

Web Archiving and Digital Heritage Preservation: Approaches and Challenges

[10.35781/1637-000-196-007](https://doi.org/10.35781/1637-000-196-007)

د. طارق الحبيب الورفلي*

*جامعة الشرقية – سلطنة عمان

عضو في مخبر البحث علم المعلومات بالمعهد العالي للتوثيق – جامعة منوبة – تونس

ملخص

الهائل والنمو المتسارع لصفحات الويب، مما يتطلب بنية تحتية متطورة، إضافة إلى تنوع المحتوى وتشعبه، مما يستوجب تنوع الأدوات والتقنيات لضمان إمكانية الوصول إليه وتشغيله لاحقاً بالطريقة نفسها التي كان عليها عند أرشفته. وتتلخص التحديات القانونية بمسألة حقوق التأليف والإيداع القانوني الرقمي.

الكلمات المفتاحية: أرشفة الويب، حفظ التراث الرقمي، التحديات التقنية، التحديات القانونية

تهدف هذه الدراسة إلى إبراز أهمية أرشفة مواقع الويب في حفظ التراث الرقمي. واعتمدت المنهج الوصفي التحليلي من خلال قراءة تحليلية لمجموعة من التجارب الدولية في أرشفة الويب وفق مقاربات مختلفة: المقاربة الشاملة والمقاربة الانتقائية والمقاربة التكاملية. وأفضت إلى التأكيد على دور أرشفة الويب في حفظ التراث الرقمي المنشور عبر شبكة الإنترنت. كما أبرزت التحديات المتعلقة بأرشفة الويب سواء أكانت تحديات تقنية أم قانونية. وتتمثل التحديات التقنية في طبيعة الويب الديناميكية والحجم

Web Archiving and Digital Heritage Preservation: Approaches and Challenges

Dr. Tarek Habib Ouerfelli

A'Sharqiyah University – Sultanate of Oman

This study aims to highlight the importance of web archiving in preserving digital heritage. It adopts a descriptive-analytical approach based on a review of a range of international experiences in web archiving, using three different approaches: the comprehensive approach, the selective approach, and the integrative approach. The study underscores the role of web archiving in preserving digital heritage published on the Internet. It also highlights the challenges associated with web archiving, whether technical or legal. Technical challenges stem from the

dynamic nature of the Web, the enormous volume of web pages, and their rapid growth, all of which require advanced infrastructure. In addition, the diversity and complexity of web content necessitate a variety of tools and technologies to ensure that it can be accessed and used later in the same manner as when it was archived. Legal challenges relate to copyright and digital legal deposit.

Keywords: Web archiving, digital heritage preservation, technical challenges, legal challenges

المقدمة

الإنترنت بطبيعته نظام شديد الديناميكية والزوال. فالإحصاءات تشير إلى أن متوسط عمر صفحة الويب العادية لا يتجاوز بضعة أشهر قبل أن يتم تعديلها أو حذفها. ومن هنا تبرز أهمية الأرشفة لحفظ الذاكرة التاريخية والثقافية وحماية التراث الرقمي الذي يوثق التطور الفكري والثقافي للأمم. كما تساهم عملية أرشفة الويب في دعم البحث العلمي من خلال توفر مصادر للبيانات والبحوث على مر الزمن.

يرتبط تعريف التراث الرقمي بالإنترنت نفسه، باعتباره فضاءً يحتوي عدداً هائلاً من المواقع والمعطيات، تُوظف فيه الوسائل الإلكترونية لجمع المواقع وتحويلها إلى مجموعات أرشيفية تُحفظ ويُتاح لاحقاً (Chautemps, 2020)، مما يرسّخ فكرة أنّ التراث الرقمي ليس مخزوناً ثابتاً، بل حقل موارد متجدداً يتطلب عمليات جمع منظمة من خلال مناهج مختلفة. ستمثل هذه المقاربات لأرشفة مواقع الويب المحور الأساسي لدراستنا بالإضافة إلى استعراض مختلف التحديات التي يمكن أن تواجه كل مشروع لأرشفة مواقع الويب.

الإشكالية

تتمحور إشكالية هذا البحث حول أرشفة مواقع الويب وذلك في إطار ضخامة الإنتاج الرقمي على الويب من جهة، وهشاشة هذا الإنتاج وصعوبة حفظه على المدى الطويل من جهة أخرى. فقد أدى الانتقال إلى النشر الرقمي إلى بروز ما يُسمّى بالتراث الرقمي، الناتج عن نشر المزيد من الوثائق بصورة رقمية، وهو تراث يحتاج إلى الجمع والحفظ لضمان تعداد كامل للإنتاج الثقافي والفكري والعلمي للبلدان. هذا التراث يأخذ الأهمية نفسها التي يحظى بها التراث الورقي، غير أنّ الطبيعة الرقمية للوثائق خلقت مخاطر تهدد هذا التراث بالاندثار، وهو ما يستدعي حلولاً عملية، تأتي أرشفة الويب في مقدمتها كاستجابة مباشرة لمشكلة هشاشة الرقمنة.

وتطرح الإشكالية من خلال الأسئلة التالية :

- ماهو دور أرشفة الويب في حفظ التراث الرقمي للبلدان؟
- ما هي مناهج أرشفة الويب؟
- ما التحديات التي تواجه عملية أرشفة الويب؟

أهداف الدراسة وحدودها

تتمثل الأهداف الرئيسية لهذه الدراسة في تحليل واقع أرشفة الويب في السياقات الدولية، مع التركيز على المناهج والتقنيات المعتمدة في جمع الموارد الرقمية وحفظها، وفهم حدود هذه المناهج من

حيث التغطية والشمولية. وتسعى الدراسة أيضاً إلى حصر التحديات التي يمكن أن تؤثر على أرشفة مواقع الويب.

أما من حيث الحدود، فتقر الدراسة بأن اعتمادها يقوم حصراً على ما توثقه المصادر المتاحة حول مشاريع وتجارب محددة، مثل أرشفة مواقع رؤساء الوزراء الفرنسيين ومشروع Minerva لمكتبة الكونغرس وتجارب أخرى. وبالتالي لا تغطي جميع مبادرات أرشفة الويب عالمياً. كما تقف الحدود أيضاً عند الإطار الذي ترسمه المعايير التقنية والتنظيمية الموصوفة في هذه الدراسات، خاصة ما يتعلق بالقيود التنظيمية والقانونية، مما يجعل نتائج البحث مرتبطة بهذه الحالات المدروسة دون تعميم يتجاوزها.

