

حين تكون بياناتك تسجيلًا لفناء أجدادي: رقمنة السجلات الثقافية وإتاحتها في أرشيف (PARADISEC)

نيك ثيرغر، أماندا هاريس



المخلص:

يعرض هذا المقال تجربة أرشيف (PARADISEC)، وهو مستودع بحثي متخصص يحفظ تسجيلات اللغات الصغيرة في العالم ونصوصها المدونة وأغانيها وحكاياتها، ويضم أكثر من 390000 ملف بحجم 150 تيرابايت تمثل 1312 لغة، كثير منها من بابوا غينيا الجديدة والمحيط الهادئ، ويبيّن المؤلفان أهمية المستودعات الخاصة بالتخصص للبحث القابل للتكرار وإدارة بيانات البحث، ويصفان مهمة رقمنة الأشرطة التناظرية المعرضة للتلف، ونظام الإيداع الآلي وضبط الجودة وشروط الوصول والمعرفات الرقمية والتحقق اليومي من سلامة الملفات، ويناقشان كلفة عدم أرشفة المواد البحثية، والتخطيط لانتهاء المشروعات، والمجموعة ذاتية الوصف التي تتيح نقل أجزاء منها إلى المراكز الثقافية، والانتقال إلى معياري (OCFL) و (RO-Crate)، ويخلصان إلى أن هذه البيانات تراث ثقافي لأصحابه وذريتهم، وأن الباحثين مسؤولون عن إعادتها إليهم، داعيين إلى مزيد من أرشيفات اللغات حول العالم.

الكلمات المفتاحية: أرشفة اللغات، الأرشيفات الموسيقية، اللغات المهددة، الرقمنة، إدارة بيانات البحث

مقدمة:

نناقش في هذه الورقة مستودعًا بحثيًا يهدف صراحةً إلى أن يكون قناةً توصل المخرجات البحثية إلى طائفة من الجماهير داخل الأوساط الأكاديمية وخارجها. وينصب تركيزنا على التسجيلات والنصوص المدونة بلغات العالم الصغيرة الكثيرة، أي على الأغاني والحكايات التي تمثل تعبيرات ثقافية فريدة. ومع أن هذه البيانات تُنشأ عادةً لمشروع بحثي بعينه، فإن لها قيمة هائلة تتجاوز البحث الأكاديمي، لأنها في الغالب تراث شفهي سُجِّل في أماكن لم يُسجَل فيها إلا القليل غيره.

وأرشفيات اللغات والموسيقى الرقمية ظاهرة حديثة (Barwick 2004)، تقدم أوصافاً لمثل هذه المواد بمصطلحات معيارية ثم تتولى إدارة هذه المواد ذات الأهمية الثقافية وتنظيمها (curate)، وكثيراً ما تكون هذه المواد حصيلة رقمنة تسجيلات تناظرية. وهذه التسجيلات التناظرية معرضة للخطر، ولم يعد من الممكن الاستماع إليها في المجتمعات التي جاءت منها لعدم توافر أجهزة التشغيل. وثمة شبكة دولية لأرشفيات اللغات تُسمى شبكة أرشفيات اللغات والموسيقى المهددة بالاندثار الرقمية (Digital Endangered Languages and Musics Archives Network, DELAMAN)، تتبادل المعلومات وتوفر مرجعاً لكل من يبحث عن مجموعات لغوية. ومشروعنا، أرشيف المحيط الهادئ والمنطقة للمصادر الرقمية في الثقافات المهددة بالاندثار (PARADISEC)، عضو في (DELAMAN)، ويعمل منذ 19 عامًا. وقد نما حتى صار يضم أكثر من 390000 ملف يبلغ حجمها حاليًا 150 تيرابايت وتمثل 1312 لغة، كثير منها من بابوا غينيا الجديدة والمحيط الهادئ. وكان التركيز الرئيس للمجموعة على الأشرطة الصوتية المسجلة منذ خمسينيات القرن العشرين، ويُكرّس كثير من جهدنا لتحديد أماكن الأشرطة التناظرية الفريدة ورقمنتها.

