أفضل الطرق لأرشفة السجلات الرقمية

د. أشرف عبد المحسن الشريف
مدرس الوثائق والمعلومات
جامعة بنى سويف

إن مشكلة الأرشفة الرقمية تكمن في كيفية الاحتفاظ بالسجلات الأصلية كمـا هـي مع جعلها متاحة باستمرار ولمدد زمنية طويلة , ويمكن تحقيقه من خلال حفظ السجلات الرقمية في أشكال رقمية مفتوحة المصدر Open Source - أي أن عناصر تكوينها لا تتغير بمرور الزمن وبالتالي تحتفظ بالخصائص الأصلية للسجل الرقمي , كما انه لا علاقة لها بحقوق الملكية الفكرية فهذه الأشكال الرقمية مفتوحة المصدر ليست من إبداع احد الشركات , وإنما هي عبارة عن مجموعة من والتيجان والاكواد التي من الممكن لأي شخص أن يكون بها شكلا رقميا مفتوح المصدر . وعليه فان حفظ السجلات الرقمية على الأشكال مفتوحة المصدر يعد حلا جيدا لمشكلة الأرشفة الرقمية .

وحين تبحث الارشيفات الوطنية عن أشكال وصيغ رقمية Object Digital تتناسب مع الحفظ الارشيفى طويل الأمد , عليها أن تستعين بالمبادئ التالية :

1. **أن تكون الأشكال الرقمية متاحة على الدوام ولا تتأثر بالتغيرات التكنولوجية والاصدارت التي تنتجها الشركات كل فترة .**
2. **ألا تسمح الأشكال الرقمية بالتعديل والتبديل فى محتواها .**
3. **أن تكون الأشكال الرقمية مفتوحة المصدر بحيث لا تتقيد بحقوق الملكية الفكرية .**
4. **أن تحتفظ بالسياق الذي أنتجت فيه السجل الرقمي – بياناته الوصفية Metadata – وذلك لإعطاء مزيد من الصحة والموثوقية للسجلات الرقمية .**
5. **أن يحمل الشكل الرقمي من الخصائص ما يؤهله لان يحمل جميع عناصر السجل الرقمي الذي يحويه حتى يظل السجل صالح للاستخدام على الدوام .**

أنواع الأشكال الرقمية المناسبة للأرشفة

هناك من يفضل استخدام ملفات Portable Document Format PDF فى أرشفة السجلات الرقمية للأسباب الآتية :

1- **الارتباطات التشعبية داخل النص وبين الوثائق ومواقع الانترنت .**
2- **إمكانية تصغير أو ضغط الصفحات عند العرض .**
3- **تضمين الخطوط مع الوثيقة مما يضمن عرض الوثيقة بالخطوط نفسها دون التقيد بوجودها على الحاسب الخاص بالعرض**
4- **إمكانية تحميل صفحة بصفحة .**

5- صغر الحجم الناتج عن ضغط الملف الذي قد يصل لحجم أصغر من الملف النظير له في صيغة PostScript .
6- صعوبة إحداث تغيير في نص الوثيقة بالإضافة إلى انه يحتفظ بشكل الوثيقة نفسه كما هو أيا كان برنامج المستخدم في العرض .
7- تأمين الدخول إلى نص الوثيقة من خلال تحديد كلمة مرور, وتقييد صباغة النص أو أجزاء منه .

إلا أن من عيوبه ما يلي :

1- ليس بالشكل الجيد لعمليان الحفظ التاريخي طويل المدى بسبب التغيير والتطوير المستمر في لغة توصيف الصفحة من قبل الشركة المنتجة لهذا البرنامج .
2- من الصعب تصميم برمجيات تحويل محتوى ملفات PDF إلى صيغ أخرى .
3- لا تتيح الصيغة بنية الملف وبالتالي يصعب البحث عن جزء أو فقرة في النص .

وللتغلب على مشكلة الحفظ التاريخية فانه يمكن حفظ السجلات ومحتواها في صيغة تتناسب مع الحفظ طويل المدى مثل ASCII ,TIFF, SGML ثم تحويلها إلى صيغة DDF لإغراض الإتاحة والعرض فقط .

الأشكال مفتوحة المصدر Open Source

وهى أشكال وصيغ رقمية يستخدم في إنشاءها مجموعة من الواصفات المعيارية – مجموعة من الرموز والتيجان – التي حددتها المنظمات الدولية للتوحيد القياسي مثل معيار ISO 10646 وهى أشكال موثقة لا تعتمد على منتج واحد مثل ملفات XML, SGML , ومتاحة للعامة دون التقيد بحقوق الملكية الفكرية .