منهجية الدراسة

تعتمد الدراسة على المنهج الوصفي التحليلي من خلال مراجعة الأدبيات المتخصصة التي تناولت موضوع أرشفة الويب والسياسات المؤسسية ذات الصلة، مع التركيز على تتبع سياقات إنتاج هذه المشاريع ومنهجية الأرشفة وآليات الوصول إلى الموارد الرقمية. الخطوة التالية في المنهجية تتمثل في تحليل مشاريع أرشفة محددة بوصفها دراسات حالة التي تعكس مراحل تطور مشاريع أرشفة الويب في العالم منذ سنة 1996 وذلك لحصر مختلف المقاربات المتبعة في أرشفة مواقع الويب.

أهمية البحث

النظرية : تتمثل أهمية البحث في إستجلاء طرق ومناهج أرشفة مواقع الويب وإثراء المكتبة العربية في هذا المجال، إذ يلاحظ، من خلال مراجعة الأدبيات العربية، غياب البحوث في هذا الإطار.

التطبيقية : تمكن الدراسة من إستكشاف التجارب الدولية في أرشفة الويب مما يساعد مؤسسات الذاكرة في الوطن العربي، التي تعتزم أرشفة مواقع الويب، من اختيار المقاربة التي تتماشى وطبيعة مشروعهم لحفظ التراث الرقمي المنشور عبر الإنترنت والإحاطة بمختلف القيود التي يمكن تعيق هذا المشروع.

تعريف المصطلحات

الويب : الويب، أو World Wide Web بالإنجليزية، هو نظام عالمي من الصفحات والموارد الرقمية المترابطة عبر روابط نصية تشعبية، ويمكن الوصول إليها عبر الإنترنت باستخدام متصفح (Barats, 2014).

أرشفة الويب : هي عملية جمع الصفحات والمواقع الإلكترونية وتسجيلها وحفظها من أجل الحفاظ على الذاكرة الرقمية.

وتشير أرشيفات الويب إلى مجمل المحتويات الرقمية التي يتم جمعها وتخزينها وحفظها بهدف الحفاظ على ذاكرة الإنترنت. وتتيح هذه الأرشيفات الوصول إلى الإصدارات السابقة من المواقع الإلكترونية والوثائق الرقمية والموارد الأخرى المتاحة عبر الإنترنت، حتى بعد اختفائها (Chaimbault, 2008).

التراث الرقمي : يشير هذا المصطلح إلى مجموعة الموارد الثقافية والتاريخية والمتعلقة بالذاكرة التي يتم الحفاظ عليها وإثرائها ونشرها بفضل التقنيات الرقمية: مثل الخرائط التفاعلية وأنظمة إدارة المحتوى متعددة اللغات وقواعد البيانات ومواقع ومنصات الويب المتاحة للجمهور (Molinaro, 2005).

الإطار النظري : الويب والتراث الرقمي

• الويب وتطوره

بدأت شبكة الإنترنت مشروعاً عسكرياً سنة 1969 من خلال شبكة ARPANET الأمريكية لتبادل المعلومات قبل أن تصبح أداة للباحثين، ثم شبكة عالمية يمكن لأي شخص استخدامها. وتعمل شبكة الإنترنت وفق بروتوكولات الاتصال لتبادل البيانات (بومرة، حجار، 2005)، كما توفر مجموعة من التطبيقات كالبريد الإلكتروني الذي ظهر سنة 1971 حيث أرسل المهندس راي توملينسون أول بريد إلكتروني. وقد استخدم الرمز “@” لفصل اسم المستخدم عن الخادم.

وفي سنة 1989 جاء تيم بيرنرزلي بفكرة الشبكة العنكبوتية العالمية، وهو نظام لربط صفحات الويب ببعضها البعض وما يعرف الآن باسم شبكة الويب العالمية (WWW). والويب هو نظام معلومات موزع ومجموعة من الوثائق المترابطة عن طريق روابط النص التشعبي، ويمكن الوصول إليها عبر الإنترنت. وتعتمد شبكة الويب على ثلاث ركائز تكنولوجية أساسية:

1. **HTML** (لغة ترميز النص التشعبي): لغة الترميز لتهيئة صفحات الويب وإنشائها.

2. **HTTP** (بروتوكول نقل النص التشعبي): البروتوكول الذي يسمح لمتصفحات الويب بالاتصال بخوادم الويب لاسترداد الصفحات.

3. **URL** (محدد موقع الموارد الموحد): نظام العنونة الذي يحدد عنوان كل موقع على الويب. وينقسم تاريخ الويب إلى عدة “أجيال”، يمثل كل منها طفرة كبيرة من حيث التكنولوجيا والاستخدام وتجربة المستخدم. ولكل جيل من الويب خصائصه التي تمثل مرحلة رئيسية في تطوره.

الويب 1.0: الويب الثابت (1990-2004)

كان الويب 1.0 للقراءة فقط و يتكون بشكل أساسي من صفحات ثابتة، تم تصميمها ووضعها على الإنترنت من قبل الشركات أو الأفراد لنشر المعلومات. وكان المستخدم في المقام الأول مستهلكاً للمحتوى فقط دون تفاعل (Raman, 2025).

الويب 2.0: الويب الاجتماعي والتشاركي (2004-2016)

كان الانتقال إلى الويب 2.0 ثورة في تحويل الويب من مجرد مجموعة بسيطة من المعلومات إلى منصة تفاعلية وتعاونية (العيصري، الحاج، 2019). ويشدد المصطلح، الذي شاع من قبل شركة أورابلي ميديا، على "المشاركة" و"التفاعل"، إذ يشترك الأفراد في إنشاء المحتوى، ومشاركته، وتعديله. ويشمل ذلك استخدام أدوات مثل الويكي، والمدونات، ومنصات التواصل الاجتماعي، وغيرها من التطبيقات التي تتيح التفاعل والتعاون بين المستخدمين.

الويب 3.0: الويب الدلالي (بعد 2016)

يميل الويب 3.0 نحو اللامركزية في الإنترنت، ولم يعد الأمر يتعلق بتبادل المعلومات أو التفاعل فحسب، بل أصبح يتعلق كذلك بإعطاء الويب القدرة على "فهم" معنى البيانات. ويوفر الويب 3.0 مزايا مثل سرية البيانات والتحكم فيها استناداً إلى سلسلة الكتل والذكاء الاصطناعي وقابلية التشغيل البيئي (Decima, 2026).