ومن المهم للممارسة البحثية أن تكون هناك مستودعات خاصة بالتخصصات. فالبيانات القابلة للاستشهاد جزء لا يتجزأ من البحث السليم (Berez-Kroeker et al. 2018)، إذ تتيح التحقق من الوقائع التي تُبنى عليها الادعاءات التحليلية، وتكرار التحليل (cf Himmelmann 1998). لكن البيانات في عملنا هي أيضًا مادة ثقافية ذات قيمة للأشخاص الذين سَجَلُوا ولذريتهم؛ إنهم أجدادهم، ولذلك تقع علينا، نحن الباحثين القادمين من الخارج، مسؤوليات خاصة في التعامل مع المواد باحترام وفي ضمان إتاحتها للأشخاص الذين عملنا معهم.

وفي العالم نحو 7000 لغة (Eberhard et al. 2022)، ولا توجد عن معظمها إلا سجلات قليلة. وكثيراً ما تكون أكثر السجلات تفصيلاً هي تلك التي أعدها باحثون أكاديميون - لغويون وعلماء موسيقى وعلماء أنثروبولوجيا - أمضوا وقتاً في دراسة الأداء بتلك اللغات. لكن هذه السجلات الفريدة معرضة للضياع إن لم يوجد مستودع رقمي لحفظها. وفي الوقت نفسه، ثمة تركيز متزايد في الأوساط الأكاديمية على البحث القابل للتكرار وإدارة بيانات البحث (Corti et al. 2014)، والمستودعات هي مفتاح الإدارة الناجحة للبيانات. وفي العلوم الإنسانية خصوصاً، يُرَجَّح أن تثير السجلات البحثية اهتمام جمهور أوسع مما تثيره معظم البيانات العلمية، ولذلك تشتد الحاجة إلى بنية تحتية لوصف هذه المواد وإدارتها وإتاحتها.

وكما سنبين، تستطيع المستودعات الخاصة بالتخصصات والمتوافقة مع المعايير السليمة أن تلبي احتياجات الباحثين، وأن توفر في الوقت نفسه تدفقات للبيانات الوصفية (OAI-PMH, RIF-CS) لنظم البيانات الوطنية، كما في حالتنا لمكتبة أستراليا الوطنية (National Library of Australia) أو لـ (Research Data Australia). ولعل الأهم من ذلك أن السجلات يمكن أن تُعاد إلى الأشخاص المسجلين فيها، وأن يحدث ذلك مراراً إلى أي مكان يتوافر فيه اتصال بالإنترنت. وتوفر الهواتف المحمولة على نحو متزايد الوصول إلى الإنترنت حتى في أكثر المناطق نائية، ولذلك نأمل أنه حين يبحث أحد عن معلومات بلغته، سيتمكن من العثور عليها في مجموعتنا. ونود التأكيد على أن هذا العمل يهدف إلى ردم الفجوة الرقمية بأكثر الطرق عملية، ببناء مستودع ذي نظام إيداع (ingestion) بسيط ومخطط بيانات وصفية مناسب، لجعل إضافة مواد جديدة، وترخيصها وإتاحتها، أيسر ما يمكن (Barwick and Thieberger 2018).

ويوفر هذا النوع من الأرشفيات نسخة ثابتة من بيانات البحث يمكن الاستشهاد بها، ومن ثم فهو مركز للنشاط، ومكان يمكن انطلاقاً منه التعليق على المواد البحثية وإعادة استخدامها بطرق جديدة، مع انعكاس المعرفة الجديدة على المجموعة بمرور الوقت. فالأرشيف، بعيداً عن أن يكون نقطة النهاية للبحث، يعيد إدخال هذه المواد

في علاقة مستمرة وحوارية مع الأشخاص المسجلين ومع الباحثين في المستقبل (Barwick 2004). ولولا جهد الأرشفة لظلت هذه المواد غير متاحة بعد انتهاء المشروع الذي أنتجها. وتجربتنا أن هذه المواد، لولا الأرشفة، ستكون على الأرجح غير متاحة أيضًا إلا لأكثر الباحثين تنظيمًا، إن لم تضع بسبب عطل في القرص الصلب أو الحاسوب.

