

# مستودع أكسفورد (ORA): أتمتة جمع النصوص والبيانات الوصفية

جيسون بارتريدج



## الملخص:

يعرض المقال تجربة مستودع أكسفورد (ORA) في أتمتة جمع المنشورات والنصوص الكاملة والبيانات الوصفية لتخفيف العبء الإداري ودعم الامتثال لسياسات الوصول المفتوح، ويشرح الربط مع (Jisc Publications Router) و (CORE)، وتبادل السجلات مع (Symplectic Elements)، واستخدام المعرفات لتيسير إسناد الأعمال إلى الباحثين. ويناقش مواءمة أنواع المواد والملفات والتواريخ، ومنع التكرار، وتحديد حالات الوصول والحظر، والموازنة بين النشر الآلي والمراجعة البشرية. كما يعرض نتائج التنفيذ وحدود التفطية وجودة البيانات والواجهات، وخطط توسيع مصادر الحصاد، ويؤكد أن نجاح الأتمتة يرتبط بتحسين البيانات والتواصل والسياسات، لا بالحل التقني وحده، مع السعي إلى مستودع أشمل وأكثر كفاءة واستدامة يدعم ظهور الإنتاج العلمي ومتطلبات المؤسسة والممولين المتغيرة.

**الكلمات المفتاحية:** المستودعات المؤسسية، أتمتة الإيداع، البيانات الوصفية، الوصول المفتوح، حصاد المعلومات

## تمهيد المؤلف:

أصبح الوصول المفتوح مكونًا مهمًا للنشر الأكاديمي، ويضطلع مستودع جامعة أكسفورد (ORA) بجمع الإنتاج العلمي وحفظه ونشره. لكن الوفاء بسياسات الممولين وإطار التميز البحثي (REF) يفرض عبئًا إداريًا على الباحثين وموظفي الدعم. ويعرض المقال تحول الممارسة نحو الوصول الذهبي، وإدماج الأتمتة لتخفيف العبء، والاتجاهات اللاحقة لتحسين الامتثال والإتاحة.

## مقدمة:

يجمع (ORA) نتائج أعضاء الجامعة ويحفظه ويتيحها [1]. ولتكوين صورة شاملة للنشر، يودع الباحثون أعمالهم أو يوفرون معلومات الربط والتقارير في نظام معلومات البحث (Symplectic Elements) [2]. ويجب خلال ذلك مراعاة سياسة الجامعة [3]، وشروط الممولين، وفترات حظر الناشرين، ومتطلبات (REF) [4].

## معالجة فجوات السياسة ومشاركة الباحثين:

انسجامًا مع (REF2021)[5] وشروط التمويل، طبقت الجامعة سياسة «التصرف عند القبول»، التي تدعو إلى إيداع المخطوطات المقبولة للمقالات وأوراق المؤتمرات فور قبولها. لكنها، رغم دعم الامتثال وظهور البحث، لا تستوعب تمامًا النشر الذهبي[6]. ويستدعي تغير السياسات، ومنها مشاورات (REF2029)[7] وازدياد اتفاقيات التحول أو القراءة والنشر، تعديل الممارسة والرسائل الموجهة للباحثين.

تهدف أكسفورد إلى تيسير الإجراءات وتقليل العبء، مع الامتثال وتعظيم انتشار النتاج. ولذلك أدخلت جمعًا آليًا للمحتوى والبيانات عبر (Jisc Publications Router)[8] و (CORE)[9]. ويعرض المقال عمليات الربط ونجاحها وتحدياتها وأهداف التوسع إلى مصادر أخرى، لإدارة شاملة وفعالة.

## الوضع الراهن للوصول المفتوح:

يتحقق جزء كبير من الامتثال الآن بالنشر الذهبي، حيث يتيح الناشر العمل فورًا، غالبًا مقابل رسوم معالجة المقال (APCs)[10]. وفي المقابل يتراجع المسار الأخضر، أي الأرشفة الذاتية للمخطوط المقبول في (ORA)[11]. وأظهرت تقارير داخلية أن نسبة المواد الداخلة في نطاق سياسة التصرف عند القبول والممتثلة لها لا تتجاوز 31% [12].