الويب 4.0

غالباً ما يُوصف الويب 4.0 بأنه «الويب المستقل»، حيث يمكن للآلات وتطبيقات الذكاء الاصطناعي القراءة والكتابة وتنفيذ المعاملات بشكل مستقل نيابة عن المستخدم. ويعتبر الويب 4.0 نظام بيئي رقمي يتقل فيه وكلاء الذكاء الاصطناعي ويتفاعلون مع تطبيقات الويب ويؤدون المهام عبرها دون تدخل بشري مباشر. ويحاكي وكلاء الذكاء الاصطناعي هؤلاء سلوكيات التصفح البشرية، ولكن بسرعات وكفاءات تفوق قدرات البشر بكثير (Raman, 2025).

نحو الويب 5.0

يتطلع البعض إلى المرحلة التالية من تطور الويب نحو «الويب 5.0». ويتصور البعض «الويب 5.0» على أنه «الويب العاطفي»، حيث تتيح تقنيات الذكاء الاصطناعي المتقدمة وتحليل المشاعر للآلات فهم المشاعر البشرية والاستجابة لها. في هذا السيناريو، يصبح الويب أكثر بديهية وتخصيصاً، حيث يتكيف في الوقت الفعلي مع الحالات العاطفية للمستخدمين والسياق الذي يتواجدون فيه (Decima, 2026).

لقد فرض الويب نفسه في العقود الأخيرة وسيلة مفضلة لنشر المحتويات والتجارب ومشاركتها. في البداية، كان يُفكر في الويب أداة تسمح بجعل الوثائق الموجودة في أماكن أخرى متاحة وتسهيل الوصول إليها من أي مكان. لكن الويب تجاوز بسرعة هذا الوضع ليصبح مكاناً أصيلاً وأصلياً لإنشاء المحتوى: أصيل لأنه يتم إنشاء المحتويات فيه من البداية ضمن هذا الإطار التقني، وأصلي لأن المحتويات لا يمكن أن توجد في مكان أو إطار آخر لأنها تستفيد من الخصائص التي يقدمها الويب. وإذا كان الإنترنت يشكل الآن مكاناً افتراضياً لنشر المحتويات وتوزيعها، فإنه يتميز أيضاً بوجوده الزمني الذي

يمكن أن يكون متغيراً للغاية: من بضع دقائق أو ساعات إلى عدة سنوات. ومع التطور المتسارع لاستخدام الإنترنت في نشر المعلومات، أصبحت أرشفة هذه المعلومات ضرورة ملحة للحفاظ على ذاكرة المؤسسات للرجوع إليها على المدى الطويل (عمرون، قوميدي، 2022).

• التراث الرقمي وأرشفة الويب

يُعرّف التراث الرقمي في بعض الدراسات بوصفه النتاج المتزايد لنشر الوثائق في صيغ رقمية على الويب، بحيث يشكل هذا النتاج في حد ذاته تراثاً يستلزم الجمع والحفظ ضماناً لتعداد كامل للإنتاج الثقافي في البلدان المختلفة (قوميدي، 2024). وتؤكد هذه الدراسات أن هذا التراث يأخذ الأهمية نفسها التي يحظى بها التراث الورقي، غير أن طبيعته الرقمية تجعله عرضة لمخاطر اندثار جدية، ما يجعل أرشفة الويب حاجة ملحة لمعالجة هشاشة الرقمنة وما يرافقها من احتمالات فقدان الصفحات والنصوص والوثائق السمعية البصرية.

واعتمدت اليونسكو سنة 2003 ميثاقاً بشأن حفظ التراث الرقمي. ويهدف هذا الميثاق إلى حماية الذاكرة الرقمية العالمية المهددة بالخطر، كما يعكس إدراكاً متزايداً لأهمية التراث الرقمي وضرورة اتخاذ إجراءات دولية لحمايته. وحسب هذا الميثاق، يتكون التراث الرقمي من موارد المعرفة أو التعبير الإنساني، سواء كانت ثقافية أو تعليمية أو علمية أو إدارية، أو تشمل المعلومات التقنية والقانونية والطبية وغيرها من أنواع المعلومات، يتم إنشاؤها بشكل متزايد رقمياً، أو تحويلها إلى شكل رقمي من الموارد التناظرية الموجودة في البداية (Bermes, 2020). فقد أشار هذا الميثاق إلى نوع جديد من التراث وهو "التراث الرقمي الأصلي" الذي يتضمن جميع المحتويات والبرمجيات والصيغ التي تم إنشاؤها مباشرة على وسائط رقمية، كمواقع الويب التي يجب حفظها لضمان إمكانية الوصول إليها في المستقبل بإعتبارها رمزا لذاكرة المؤسسات.

ومن خلال هذه الخطوة، تليي اليونسكو الحاجة المتزايدة إلى تنويع مفهوم التراث، الذي كان يقتصر فقط على التراث المادي، مثل المباني أو الكتب أو المخطوطات وغيرها، ليكون أكثر شمولية بإدراج التراث الرقمي.

لمحة تاريخية عن تجارب أرشفة مواقع الويب في العالم

تتضمن أرشفة الويب التقاط محتوى المواقع الإلكترونية وحفظه وإتاحة للوصول إليه، ويستجيب ذلك للحاجة الأساسية لحفظ الذاكرة الرقمية على أساس دائم، نظراً لطبيعة صفحات الويب سريعة الزوال والتغيير، والتي هي عرضة للتعديل أو النقل أو الحذف.

ويمثل عام 1996 السنة الأولى لأرشفة الويب إذ بدأت أولى بوادر الوعي بأهمية حفظ المحتوى المتاح على الويب. فقرر بعض الأشخاص التحرك لمواجهة ضياع هذه المعلومات. وكان بروستر كاهل

(Brewster Kahle) أول من طرح فكرة أرشفة الويب سنة 1995 (Musiani, 2019)، وبعد هذا التاريخ ظهرت عديد المبادرات نذكر منها:

أرشيف الإنترنت Internet Archive

منذ عام 1995، بدأت فكرة أرشفة الويب العالمي تتبلور في ذهن بروستر كاهل الذي كان شديد الحساسية لمشكلة أخطاء 404 التي أصبحت أكثر شيوعاً، والتي كانت الدليل الوحيد المتبقي على موقع أو صفحة ويب اختفت. فكر كاهل في وسيلة لتجنب هذا الفقدان وبدا له أن أرشفة هذه الصفحات قبل اختفائها هي الحل الأكثر وضوحاً وفعالية للحفاظ عليها. ومن هذا المنطلق أنشأ مؤسسة أرشيف الإنترنت Internet Archive. أطلقت هذه المؤسسة رسمياً في إبريل 1996 في سان فرانسيسكو بهدف جمع وأرشفة كل ما يمكن للبشرية نشره وابتكاره على الإنترنت حتى يتمكن الجميع من الاستفادة من الوصول إلى هذه المعرفة. وكان مشروع أرشيف الإنترنت يعتمد على مبدأ أن كل ما يُنشر على الإنترنت يُعتبر عاماً، وبالتالي يمكن جمعه. فيتم جمع كل شيء دون أي اختيار مسبق.