ومن المهم التمييز بين مجموعة منظمة ومُدارة من الملفات الأرشيفية وبين ملفات على نظام تخزين. فطلبتنا من مديري النظم في الجامعة بشأن الأرشفة طويلة الأمد تُقَابَل عادةً باقتراحات لاستخدام نظم تخزين الملفات، وحتى في هذه الحالة لا يوجد عادةً ضمان للاستمرار بعد دورة التمويل الحالية. وتخزين الملفات خطوة أولى ضرورية، ويمكن أن يوفر النسخ الاحتياطي اللازم أثناء تأسيس المجموعة. غير أن تخزين الملفات وحده لا يوفر البيانات الوصفية لوصف المواد، ولا يطبق أي تراخيص تُعلم الآخرين بكيفية استخدام المواد. كما أن الأرشفة يفحص الملفات بانتظام، فيحسب المجاميع الاختبارية (checksums) لضمان سلامة البيانات. ولكل هذه الأسباب، لن تكون الملفات في نظام التخزين قابلة للعثور عليها، ولا للوصول إليها، ولا للتشغيل البيئي، ولا لإعادة الاستخدام، كما ينبغي أن تكون وفقًا لمبادئ (FAIR). [2] وفضلًا عن ذلك، نلتزم بمبادئ (CARE) [3] قدر الإمكان. ونذكر أن مسؤوليتنا هي ضمان سلامة التسجيلات التناظرية واستمرارها، وهي تسجيلات أبعدت عن التداول وظلت غير متاحة للأشخاص الذين سُجِّل أسلافهم فيها. ولأننا نتعامل مع طائفة كبيرة من اللغات (1312 لغة ممثلة في المجموعة حاليًا)، فإننا لا نستطيع استشارة المتحدثين، الذين كثيرًا ما يكونون في أماكن نائية. وعندئذ يكون علينا الاختيار بين إتاحة السجلات أو الانتظار حتى نحصل على الإذن. وقد رأينا مؤسسات تغلق الوصول في هذه الحالة، ونرى أن هذا النهج يفاقم المشكلة. ونحن نعمل بمبدأ الإزالة عند الطلب (takedown principle)، فننصح مستخدمي الفهرس بالتواصل معنا إن رأوا أن مادة ما لا ينبغي إتاحتها، ونشجع المتحدثين الذين تربطهم صلة بتسجيل ما على التواصل معنا.

ماذا يفعل (PARADISEC):

(PARADISEC) أرشفة يستكشف إمكانات جديدة مثيرة للمواد البحثية. ومع أن الدافع الأول كان رقمنة الأشرطة التناظرية من مرحلة سابقة من البحث (منذ الخمسينيات)، فإننا بعد أن بنينا البنية التحتية اللازمة صرنا نتولى أيضًا إدارة الملفات المولودة رقميًا (born-digital) الناتجة عن العمل الميداني الجاري، وهي تمثل نحو نصف المجموعة الحالية.

وعلى أن نكون واضحين في أن هذا المشروع لا «ينقذ اللغات» ولا «ينقذ الموسيقى». فنحن ننقذ سجلات الأداء، ويساعد ذلك على إبراز تنوع اللغات والأداء الموجود في العالم. كما يمنح حضورًا لأصوات تكون عادةً مهمشة ومستبعدة من الإنترنت. وفيما يمكن عدّه ممارسة تطبيقية لما بعد الاستعمار، نهدف إلى إتاحة هذه السجلات مجددًا للأشخاص الذين جاءت منهم (Thieberger 2020).

