Jupyter Notebooks - a publishing format for reproducible computational workflows. In: Loizides, F and Schmidt, B (eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas, 87-90. Amsterdam, Netherlands: IOS Press. DOI: <https://doi.org/10.3233/978-1-61499-649-1-87>
- Swiss National Science Foundation. 2017. Open Research Data. Bern, Switzerland: SNF. Available at http://www.snf.ch/en/theSNSF/research-policies/open_research_data/ [Last accessed 7 August 2020].