في نهاية عام 1997، طور بروستر كاهل برنامجاً يتيح تصفح صفحات الويب وجمعها وفهرستها تحت إسم أليكسا. كما يجمع هذا البرنامج معلومات حول زيارات المواقع، وتكرار تحديثها، وعدد الروابط المرتبطة بها، وغيرها. كما يتيح برنامج أليكسا الوصول إلى الإصدارات السابقة من المواقع من خلال البحث عنها مباشرة في أرشيف الإنترنت (Khouaja, Couture, 2004).

ومنذ عام 1998، لم يتوقف حجم الويب عن الازدياد، مما جعل من الصعب بشكل متزايد إجراء جمع يدوي شامل. فقد كان عدد الصفحات يتضاعف كل ثلاثة إلى ستة أشهر، وكذلك عدد المستخدمين. ولمواكبة هذا الارتفاع في عدد الصفحات المطلوب جمعها، تعاونت أليكسا مع اثنين من كبار الفاعلين في مجال المعلوماتية والويب: مايكروسوفت ونتسكيب. وبفضل هذه الشراكة، تمكنت أليكسا من دمج الإضافة البرمجية الخاصة بها في محركات البحث إنترنت إكسبلورر ونتسكيب، مما سمح لها بتغطية نحو 90٪ من أجهزة الحاسوب في العالم في تلك الفترة.

وفي سنة 2001 تم إحداث تقنية جديدة "آلة العودة بالزمن" (Wayback Machine). وأنشئت هذه التقنية عن طريق شركة أليكسا إنترنت، بموجب عقد مع أرشيف الإنترنت (Internet Archive).

إن Wayback Machine هي أداة لأرشفة الويب تعمل من خلال أخذ لقطات دورية لمواقع الويب. يتم تخزين هذه اللقطات في قاعدة بيانات واسعة، مما يمكن المستخدمين من الرجوع إلى الإصدارات السابقة من المواقع الإلكترونية، حتى لو تم تعديلها أو حذفها أو لم يعد من الممكن الوصول إليها. وتقوم أداة Wayback Machine بتحديث الأرشيف باستمرار بالتقاط صور جديدة، مما يوفر تاريخاً

متطوراً للويب (Chebbi, 2013). وقد انتشرت هذه الأداة في العديد من المبادرات الخاصة بأرشفة الويب، لا سيما في المكتبة الوطنية الفرنسية (BnF).

باندورا Pandora

أطلقت المكتبة الوطنية في أستراليا، بالشراكة مع الأرشيفات الوطنية، مشروع باندورا سنة 1996، وهو مشروع لحفظ وإتاحة الموارد الوثائقية الشبكية لأستراليا. حتى نهاية عام 1997، ركز فريق العمل على تطوير سياسة وثائقية وتحديد إجراءات الاختيار والجمع والأرشفة التي سيتم استخدامها. وأدى هذا العمل إلى إجراء أول عملية جمع في نهاية العام، حيث تمت أرشفة 229 وثيقة. وحسب آخر إحصائيات المشروع سنة 2023 وفق موقع المكتبة الوطنية الأسترالية، يبلغ عدد المواقع والصفحات المؤرشفة 16.486.594.851 ما يعادل 774 تيرا بايت.

KulturarW3

في سبتمبر 1996 أطلقت المكتبة الملكية بالسويد مشروع KulturarW3، ويعني kulturarv التراث الثقافي باللغة السويدية وwww الخاصة بشبكة الويب. وفي هذا المشروع تم اختيار المقاربة الشاملة لجمع كامل محتوى الويب الوطني، أي جميع المواقع التي تحمل نطاق.se، دون معايير للجودة أو الملاءمة (Khouaja, Couture, 2004). وقد تم تبرير هذا الخيار الشامل بصعوبة التنبؤ بأي محتوى سيكون مهماً للبحث أو للأرشفة في غضون 20 أو 80 عاماً. ويمكن هذا المسح الشامل والمنظم لمواقع الويب من أرشفة الويب الوطني بتكلفة منخفضة. وحتى سنة 2003 كان المشروع جزءاً من أرشفة الويب الإسكندنافية، الذي جمع مبادرات الدول الإسكندنافية. وبعدها انضم المشروع السويدي إلى التكتل الدولي لحفظ الإنترنت IIPC.

من خلال هذه التجارب وبمراجعة لأدبيات المتعلقة بأرشفة مواقع الويب تبين أن البحوث الأولى ركزت في المقام الأول على تقنيات جمع صفحات الويب الثابتة وحفظها، وغالباً ما كان ذلك عن طريق برامج الزحف crawlers. ومع ظهور الويب الديناميكي والشبكات الاجتماعية والمحتوى المتعدد الوسائط تطورت الأساليب والمناهج في أرشفة مواقع الويب. وفي هذا الإطار، تستكشف الأدبيات الحالية تحديات مثل إدارة خصوصية تقنيات الويب الحديث، والحفاظ على التفاعل والدقة البصرية، والجمع الانتقائي وفقاً لمعايير محددة (المواضيع، والفترات، والأحداث)، والقضايا القانونية والأخلاقية التي ينطوي عليها جمع محتوى الويب وإتاحته في ظل حماية البيانات الشخصية. وبالإضافة إلى ذلك، هناك اهتمام متزايد بتحليل البيانات الواردة في أرشيفات الويب واستخدامها لأغراض البحث الأكاديمي، لا سيما في التاريخ والعلوم الاجتماعية والاتصالات (Gebiel, 2024).