وقد بنينا نظامًا يؤتمت معظم عمليات إيداع الملفات، وضمان الجودة، وإدارة المستخدمين، والوصول، لمجموعات المواد البحثية، ولا سيما تسجيلات الوسائط (media recordings) والنصوص المدونة والمواد المرتبطة بالعمل الميداني اللغوي أو الموسيقي. ونقدم على صفحتنا الإلكترونية نصائح بشأن إدارة البيانات وتسمية الملفات، ونعقد جلسات تدريبية منتظمة للتشجيع على التفكير في الأرشفة منذ بداية العمل الميداني. كما نعقد ورش عمل تدريبية على استخدام أدوات مناسبة يمكن أرشفة مخرجاتها ولا تكون حبيسة صيغ مملوكة. وبمجرد دخول الملفات إلى المجموعة، نسند إليها معرفات الكائنات الرقمية (DOI)، ويفرض نظامنا شروط الوصول. ويقبل كل مستخدم مسجل في الفهرس مجموعة من الشروط، [4] ويحدد كل مودع كيفية استخدام مواده. وحتى إذا أُعطيت مادة ما حالة

«مغلقة»، أي لا يستطيع أحد الوصول إلى ملفاتها، يستطيع المودع أن يمنح حقوقاً فردية لمستخدمين مسجلين آخرين لاستخدام تلك المادة. كما يتيح حالة «خاصة» أثناء بناء المجموعة، تحجب حتى البيانات الوصفية عن العرض العام، ولا يُسند أي (DOI) حتى تنتهي تلك الحالة الخاصة. أما المواد المدرجة بوصفها «مفتوحة»، فيستطيع المستخدم المسجل تنزيل ملفاتها.

ويُعالج كل ملف يدخل المجموعة بحسب نوعه، فُتُرفض أنواع الملفات غير المسموح بها فوراً ويُولد تقرير بالبريد الإلكتروني. وتعكس صيغة تسمية ملفاتنا التسلسل الهرمي: معرف المجموعة-معرف المادة-اسم الملف. الامتداد (collection id-item id-filename.extension) (مثل NT1-98007-01.wav). وبالنسبة إلى أنواع الملفات المسموح بها، تبدأ سلسلة معالجة، فمثلاً يُغلف ملف (wav) بمجموعة من مصطلحات البيانات الوصفية المنسوخة من الفهرس في ملف (xml) لإنتاج ملف بصيغة (broadcast wav format, BWF)، وتُنشأ نسخة (mp3) للتقديم عبر الإنترنت. ويحوّل ملف (tif) إلى (jpg) للتقديم. ونعالج ملفات الفيديو خارج مسار المعالجة المباشرة (offline) لأنها تتطلب قدرًا كبيرًا من قدرة المعالج، ونتج ملف (mxf) أرشيفيًا وملف (mp4) للتقديم. وتُرسل النسخة الأرشيفية للملف ومشتقه المضغوط إلى الدليل المناسب في الأرشيف بناءً على اسم الملف.

وتُصدر تقريرًا يوميًا عن حالة المجاميع الاختبارية للملفات تحسبًا لحدوث أي تغيير. ويُنسخ فهرسنا احتياطيًا يوميًا إلى موقع آخر، وكذلك المجموعة، ونجري دوريًا تمرينًا للتعافي من الكوارث نسترجع فيه ملفات من النسخة الاحتياطية للتأكد من سلامتها.

وفي عام 2019 حصلنا على ختم البيانات من (World Data System) [5]، وهو ما يدل على أننا نلتزم بجميع المعايير المطلوبة. وفي عام 2013 أُدرجت مجموعات (PARADISEC) في سجل ذاكرة العالم الأسترالي لليونسكو (UNESCO Australian Memory of the World) [6].

مهمة عاجلة:

سُجّلت التسجيلات الفريدة التي ركزنا عليها في (PARADISEC) على أشرطة تناظرية حتى التسعينيات، حين صارت التسجيلات الرقمية هي القاعدة. وللشريط التناظري، سواء على البكرات أو في الكاسيت، مشكلة كبرى، هي أنه يُرجّح أن يصبح غير قابل للتشغيل في السنوات القليلة المقبلة. [7] وعدم توافر أجهزة تشغيل هذه الأشرطة عامل من عوامل تعذر الوصول إليها، ولا سيما البكرات المفتوحة، وكذلك الكاسيت. والأخطر من ذلك أن الأشرطة نفسها ستبدأ في التفكك بعد أن تبلغ نهاية عمرها. والمفارقة أنه مع هشاشة الشريط التناظري، فإننا نعلم أن السجلات الرقمية أشد هشاشة، ومع ذلك فالرقمنة هي حاليًا الوسيلة الموصى بها لحفظ الصوت التناظري. [8] فالأشرطة التناظرية المسجلة في الخمسينيات لا تزال قابلة للتشغيل، لكن لعلنا جميعًا مررنا بتجربة العجز عن فتح ملفات رقمية أنشئت قبل عشر سنوات فقط، بسبب تغير الصيغ والبرمجيات. ومن الحلول الجزئية ضمان تحويل جميع الملفات إلى صيغة نعلم أن فرص بقائها أكبر، وهنا تتبع ببساطة المعايير الراسخة. ولذلك نُؤرشف بصيغ (wav) و (txt) و (xml) و (tif)، لكننا نخزن أيضًا نسخًا أقل دقة بصيغ (mp3) أو (pdf) أو (jpg) للتقديم. ونُعد نسخًا احتياطية يومية في مواقع مادية مختلفة.

كلفة عدم أرشفة المواد البحثية:

يمكن حساب قيمة بيانات البحث إما بوصفها كلفة إنشائها، وإما بوصفها قيمتها للأجيال اللاحقة مجموعة متاحة من المواد. وبالنسبة إلى نوع المواد الذي ناقشه هنا، فإن القيمة حائلة كلفة تنفيذ المشروعات البحثية، ومقدار الجهد المبذول في التسجيلات وفي شرحها (annotation) وتدوينها. وفي العمل الميداني في أماكن بعيدة عن مقر الباحث، ثمة تكاليف السفر، والوقت اللازم لبناء العلاقات مع المؤدين والمتحدثين والمجتمعات التي يعيشون فيها. وثمة قيمة ثقافية للتسجيلات الوحيدة بلغة معينة يجب أيضًا أخذها في الحسبان عند النظر في كيفية تقدير التكاليف. والتسجيلات المُنجزَة ثمرة علاقات تعاونية قائمة على الثقة نُسجت بمرور الوقت. وضمان إمكانية الوصول مستقبلاً إلى السجلات المنشأة في هذا السياق جزء من العملية البحثية، وتعترف به جهات التمويل على نحو متزايد بوصفه جزءاً لا يتجزأ من الممارسة الجيدة، إذ تتوقع من الحاصلين على المنح اتباع خطط إدارة البيانات. وكل ذلك معرض للخطر إن لم يتخذ ما يلزم للعناية بالسجلات على المدى الطويل.

التخطيط لانتهاء المشروع:

إن ممارسة إعداد مجموعة للأرشفة في (PARADISEC) تمثل في حد ذاتها تقدماً كبيراً مقارنة بالتجميع المعتاد للمواد الذي يحتفظ به الباحث عادةً. فالباحثون لا يجرّدون ما لديهم ويحددون ما يملكونه، وما عاجوه (مثل ما دونوه أو شرحوه)، وما تبقى من عمل، إلا حين يواجهون ضرورة وصف الملفات، مهما كان الوصف موجزاً، لإيداعها في المجموعة. ويلبي الأرشيف احتياجات اختتام المشروع البحثي، لكنه يوفر أيضًا أساساً للتفاعل المستمر مع البيانات. وينشئ الأرشيف معرفات دائمة لهذه المواد، بحيث يمكن الاستشهاد بها في الأمثلة في المنشورات، ويستطيع آخرون الوصول إليها للتحقق من الادعاءات المبنية على هذه الأمثلة. وفضلاً عن ذلك، تصبح المواد الآن متاحة للآخرين لاستخدامها، لأن عملية الإيداع تتضمن تحديد نوع شروط الوصول التي ستنطبق. وبهذه الطريقة تترسخ منهجية بحثية أمتن، ولأن السجلات التي تُعنى بها ذات قيمة تراثية، فإن من الفوائد الإضافية قدرة المتحدثين بهذه اللغات على الوصول إلى التسجيلات.