ويتأثر تفضيل الذهبي بسياسات (Wellcome Trust)[13] و (UKRI)[14] ومنحهما التمويلية للجامعة، وبانتشار اتفاقيات القراءة والنشر، التي تقلل الحاجة إلى رسوم فردية وتوسع تغطية الباحثين، بما في ذلك غير المستفيدين من المنح المجمعة لدفع الرسوم.

ورغم الحث على الإيداع في المسارين، يبقى المعدل أقل من المرغوب، وتظهر فجوات في تتبع الامتثال وأهلية تقديم الأعمال إلى (REF). ويرى باحثون أن إيداع مقال منشور ذهبيًا عبء غير ضروري. ويبرز ذلك عند إعداد التقارير الحاجة إلى التقاط النص والبيانات آليًا بكفاءة.

## حلول آلية للامتثال: (Jisc Publications Router):

طورت أكسفورد الربط مع (ORA) بالتعاون مع (Cottage Labs)[15]. وتوفر الخدمة إشعارات تلقائية للمقالات ذات الانتماء إلى الجامعة، وتتيح حصاد بياناتها ونصوصها.

وتستقي، وقت الأصل، من 25 مزودًا [16]، بينهم ناشرون و (Crossref)[17] و (PubMed)[18] و (Europe PubMed Central)[19]. ويكاد المحتوى يقتصر على النشر الذهبي، لكن مقارنة قائمة المزودين بأماكن نشر باحثي أكسفورد توحى بتغطية نسبة مرتفعة من إنتاج الجامعة المفتوح.

وتقدر بيانات (Elements) لسنتي 2021 و2022 أن الناشرين المشمولين يمثلون 73% من المقالات التي سُجلت ناشرها، وأن 13 من أكبر 20 ناشرًا لأكسفورد يزودون الخدمة بالمحتوى. ويتيح ذلك حصادًا آليًا كبيرًا دون إيداع يدوي، بشرط إتاحة المزودين للمحتوى فعليًا.

## الربط مع (ORA) و (Elements):

لم يكن الربط جاهزًا للتشغيل بمجرد التوصيل. فقد بُني (ORA) على (Fedora)[20] لإدارة المحتوى، و (Hyrax)[21] ضمن (Samvera) للواجهة، ولم يوجد رابط سابق مع الخدمة؛ فطُور باستخدام (API) ونقاط (SWORD)[22]. وكان الفريق يعرف النمط من ربط (Elements) عبر (Repository Tools2) أو (RT2).

اختير الربط بـ (ORA) لا (Elements)، لغياب تغذية مباشرة من (Jisc) إلى (Elements). وبفضل (RT2) تُحدد السجلات الجديدة أو المحدثة من (ORA) إلى (Elements)، فتظهر إلى جانب سجلات المصادر الأخرى ليؤكد الباحث نسبتها إليه.

وإذا فُقد الباحث الإسناد التلقائي بمعرّف مثل (ORCID) [23] أو البريد، ووُجد في البيانات الواردة، أمكن إلحاق العمل بحسابه دون إبداع أو إجراء إضافي، فتتيسر التقارير والامتنال.

### تحسين جمع البيانات والنص الكامل:

استغرقت مواءمة بيانات (Jisc) مع نموذج (ORA) [24] وقتًا طويلاً، لتجميع القيم العديدة في قيم ذات معنى للعرض. فمثلاً كشف تحليل 60500 إشعار نحو 230 قيمة مختلفة لنوع المادة، احتاجت إلى مواءمتها مع أنواع (ORA) الرئيسية والفرعية.

ولزم افتراض بعض المعلومات الناقصة أو الفاقصة. فعند وصول ملفات متعددة، لا تبين بيانات الملف الثنائي هل هو نسخة السجل النهائية أم ملف تكميلي. لذلك استُخدمت كلمات في أسماء الملفات مثل (VoR) و (Supp) و (Article) لاستنتاج النسخة وإضافة قيم وفق ملف (RIOXX) [25].

وعدّل الاستيعاب ليتحقق من السجلات القائمة، ويتجنب التكرار، ويحدد حالة الوصول وفترات الحظر [26] لما هو محفوظ أو وارد.

ولم يقتصر الهدف على إنشاء سجلات جديدة، بل شمل تحديث القائم بتاريخ النشر والنسخ النهائية، لزيادة الاكتمال والدقة وتقليل تحديثات الموظفين اليدوية.