ومع زيادة الوعي بأهمية أرشيف مواقع الويب جاءت عديد المقترحات لتقدم مناهج وأساليب للأرشفة. فقد أجرى تشارلز دولار دراسة لصالح أرشيف معهد سميثسونيان في واشنطن عرض من خلالها استراتيجية لإدارة موارد الويب (Dollar 2001)، إذ اقترح عدداً من النماذج لاختيار صفحات الويب الخاصة بالمؤسسة وإدارتها وحفظها. ومن جانبها، حاول مكور وسبريهي وضع دليل لإدارة السجلات الإدارية من المواقع الإلكترونية للإدارات الاتحادية وإدارات الولايات في الولايات المتحدة. وكان الهدف من هذه الدراسة هو وضع استراتيجية لإدارة وحفظ أرشيفات الويب من خلال توفير معلومات عن كيفية تحديد هذه الوثائق، وأساليب الإدارة حسب نوع موقع الويب (ديناميكي أو ثابت) ومدى تعرضه للخطر. وقد تناول الموضوع نفسه ليفورجي، من بين آخرين. فهو يرى أن حفظ جزء من موقع ويب مؤسسي ليس حلاً، لأنه لا يوجد ضمان لحفظ الجزء الأكثر قيمة (Khouaja, Couture, 2004).

كما تم نشر دليلين: الأول بعنوان "أرشفة موارد الويب: مبادئ توجيهية لحفظ سجلات النشاط القائم على الويب في حكومة الكومنولث"، أصدره الأرشيف الوطني الأسترالي؛ والثاني، من مكتب السجلات العامة في المملكة المتحدة، بعنوان إدارة السجلات الإلكترونية على المواقع الشبكية والإنترنت: أداة لإدارة موارد الويب. وقد كانت هاتان المؤسستان أول من وضع دليلاً لإدارة المواقع الإلكترونية المؤسسية في بلديهما. وفي كندا، تم إصدار دليل لأرشفة المنشورات الإلكترونية للإدارات الحكومية التي يجب إيداعها لدى المكتبة الوطنية وفقاً لأحكام قانون الإيداع.

إن الاهتمام الذي أبداه هؤلاء المؤلفون بمسألة إدارة المواقع الإلكترونية وأرشفتها قد نشأ نتيجة لزيادة الوعي في عدد من الدول التي أدركت أهمية الإنترنت والدور الكبير الذي تلعبه في نشر المعلومات في جميع مجالات المعرفة. ومن ثم بدأت هذه الدول في تجربة إدارة تراثها على الإنترنت وحفظه.

التكتل الدولي لحفظ الإنترنت IIPC وأرشفة الويب

(International Internet Preservation Consortium)

تأسس التكتل الدولي لحفظ الإنترنت سنة 2003، ويجمع المؤسسات التراثية الرئيسية كالمكتبات الوطنية من جميع أنحاء العالم. ويعمل كمنصة لتبادل الخبرات بين الخبراء من المؤسسات الأعضاء فيه بهدف حل المشاكل المشتركة ووضع مبادئ توجيهية للحفظ الفعال طويل المدى لمواقع الويب.

كما يهدف التكتل الدولي لحفظ الويب إلى مساعدة أعضائه على تطوير استراتيجيات وأدوات للزحف وجمع المليارات من صفحات الويب واستيعابها، إضافة إلى تطوير حلول تقنية وأطر أخلاقية وقانونية لجعل أرشيفات الويب قابلة للبحث من قبل الباحثين والمؤرخين وعامة الناس، مع احترام حقوق النشر والخصوصية.

ومن أجل الحفاظ الفعال والقابل للتشغيل المتبادل لأرشفات الويب، يلعب IIPC دوراً حاسماً في تطوير واعتماد تنسيقات مثل تنسيق ملفات WARC (Web ARChive)، الذي أصبح معياراً دولياً لأرشفة الويب. كما يدعم التكتل أيضاً تطوير أدوات برمجية، مثل برامج الزحف (Heritrix) وواجهات الإطلاع (OpenWayback).

باختصار، يعمل التكتل الدولي لحفظ المواقع الإلكترونية بمثابة ركيزة أساسية في الجهود العالمية لحماية تراثنا الرقمي. ومن خلال مهامه في التنسيق وتطوير التكنولوجيا واحتضان أفضل الممارسات يحافظ على الذاكرة الجماعية والتنوع الثقافي والتراث الفكري لعصرنا الرقمي. وبالتالي يضمن أن تكون أرشفات المواقع متاحة للدراسة والفهم من قبل الأجيال القادمة.

مناهج أرشفة مواقع الويب

تعتمد أرشفة مواقع الويب على مبدأ تجميع، عادةً ما يكون تلقائياً ودورياً، عناوين URL التي تقوم بها برامج، ويتم تسجيل صفحات الويب وحفظها. لذا نحصل على مجموعة من الصفحات التي يمكننا الإطلاع عليها كما لو كنا نفوس في الويب من الماضي. وفي الواقع، نظراً لأن الجمع لا يكون أبداً شاملاً أو دائماً، فإن الأرشيف الذي يتم الحصول عليه هو عينة زمنية ومكانية من الويب على فترة معينة. (Bachimont, 2024) وقد استندت مختلف المبادرات لأرشفة محتوى الويب وحفظه في العالم على ثلاث مقاربات :

المقاربة الشاملة

تتمثل في أرشفة جميع محتويات الويب وحفظها باستخدام روبوت آلي لالتقاط جميع المواقع وفقاً لاختيار تم تحديده مسبقاً (حسب البلد، المجال، التخصص). تقضي هذه الطريقة على خطر فقدان الوثائق التاريخية بسبب الأحكام القيمية التي يتم إصدارها بشأنها. وتعتبر هذه المقاربة الأقل تكلفة مقارنةً بالمناهج الأخرى. ومن أشهر نتائج هذا النهج Wayback Machine الذي يعتمد عليه مشروع Internet Archive بوصفه مشروعاً يهدف إلى جمع الويب العالمي اعتماداً على لقطات آلية، وقد أتاح هذا الأسلوب تكوين نسخ متعددة من الموقع نفسه. لكن هذا النموذج الآلي لا يضمن جودة المواقع المجموعة أو صلتها، كما أنه يقتصر على الجزء المرئي من الويب الذي تستطيع الروبوتات الوصول إليه، ولا يجمع المواقع المؤمّنة أو ما يتصل بواجهات قواعد البيانات (Beis, 2023).