وقد نال (PARADISEC) إشادات كثيرة، لكننا ندرك أنه يعمل ضمن بيئة للبنية التحتية البحثية، على مستوى الجامعات وعلى المستوى الوطني معاً، لا توفر الإدارة طويلة الأمد للمخرجات البحثية. وقد كنا رواداً في نظام لوصف أنواع معينة من بيانات البحث وإدارتها، ونعلم أن أساليبنا يمكن توسيعها لتشمل أنواعاً أخرى من البيانات. وكنا قد رأينا أمثلة لنظم قُصّلت فيها البيانات عن أوصافها الوصفية، فكانت البيانات الوصفية في قاعدة بيانات والملفات في المجموعة مخزنة في مكان آخر، ولم نرد استخدام هذا النهج في مجموعتنا.

ولدينا بالفعل قاعدة بيانات تضبط إدخال البيانات الوصفية، وتتحقق من مطابقة المواد الواردة لمعاييرنا (مثل تسمية الملفات، ومصطلحات البيانات الوصفية، وأنواع الملفات، والمعرفات الفريدة، وإسناد معرفات الكائنات الرقمية). وقد دأبنا على كتابة وصف كامل لكل مادة في الدليل الذي يحوي البيانات، فهي مجموعة ذاتية الوصف، مستقلة عن برمجيات الفهرسة، ويمكن إعادة بناء فهرس المجموعة كلها من المجموعة نفسها. وتخزن كل مادة البيانات الوصفية على مستوى المجموعة وعلى مستوى المادة، وتُحدّث في كل مرة يُعدّل فيها مدخل الفهرس ويحفظ.

ولأن كل دليل في المجموعة ذاتي الوصف، نستطيع إنشاء أي مجموعات فرعية نختارها من المجموعة من دون أن نفقد البيانات الوصفية (المعلومات السياقية). ومن مهامنا المعتادة تحميل الملفات على قرص صلب لإتاحة مجموعات من السجلات للمراكز الثقافية أو المتاحف في المحيط الهادئ، ومجرد وضع الملفات على قرص صلب

يتيحها فعليًا، لكن التنقل فيها صعب من دون فهرس. وقد كتبنا تطبيقًا [9] يفحص مجموعة الملفات هذه على القرص الصلب ويكتب فهرسًا بصيغة (html) لتلك الملفات وحدها لإنشاء عارض محلي. والخدمات نفسها التي نقدمها في الفهرس الإلكتروني (مشغلات الوسائط، وعارضات الصور، وغيرها) متاحة أيضًا في هذا العارض المحلي. وننسخ حزم المجموعات الفرعية مع فهرسها إلى أجهزة (Raspberry Pi) تتيحها عبر شبكة (Wi-Fi) صغيرة، مما يتيح قراءة الفهرس والملفات على جهاز محمول.

ولأن نظامنا الحالي قد تقادم، فإننا ننتقل إلى استخدام (Oxford Common File Layout, OCFL) [10] لتخزين الملفات، و (Research Object Crate, RO-Crate) [11] لتقديم النوع نفسه من الوصف الذي كان في ملفات (XML) غير المعيارية لدينا، ولكن الآن بصيغة متوافقة مع المعايير. وكان نظامنا السابق يتطلب معرفة متخصصة بـ (Ruby on Rails)، ووجدنا صعوبة في استقطاب أشخاص يملكون المهارات اللازمة لصيانة ذلك النظام. أما (OCFL/ RO-Crate) فيستخدمان (JSON)، وهي تقنية شائعة الاستخدام، ومن ثم ينبغي أن يكونا أكثر استدامة في مرحلة التطوير التالية. وقد بنينا نموذجًا تجريبيًا (demonstrator) باستخدام هذه المعايير. [12] ونظرًا إلى أننا نحفظ بـ 150 تيرابايت، فمن غير المجدي إنشاء نسخة تجريبية معزولة (sandbox) للمجموعة كاملة، ولذلك يضم النموذج التجريبي الفهرس كاملًا ونحو ثلاثة أرباع الملفات. وينبغي أن تكون النسخة الجديدة جاهزة في أوائل عام 2023.