وعند وجود سجل، يُفحص اكتماله وإمكان تحديثه، كإضافة نسخة ملف أو تعديل البيانات، أو حاجته إلى مراجع بشري. وإذا استوفى الكائن الجديد أو التحديث معايير محددة، يُنشر تلقائياً في موقع (ORA) بعد الاستيعاب.

### ما تحقق حتى الآن:

البيانات الواردة غنية وتسمح بسجلات كاملة. ودخل الربط الخدمة منتصف يونيو 2024، وأضيف حتى كتابة الأصل 2710 منشوراً، منها 1852 استُوعبت ونُشرت آلياً دون تدخل موظف، مع مثال في المرجع [27]. أما البقية فاستُوعبت وتنتظر المراجعة قبل النشر.

ومع نمو الثقة وإغناء البيانات، يُتوقع استيفاء مزيد من السجلات شروط النشر الآلي.

### (CORE) و (ORA):

يتيح الربط بـ (CORE) جمع معلومات المقالات تلقائياً وتبسيط المراجعة والتنظيم. وهي خدمة غير ربحية تجمع من مستودعات ودوريات حول العالم. ويستفيد (ORA) من واجهتها [28] لتحسين عمله بشأن نتاج أكسفورد. ويتناول المرجع [29] العمل بمزيد من التفصيل.

### تحسين إجراءات المراجعة:

تقدم (CORE API) وصولاً إلى ملايين الوثائق من مصادر مثل (Crossref) و (DataCite)[30] و (BASE)[31] و (PubMed Central) و (arXiv)[32]، وتوحد استرجاع المعلومات، ويمكن جلب العنوان والمؤلفين والتاريخ و (DOI) [33] ورابط التنزيل.

ويستخدم (ORA) مهمة (Rake)[34] في بيئة (Ruby) لتحديد سجلات قابلة للتحديث، بالتركيز على حالات مراجعة محددة، وتُضاف معلومات النشر والبيانات ثم تُعلّم السجلات ليتحقق الموظفون من دقتها واكتمالها.

ويستجيب الربط لسياسات متغيرة، مثل (Plan S) وشروط (Wellcome) و (UKRI)، التي تتطلب معرفة تحديثات النشر لضمان الامتثال.

وبدأت أكسفورد في فبراير 2023 تجربة تتيح الاشتراك الاختياري في سياسة الاحتفاظ بالحقوق. ومن 14 أكتوبر 2024 أُدرجت في سياسة الوصول المفتوح[35]، وأصبح على المؤلف الانسحاب إن أراد استثناء منشور بعينه. ويدعم ذلك الأرشفة الذاتية، ويزيد أهمية الإتاحة الدقيقة وفي موعدها.

وتقلل أتمتة الحقوق والتببيه المبكر جهد تتبع النشر والبحث اليدوي عن معلومات ناقصة أو انتهاء حظر، فتيسر المراجعة والامتثال.

### جودة البيانات وحدود الواجهة:

تجمع (CORE) في مراحل مختلفة من دورة النشر، من التقديم والقبول إلى النشر. لذلك يلزم التأكد من دقة البيانات واكتمالها. وقد يرسل المزود تاريخاً افتراضياً أو خاطئاً، مثل أول يوم من الشهر أو السنة. ووضع (ORA) معايير تعطي الأولوية للتاريخ الأكثر تفصيلاً ودقة. ورغم التحديات، يدعم الربط الإتاحة في الوقت المناسب والامتثال والاحتفاظ بالحقوق.

وتوجد حدود تقنية ومحتوائية للاستعلام. يتعلق التقني بترويصات (HTTP) وعدد رموز البحث المسموح، 150 كل خمس دقائق كما يذكر الأصل؛ وعولج أساساً بإضافة انتظار بين النداءات. أما المحتوى فيتأثر بعدد المزودين وتواتر التحديث والحصاد. وكشف الاختبار عدم وجود بعض (DOI) الموجودة في (ORA) داخل (CORE) حتى ذلك الحين.

وتوضح (CORE) أنها تسعى إلى التغطية الكاملة بالمزامنة مع (Crossref) ومقارنة معرفات المستودعات، لكنها لم تبلغها. ولم يشكل أي من هذه التحديات عائقاً كبيراً أو متعذر التجاوز أمام استمرار الربط.

وعملياً تحقق مرحلتان أفضل توازن: تحدد الواجهة محتوى جديداً محتملاً، ثم يتحقق موظف من صحة المطابقة وجودة البيانات. ويخفض ذلك العمل اليدوي السابق لتحديد التحديثات.