المقاربة الانتقائية

تم تطبيق هذه المقاربة في دول مثل أستراليا والولايات المتحدة الأمريكية، وتستند على اختيار دقيق لصفحات الويب التي سيتم أرشفتها وفقاً لمعايير محددة مسبقاً. ويتم تنفيذه من قبل متخصصين يقيمون أهمية موارد الويب التي سيتم حفظها. وقد تم اعتماد هذا المنهج من قبل أستراليا في عام 1996 لأرشفة الويب في مشروع PANDORA. وكانت الاستراتيجية المتبعة انتقائية للغاية: فقد كانت تتألف

من حفظ المنشورات في شكل إلكتروني فقط التي ليس لها مثل ورقي والتي تعتبر ذات أهمية وطنية (Khouaja, Couture, 2004). وكان الهدف هو إنشاء مجموعة تمثيلية بدلاً من أن تكون شاملة، حيث يقوم النظام بأخذ مقتطفات من عدد من المواقع بانتظام بعد فحصها وتخزينها على الخوادم. كما تم اعتماد هذه المنهجية في مشاريع مؤسسية تضبط مجال الجمع مسبقاً كمشروع أرشفة مواقع رؤساء الحكومات الفرنسية الذي اقتصر على مواقع حكومية محددة . وفي مشروع Minerva قامت مكتبة الكونغرس بأرشفة مجموعة من المواقع تم اختيارها بناء على معايير مرتبطة بموضوع أو حدث أو مجال معين. كما نجد التجربة البريطانية التي تقوم على الجمع الانتقائي حسب الموضوع أو المجال، حيث تتولى مؤسسات مختلفة مسؤوليات متخصصة، فالمكتبة البريطانية تجمع المواقع ذات الأهمية الثقافية والتاريخية والسياسية، بينما تُعنى Welcome Library بالمجال الطبي. (Gebeil, 2024).

المقاربة التكاملية

تتمثل هذه المقاربة في جمع تلقائي لمواقع الويب مصحوبا بمتابعة فردية ويدوية لضمان شمولية العينة وتمثيليتها لمواقع الويب. وترتكز هذه المقاربة على متابعة دقيقة لبعض المواقع التي لا يمكن للروبوتات الوصول إليها، ما يسمى بالويب العميق (Castier 2002)، وذلك لأسباب تقنية، أو بسبب قيود من نوع المواقع المدفوعة أو المقيدة لفئة محددة من المستخدمين. هذا النوع من المحتوى، الذي يكون عادةً أكثر ثراءً من بقية محتويات الويب، يصعب أرشفته باستخدام المقاربة الشاملة. وفي المقابل، تتطلب عملية الاختيار البسيطة، التي تتم يدوياً بحثاً، مزيداً من الوقت وقد لا تواكب تطور الكم الهائل من المعلومات التي يتزايد حجمها باستمرار (Ouerfelli & al., 2024).

تجربة المكتبة الوطنية التونسية لأرشفة الويب

إن توسيع نطاق الإيداع القانوني في تونس، من خلال القانون رقم 37 لسنة 2015، ليشمل المصنفات الإلكترونية مكن المكتبة الوطنية من إطلاق برنامجها لأرشفة الويب. وانطلقت أولى التجارب سنة 2016 بتجميع حوالي 560 وثيقة وهي عبارة عن مقالات الرأي المنشورة على مختلف الصفحات والمواقع الإلكترونية أو على شبكات التواصل الاجتماعي أو الموزعة في الصحف الإلكترونية. وتواصل هذا العمل في عام 2017 بهدف حماية الذاكرة الجماعية حول الثورة التونسية لعام 2011. وتضمن هذا العمل جمع كل ما نشر حول أحداث الثورة على شبكات التواصل الاجتماعي، ولا سيما فيسبوك وعلى الصحف الإلكترونية ومدونات الناشطين (Chebbi, 2022).

كما قامت المكتبة الوطنية بجمع الويب التونسي خلال فترة الحجر الصحي المرتبط بجائحة كوفيد-19. وخلال هذه الأزمة، اعتمدت العديد من المؤسسات، لا سيما في مجال التعليم والبحث، أسلوب الاتصالات عن بُعد لضمان استمرارية خدماتها مع تقليل المخاطر المرتبطة بانتشار فيروس

كورونا. وقد نُظمت العديد من الفعاليات والدورات التدريبية والمناقشات بأسلوب الندوات عبر الإنترنت. ولحفظ هذا التراث العلمي والمعرفي الرقمي شرعت المكتبة الوطنية التونسية في حصره وجمعه. وتم تجميع الوثائق في الفترة ما بين مارس ويوليو 2020 بهدف إنشاء مكتبة رقمية "ندوات عبر الإنترنت أثناء الحجر" تضم أكثر من 194 ساعة من التسجيلات (أكثر من 30 جيجابايت من البيانات) بإجمالي 142 تسجيلًا (Chebbi, 2022).

تحديات أرشفة مواقع الويب

إن التحديات التي تواجه أرشفة الويب عديدة، منها ماهو تقني ومنها ما هو قانوني.

التحديات التقنية

تتمثل هذه التحديات في:

الحجم الهائل والنمو المتسارع: مع مليارات الصفحات والمواقع التي تنمو بمعدل كبير، يصبح جمع كل شيء وأرشفته مهمة شبه مستحيلة. ويتطلب تخزين هذه الكميات الهائلة من البيانات بنية تحتية ضخمة ومكلفة للغاية، ناهيك عن قوة المعالجة اللازمة لفهرسة هذا المحتوى (Zielinski, 2019).

الطبيعة الديناميكية والمحتوى المتغير: على عكس المستندات الثابتة، فإن معظم صفحات الويب الحديثة ليست مجرد ملفات HTML بسيطة، إذ تعتمد العديد منها على قواعد البيانات وتستخدم تقنيات مثل JavaScript، AJAX، و APIs لإنشاء محتوى ديناميكي يتغير باستمرار بناءً على تفاعل المستخدم أو تحديثات الوقت الفعلي. وبالتالي تتطلب أرشفة هذا النوع من المحتوى أدوات متقدمة يمكنها محاكاة تجربة المستخدم لالتقاط الحالة الصحيحة للصفحة، وهو أمر معقد تقنياً.

تنوع المحتوى وتشعبه: لم يعد الويب يقتصر على النصوص والصور بل أصبح يشمل أيضاً الفيديوهات والصوتيات والألعاب التفاعلية والمدونات ووسائل التواصل الاجتماعي والبث المباشر وغيرها. وكل نوع من هذه المحتويات له تسيقاته ومتطلباته الخاصة للأرشفة والتشغيل، مما يتطلب تنوع الأدوات والتقنيات لضمان إمكانية الوصول إليه وتشغيله لاحقاً بنفس الطريقة التي كان عليها عند أرشفته.

التحديات القانونية

تتعلق التحديات القانونية بمسألة حقوق التأليف. ففي السياق الرقمي، تشمل حقوق التأليف كل ما يُنشر على الإنترنت: النصوص والصور ومقاطع الفيديو والموسيقى والبرمجيات وحتى تصميم المواقع. وهو ما يطرح إشكالية عند أرشفة المواقع باعتبار أن نسخ محتوى من موقع ويب، أو إعادة نشره، أو عرضه للجمهور، يتطلب عادةً إذنًا من صاحب حقوق التأليف.