خاتمة:

عرضنا مشروعًا جاريًا يبني دورة عمل تمتد من إنشاء البيانات الأولية، مرورًا باستخدامها الأولي، وتسمية الملفات واختيار صيغ البيانات (لتحقيق أطول عمر ممكن للملفات)، وصولًا إلى الإيداع في مستودع. ويحدد المودع شروط الإيداع، ويخضع الوصول لمجموعة أخرى من الشروط يوافق عليها المستخدمون المسجلون. ومع نظم التحقق من البيانات الأخرى التي ناقشناها هنا، يقدم (PARADISEC) نموذجًا لإدارة بيانات البحث بطريقة مسؤولة تستجيب أيضًا لاهتمام المجتمع الأوسع بهذه المواد، وتقر بأن ما يُعد بيانات صوتية للبحث هو أيضًا تسجيل لفناء أجداد أحدهم.

ومع أن هذا المثال خاص بسياقنا وتجربتنا، فإننا نقدمه هنا نموذجًا لما يمكن إنجازه، مدركين الحاجة إلى عدد أكبر بكثير من أرشيفات اللغات حول العالم، تكون قادرة على التعامل مع أوضاع اللغات المحلية، وأقرب إلى أكثر الناس اهتمامًا بالمجموعات، أي المتحدثين وذريتهم.

إسهامات المؤلفين:

شارك كل من المؤلفين بالتساوي في العمل الموصوف في هذا المقال وفي كتابته.

المراجع:

1. DELAMAN. <https://www.delaman.org/>.
2. www.go-fair.org.
3. <https://www.gida-global.org/care>.
4. <https://www.paradisec.org.au/deposit/access-conditions/>.
5. <https://www.icsu-wds.org>.

6. <http://www.amw.org.au/register/listings/pacific-and-regional-archive-digital-sources-endangered-cultures-paradisec>.
7. NFSA (National Film and Sound Archive of Australia), 2014. Deadline 2025: collections at risk. <https://www.nfsa.gov.au/collection/curated/deadline-2025-0>.
8. International Association of Sound and Audio-visual Archives, Guidelines on the Production and Preservation of Digital Audio Objects, <https://www.iasa-web.org/tc04/audio-preservation>.
9. <https://language-archives.services/about/data-loader>.
10. <https://ocfl.io>.
11. <http://www.researchobject.org/ro-crate>.
12. <https://mod.paradisec.org.au>.
13. Barwick, L. 2004. Turning it all upside down... Imagining a distributed digital audiovisual archive. *Literary and Linguistic Computing*, 19(3): 253-263. DOI: <https://doi.org/10.1093/lc/19.3.253>
14. Barwick, L and Thieberger, N. 2018. Unlocking the archives. In Ferreira, V and Ostler, N (eds.), *Communities in Control: Learning tools and strategies for multilingual endangered language communities*. Proceedings of the 2017 XXI FEL conference, 135-139. <http://hdl.handle.net/11343/220007>.
15. Berez-Kroeker, AL, Gawne, L, Kelly, BF, Heston, T, Kung, S, Holton, G, Pulsifer, P, Beaver, D, Chelliah, S, Dubinsky, S, Meier, R, Thieberger, N, Rice, K and Woodbury, A. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, 56(1): 1-18. DOI: <https://doi.org/10.1515/ling-2017-0032>
16. Corti, L, Van den Eynden, V, Bishop, L and Woollard, M. 2014. *Managing and Sharing Research Data: A Guide to Good Practice* (Paperback). London: Sage Publications.
17. Eberhard, DM, Simons, GF and Fennig, CD. (eds.) 2022. *Ethnologue: Languages of the World*. Twenty-fifth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
18. Himmelmann, N. 1998. Documentary and descriptive linguistics. *Linguistics*, 36(1): 161-195. DOI: <https://doi.org/10.1515/ling.1998.36.1.161>
19. Thieberger, N. 2020. Technology in support of languages of the Pacific: neo-colonial or post-colonial? *Asian-European Music Research Journal*, 5-3: 17-24. DOI: <https://doi.org/10.30819/aemr.5-3>