### اتجاهات مستقبلية مع (CORE):

يستكشف (ORA) تحديد النصوص الكاملة وجمعها آلياً من المستودعات والمزودين، أو توفير رابط للموظف يجمع منه نسخة مسموحة، كالمخطوط المقبول (AAM). ويزيد ذلك شمول المستودع وفائدته دون جهد إضافي من الباحثين.

وتبرز التجربة أهمية جودة البيانات والتحسين المستمر للمشاركة والحصاد، لبنية وصول أكثر كفاءة وإتاحة. وتشارك أكسفورد بوصفها عضواً داعماً في (CORE)[36]، وتقدم محتوى مستودعها وفق (RIOXX) لتوسيع الوصول إلى نتائجها.

## توسيع الحصاد الآلي:

أشارت الأقسام الجامعية وجهات الحوكمة إلى الحاجة لمزيد من الأتمتة، مستهدفة توزيعًا 80/20 بين الإجراءات الآلية والجهد اليدوي في الإيداع والتقارير.

لكن الروابط الحالية محدودة؛ فمزودون كثيرون لا يضمنون صراحة المحتوى الناتج من اتفاقيات القراءة والنشر، ويقدمون الذهبي الخالص أو الذهبي في دوريات هجينة فقط.

ولذلك حُطت لخطط لمصادر بيانات إضافية وآليات لجمع النصوص من المستودعات والمجمعات. ويجري استكشاف (OpenAlex)[37] الذي يضم معلومات من (Internet Archive Scholar)[38] و (Unpaywall)[39]، وتوسيع الإسناد الآلي في (Elements) بـ (ORCID) والبريد ومعرّف مؤلف (Scopus). ويرفع ذلك دقة السجلات وشمولها ويقلل الإيداع اليدوي.

## مراجعة التواصل والخاتمة:

لا تنسج سياسة «التصرف عند القبول» تمامًا مع متطلبات الإتاحة الفورية الأحدث للممولين. ومع تراجع تفاعل الباحثين، تلزم مراجعة السياسة والرسائل. ويزيد عدم اليقين بشأن منح التمويل بعد 2024 لدعم اتفاقيات التحول[40] تعقيد المشهد، والحاجة إلى توازن مستدام بين المسارين الأخضر والذهبي.

وحُطت لتقييم شامل تقوده جهة خارجية، يأخذ ملاحظات الكليات والأقسام وهيئات الحوكمة. والهدف سياسة تعكس واقع النشر وتدعم الاستدامة والامتثال.

وتشير الملاحظات إلى أن تقليل العبء الإداري يزيد تقبل الباحثين للسياسات المحدثة. والإيداع الآلي مهم، وكذلك سياسات تكامل المعرفات، كإلزام (ORCID)، لتيسير الإسناد في (Elements).

يمثل التكامل مع (Jisc) و (CORE) تقدمًا مهمًا. ومع مراجعة السياسة وتطوير الخدمة، يدعم استدامة الوصول وظهور نتاج أكسفورد وأثره، والوفاء بمتطلبات المؤسسات والممولين المتغيرة.

## المراجع:

قائمة الاختصارات المهنية: <http://www.uksg.org/publications#aa>

1. "The Oxford University Research Archive (ORA)," University of Oxford, <https://ora.ox.ac.uk/> (accessed 14 January 2025).
2. "Symplectic Elements," Digital Science, <https://www.symplectic.co.uk/theelementsplatform/> (accessed 14 January 2025).
3. "Oxford University open access publications policy: summary," University of Oxford, <https://openaccess.ox.ac.uk/oapp>, (accessed 14 January 2025).
4. "Home," Research Excellence Framework (REF), <https://www.ref.ac.uk/> (accessed 14 January 2025).