وفي هذا الإطار، وللتوفيق بين حماية حقوق التأليف وضرورة الأرشفة، تطورت مجموعة من الحلول والاعتبارات القانونية والتكنولوجية (Beis, 2023). وتتمثل هذه الحلول في:

الإذن والتفويض : يعتبر الحل الأمثل قانونياً، حيث يتم الحصول على إذن صريح من صاحب حقوق التأليف لنسخ محتواه وأرشفته. ولكن يبقى هذا الإجراء غير عملي على نطاق الويب الواسع.

الإيداع القانوني الرقمي: في بعض الدول، تم توسيع قوانين الإيداع القانوني - التي تُلزم الناشرين بإيداع نسخ من مطبوعاتهم في المكتبات الوطنية - لتشمل المحتوى الرقمي، بما في ذلك الويب. هذا ما يوفر أساساً قانونياً قوياً لأرشفة محتوى معين كما هو الحال في فرنسا. فالإيداع القانوني للمواقع الإلكترونية بفرنسا موزع بين المكتبة الوطنية التي تقوم بأرشفة مواقع المؤسسات وكذلك الأشخاص، والمعهد الوطني السمعي البصري المكلف بالإيداع القانوني لمواقع وسائل الإعلام السمعية البصرية، مثل القنوات التلفزيونية والإذاعية والتلفزيون والراديو على شبكة الإنترنت.

إن المقاربة الفرنسية للإيداع القانوني عبر الإنترنت - والتي لا تختلف في هذا الصدد عن تلك المتبعة في بلدان أوروبية أخرى مثل المملكة المتحدة أو الدنمارك - هي مقارنة غير انتقائية فكل المحتوى على الإنترنت، مهما كانت "القيمة" العلمية أو الفكرية أو الثقافية التي يمكن أن تنسب إليه، يمكن أن يتم إيداعه في مؤسسات التراث الوطني.

وتشير بعض التشريعات إلى استثناءات وقيود في التعامل مع حقوق التأليف إذ نجد مبدأ الاستخدام العادل الذي من خلاله تسمح بعض التشريعات (خاصة في الولايات المتحدة وكندا) بالاستخدام المحدود للمواد المحمية بحقوق التأليف لأغراض مثل البحث العلمي، والتعليم، والنقد، والتقارير الإخبارية، دون الحاجة إلى إذن. وفي هذا الإطار، يمكن للمؤسسات الأرشيفية الاستناد إلى هذا المبدأ، خاصة عندما يكون الغرض غير تجاري من أرشفة مواقع الويب وحفظها. كما نجد استثناءات خاصة للمكتبات والأرشفيات تسمح لها بنسخ المواد لأغراض الحفظ والبحث دون انتهاك حقوق التأليف.

وكذلك لتقليل المخاطر القانونية، تختار بعض جهات الأرشفة تقييد الوصول إلى المحتوى المؤرشف. على سبيل المثال، قد تسمح بالوصول إلى المحتوى المؤرشف فقط من داخل مباني المؤسسات، أو بعد مرور فترة زمنية معينة. كما هو الحال في المكتبة الوطنية بلكسنبورغ التي تتيح الإطلاع على أرشفات مواقع الويب فقط من أجهزة الكمبيوتر المخصصة لذلك داخل مبنى المكتبة.

كما يمكن استخدام بعض القيود التقنية من خلال ملف Robots.txt، وهو عبارة عن ملف نصي يمكن لأصحاب المواقع وضعه على خوادمهم لتوجيه برامج البحث والأرشفة (Web Crawlers) بعدم أرشفة أجزاء معينة من الموقع أو الموقع بأكمله.

النتائج

تعتبر أرشفة الويب أداة رئيسية لحماية التراث الرقمي والذاكرة التاريخية والثقافية من الاندثار. ويظهر موقع أرشفة الويب بوضوح بوصفه أحد الأذرع العملية للحفظ الرقمي، من خلال مشاريع تستهدف النقاط نسخ من المواقع وتخزينها في مستودعات متخصصة للاستفادة منها في الدراسات والبحوث. من الناحية المنهجية تتحدد سياسات الجمع والاختيار في أرشفة الويب وفق نماذج متباينة، يتراوح فيها القرار بين السعي إلى الاتساع الآلي (المقاربة الشاملة) وبين الانتقاء الموجه (المقاربة الانتقائية). وتواجه أرشفة الويب تحديات تقنية تتبع من طبيعة الويب نفسه بوصفه موضوعاً أرشيفياً غير متجانس من حيث البنية والتمثيل. فنجد مواقع ثابتة أو إعلامية تتكون من ملفات وروابط يمكن تمثيلها بوصفها وحدات وثائقية بسيطة نسبياً، وبين مواقع تفاعلية توفر خدمات ونماذج واستجابات للمستخدم، وبين مواقع ديناميكية يتغير محتواها باستمرار لارتباطه بقواعد بيانات محدثة. كما تواجه أرشفة الويب تحديات قانونية متعلقة بحقوق التأليف والنشر التي تفرض قيوداً على نسخ محتوى المواقع وإاحتها للجميع، إضافة إلى غياب نموذج موحد لتنظيم مناهج الجمع وحدود التفويض بين المؤسسات المهتمة بحفظ الذاكرة الوطنية. فالممارسات تُظهر أن بعض برامج الأرشفة تعتمد الجمع الانتقائي الموزع بين مؤسسات ذاكرة متعددة، كما في المملكة المتحدة حيث تتولى ست مؤسسات أعضاء في UK Web Archiving Consortium مسؤوليات متخصصة بحسب الموضوع أو المجال، مثل تكفل British Library بالمواقع ذات الأهمية الثقافية والتاريخية والسياسية، وتكفل Wellcome Library بالمواقع الطبية.

التوصيات

بناء على ما تم التوصل إليه، توصي الدراسة بما يلي :

لابد من تناول موضع أرشفة الويب في إطار مقارنة متكاملة بين مختلف الجوانب التقنية والتشريعية والتنظيمية والأرشيفية وعدم الاكتصار على الجانب التقني فحسب.