5. "Home," Research Excellence Framework (REF) 2021, <https://2021.ref.ac.uk/> (accessed 14 January 2025).
6. "Gold open access definition," Open Access Network, <https://open-access.network/en/information/glossary#c6220> (accessed 14 January 2025).
7. "REF 2029 Open Access Policy consultation," REF 2029, <https://www.ref.ac.uk/guidance/ref-2029-open-access-policy-consultation/> (accessed 14 January 2025).
8. "About," Jisc Publications Router, <https://pubrouter.jisc.ac.uk/> (accessed 14 January 2025).
9. "Home," CORE, <https://core.ac.uk/> (accessed 14 January 2025).
10. "Article Processing Charge (APC) definition," Open Access Network, <https://open-access.network/en/information/glossary#c6188> (accessed 14 January 2025).
11. "Green Open Access definition," Open Access Network, <https://open-access.network/en/information/glossary#c6221> (accessed 14 January 2025).
12. Open Access Report, 6 September 2024.
13. "Open access policy," Wellcome Trust, <https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy> (accessed 17 January 2025).
14. "Open access policy," UK Research and Innovation (UKRI), <https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-article-open-access/> (17 January 2025).
15. "Home," Cottage Labs, <https://cottagelabs.com/> (accessed 14 January 2025).
16. "Current Content Providers," Jisc Publications Router, <https://pubrouter.jisc.ac.uk/about/providerlist/> (accessed 14 January 2025).
17. "Home," Crossref, <https://www.crossref.org/> (accessed 14 January 2025).
18. "Home," PubMed, <http://www.ncbi.nlm.nih.gov/pubmed> (accessed 14 January 2025).
19. "Home," Europe PubMed Central, <https://europepmc.org/> (accessed 27 July 2024).
20. "Home," Fedora, <https://fedora.lyrasis.org/> (accessed 17 January 2025).
21. "Home," Hyrax, <https://hyrax.samvera.org/> (accessed 14 January 2025).
22. "SWORD (Simple Web-service Offering Repository Deposit) (protocol)," Wikipedia, [https://en.wikipedia.org/wiki/SWORD\\_\(protocol\)](https://en.wikipedia.org/wiki/SWORD_(protocol)) (accessed 14 January 2025).
23. "Home," ORCID, <https://orcid.org/> (accessed 17 January 2025).
24. J. Partridge, and T. Wrobel, "ORA Data Model 2.2," University of Oxford, 2021, DOI: <https://doi.org/10.5287/bodleian:pr22x1bjE> (accessed 14 January 2025).

25. "About," RIOXX, <https://www.riox.net/about/> (accessed 14 January 2025).
26. "Embargo Period definition," Open Access Network, <https://open-access.network/en/information/glossary#c6209> (accessed 14 January 2025).
27. D.G. Jackson, "Lymphatic Trafficking of Immune Cells and Insights for Cancer Metastasis," *Clinical and Experimental Metastasis* 41, no. 4 (2023): 381-86, <https://ora.ox.ac.uk/objects/uuid:774e9e58-161c-4170-b7d7-a631e7979275> (accessed 14 January 2025).
28. "CORE Application Programming Interface (API)," CORE, <https://api.core.ac.uk/docs/v3> (accessed 14 January 2025).
29. J. Partridge, "CORE and the Oxford University Research Archive (ORA)," ORA, (2024), <https://ora.ox.ac.uk/objects/uuid:e3c457de-554c-4ad9-bcd1-c4cfb5ce90ac> (accessed 14 January 2025).
30. "Home," DataCite, <https://datacite.org/> (accessed 17 January 2025).
31. "Home," BASE Bielefeld Academic Search Engine, <https://www.base-search.net/> (accessed 14 January 2025).
32. "arXiv", Cornell University, <https://arxiv.org/> (accessed 14 January 2025).
33. "Digital Object Identifier (DOI) definition," Open Access Network, <https://open-access.network/en/information/glossary#c6205> (accessed 14 January 2025).
34. "What is Rake in Ruby & How to Use it," Ruby Guides, <https://www.rubyguides.com/2019/02/ruby-rake/> (accessed 14 January 2025).
35. "Rights Retention," University of Oxford, <https://openaccess.ox.ac.uk/rights-retention/> (accessed 14 January 2025).
36. "CORE Membership for data providers," CORE, <https://core.ac.uk/membership> (accessed 14 January 2025).
37. "Home," OpenAlex, <https://openalex.org/> (accessed 14 January 2025).
38. "Internet Archive Scholar," Internet Archive, 2020, <https://scholar.archive.org/> (accessed 14 January 2025).
39. "Home," Unpaywall, <https://unpaywall.org/> (accessed 14 January 2025).
40. "Working with transitional agreements," JISC, 2020, <https://www.jisc.ac.uk/guides/working-with-transitional-agreements> (accessed 14 January 2025).