يجب أن تتخذ المواءمة مع المعايير الدولية في أرشفة الويب طابعاً مؤسسياً وتقنياً في آن واحد. فوجود اتحاد IIPC بوصفه إطاراً يضم عديد المكتبات الوطنية في العالم، إلى جانب مؤسسة Internet Archive، يقدم مثلاً مباشراً على بناء تنسيق دولي حول ممارسات حفظ الويب يدل على سعي مؤسسات الذاكرة إلى العمل ضمن شبكة مشتركة في مجال حفظ الويب، بما يضع الممارسات الوطنية داخل إطار دولي منظم.

ضرورة تفعيل دور المكتبات والأرشفيات الوطنية في أرشفة الويب من خلال انتقالها من حفظ المطبوعات والوثائق التقليدية إلى بناء سياسات ومشروعات مخصصة لالتقاط الموارد الشبكية وتنظيمها بوصفها مجموعات قابلة للحفظ والوصول يمكن الاستفادة منها في البحوث العلمية.

التفكير في صياغة سياسة وطنية لأرشفة الويب تحدد فيها المنهجية والجمهور المستهدف والأولويات الوطنية للمحتوى المنشور على الإنترنت الذي يجب أرشفته.

خاتمة

إن التراث مسؤولية الجميع وأرشفة الويب وحفظ التراث الرقمي واجب تماماً كما هو الشأن في المحافظة على التراث المادي كالحصون والقصور القديمة، والمخطوطات النادرة. علينا أن ندرك أن ما ينشر اليوم على الإنترنت هو خزان تراثي لذاكرة المجتمعات في المستقبل، فحماية ما كتبناه ونشرناه وتداولناه على شبكة الإنترنت هو حماية لهويتنا الإنسانية الممتدة نحو المستقبل. وهذا يؤكد أهمية أرشفة الويب كأداة رئيسية لحماية التراث الرقمي والذاكرة التاريخية والثقافية من الإندثار. وتتطلب أرشفة الويب التكامل بين مختلف الجوانب التقنية والقانونية المتعلقة بحقوق التأليف وقانون الإيداع الرقمي، إضافة إلى الموائمة مع المعايير الدولية في أرشفة الويب من خلال العمل ضمن شبكة مشتركة تساعد على وضع الممارسات الوطنية داخل إطار دولي منظم. وفي هذا السياق، تتزايد الحاجة في الوطن العربي لإنشاء مبادرات محلية وإقليمية لأرشفة المحتوى الرقمي العربي المتداول على الإنترنت. هذا التراث يتعرض لعمليات إتلاف جراء الحروب والأزمات التي تسببت في تدمير البنى التحتية للمعلومات، أو بسبب غلق المنصات والمواقع. ومن هنا تأتي أهمية هذه المبادرات لحماية التراث الرقمي العربي وصونه وديمومته للأجيال القادمة.

المراجع:

- يومهرة، نور الدين، حجار، ماجدة (2005). الانترنت مفهومها وتجلياته والآثار المترتبة عن استخدامها، مجلة العلوم الاجتماعية والانسانية، مج 6، ع 12، ص 215-232.
- عمرون، مصطفى، قوميد فتيحة (2022). أرشفة الويب وإشكالية المصطلحات والممارسات: التقارب والإنفصال بين المكتبات والأرشيف، مجلة العلوم الاجتماعية والإنسانية، مج 12، ع 1، ص 318-339.
- قوميد، فتحية (2024). أهمية حفظ التراث الرقمي وتحدياته، مجلة الكَلَم، مج 9، ع 1، ص 295-317
- Bachimont, B. (2024). Archiver le web: quelle objectivité construire pour quelles interrogations ?. Les cahiers du numérique, vol. 30-3-4, 117-131.
- Beis, C. (2023). The What, Why, and How of Web Archiving
<https://www.choice360.org/libtech-insight/the-what-why-and-how-of-web-archiving>
- Bermes, E. (2020), « Le numérique en bibliothèque: naissance d'un patrimoine: l'exemple de la Bibliothèque nationale de France (1997-2019) », thèse de doctorat, Paris, Ecole nationale des Chartes, 107 p.
<https://tel.archives-ouvertes.fr/tel-02475991>
- Chautemps, A. (2020). L'archivage du Web ou le Web comme mémoire des sociétés contemporaines, Annales des Mines - Enjeux numériques, Vol 2, n 10, 57-65.
- Chebbi, A. (2022). Entrevue avec Hajer Sahli, directrice des collections, Bibliothèque nationale de Tunisie. Revue Communication, Organisation, Société du Savoir et Information (COSSI), 11.
<https://revue-cossi.numerev.com/articles/revue-11/2790-entrevue-avec-hajer-sahli-directrice-des-collections1-bibliotheque-nationale-de-tunisie2>
- Chebbi, A. (2012). Archivage du Web organisationnel dans une perspective archivistique. Thèse de Doctorat, Université de Montréal, Canada, 357p.
<http://hdl.handle.net/1866/9203>
- Decima, C. (2026). Evolution of the Web: Web 1.0, 2.0, 3.0, 4.0, and 5.0
<https://www.wsiworld.com/blog/how-the-web-has-evolved>
- Dollar, C. M. (2001). The foundational Archival Preservation of Smithsonian Web Resources: Strategies, Principles, and Best Practices, Smithsonian Institution Archives.
- Gebeil, S. (2024). Web vivant et web archivé, aux sources de l'histoire nativement numérique. Les Cahiers du numérique, 2024, 20 (3-4), pp.181-193

- Khouaja, B, Couture, C. (2004). La gestion et l'archivage des sites Web institutionnels, Archives, Vol. 35, N 3-4, 2003-2004 – pp17-41.
- Musiani, F. (2022), Archivage du Web, un enjeu de gouvernance (d'Internet). Presses Universitaires du Septentrion.
- Musiani, F. & al. (2019). Qu'est-ce qu'une archive du web ? OpenEdition Press, 106p.
- National Library of Australia (2008). Online Australian publications: a selection guideline for archiving and preservation by the National Library of Australia .
<http://pandora.nla.gov.au/selectionguidelines.html>
- Ouerfelli T. & al. (2024). Web archiving and digital repositories in the Arab Universities, International conference “Digital Archiving in the Arab World”, Abu Dhabi 29-31 October 2024.
- Rama, I. (2025). Web 4.0 - The Autonomous Web
<https://anchorbrowser.io/blog/web-4-0>
- Zidan, A (2024). Digital Cultural Heritage: Theory and Practice, 19 February 2024
<https://www.ifla.org/news/digital-cultural-heritage-theory-and-practice/>
- Zielinski, O. (2019). L'archivage du Web français par la Bibliothèque nationale de France Une nouvelle approche des missions de la BnF? Université d'Angers, 209, 103p.