الأشكال المغلقة :

وهى أشكال رقمية منتشرة على نطاق واسع ولكن تركيبها وخصائصها التكنولوجية ليست متاحة للجمهور وتملكها احد الشركات العمالة في مجال تقنية المعلومات مثل شركة مايكروسوفت التي تملك مجموعة برامج الأوفيس , وشركة أدوبي التي تملك ملفات PDF , وهذه الملفات ليس من حق الجمهور أو أي شخص أو هيئة أن تغير من خصائصها دوت الرجوع إلى الشركة المنتجة .

والمبادئ السابقة جميعها تتوفر في الملفات مفتوحة المصدر , وقد وجدت الارشيفات الوطنية أن لغة الترميز القابلة للتوسع Exchange Mark up Language "XML" هي اللغة الأفضل التي تستخدم في بناء الملفات مفتوحة المصدر التي تحوى داخلها السجلات الرقمية , فهي لغة دولية معتمدة من المنظمة الدولية للتوحيد القياسي.

ولغة XML هي عبارة عن مجموعة محددة مسبقا من الواصفات

والعناصر سواء كانت تيجان أو رموز تستخدم في إدماج وتضمين معلومات خارجية داخل وثيقة نصية الكترونية , وقد اعترف بهذه اللغة عام 1998 وهى عبارة عن نموذج مصغر من لغة الترميز المعيارية العامة "Standard General Mark up Language "SGML , وهى تعتبر مقياس لإدارة المعلومات تم تبنيه من قبل هيئة التوحيد القياسي ISO عام 1986 كوسيلة لإنشاء وثائق قابلة للتنسيق.

ومهمة لغة XML تتركز في ترميز المعلومات وصياغتها في بناء هيكلي موحد يسهل التعامل معه بواسطة كافة الأنظمة والتطبيقات , وقد طور هذه اللغة رابطة الشبكة العنكبوتية W3C وهى عبارة عن مؤسسة مكونة من أكثر من 180 عضو من المنظمات والهيئات التجارية والأكاديمية والحكومية من مختلف أنحاء العالم ومن مميزات هذه اللغة :
1 – تقنين نشر المصادر الالكترونية – السجلات الرقمية - بشكل مستقل وبأسلوب موحد .
2 - تسليم المعلومات إلى البرامج المستخدمة من قبل المستفيدين بشكل يسهل معالجتها آليا والتعامل معها فور استلامها .
3 –سهولة معالجة المعلومات والبيانات وتبادلها عبر مختلف التطبيقات والأنظمة بأقل تكلفة
4- – سهولة دعم البيانات الواصفة Metadata لكل وثيقة مما يسهل استرجاعها ومن ثم إيجاد حلقة اتصال بين منتج أو صانع المعلومات والمستفيد النهائي منها .

ولغة XML هي لغة بسيطة ومتوافقة مع ملفات ASCII وهى رموز تتميز بمقاومتها للتلف بشكل كبير فعلى سبيل المثال لو أن كلمة Welcome نقص منها حرف "e" فتصبح الكلمة Welcom ولكنها تظل في الملف المكتوب بلغة XML مقروءة ويمكن فهمها من سياق النص , في حين إذا نقص رمز من أحد التعليمات في ملف معالجة الكلمات MS Word فانه من العب بل من المستحيل قراءة الملف ويصبح غير قابل للعمل .

كما أن هناك ميزة كبيرة في الملف المنشأ بلغة Xml وهى أن هذه اللغة ذاتية الوصف أي أنها لا تحتاج إلى برامج أو أنظمة لقراءتها وفهمها مثل ملف PDF الذي يحتاج إلى برنامج Acrobat Reader لقراءته , بينما إذا وجد ملف مكتوب بلغة XML بعد خمسين سنة من الآن وتم فتحه وكان مكتوب فيه :

< website >
< name >
Yahoo
<name/ >
< url >
www.yahoo.com
< url/ >
< website/ >

فمن قراءة هذه العناصر الموجودة فى مصدر الملف Source File فسوف يفهم منها ببساطة أنها تتعلق بموقع ياهوو Yahoo وان عنوانه على شبكة الانترنت هو www.yahoo.com بينما هذا لا ينطبق على احد الملفات مغلقة المصدر مثل ملف word 2000 , excel 97 وغيرها حيث لا يملك المستفيد بعد مرور نصف قرن البرنامج القادر على فتح هذه الملفات وقراءتها , كما انه من الممكن إذا ما تلف أي بايت Byte من الملف فإننا لا نستطيع قراءته .

وعلي الأرشيف الوطني أن يعتمد الملفات المكتوبة بلغة XML للأسباب الآتية :

1 – تسمح الملفات المكتوبة بلغة XML بسهولة وسرعة إنشاء الأشكال الأرشيفية التي ليس لها حق الملكية الفكرية , وبالتالي لا تقع الأرشيفات تحت طائلة القانون .
2 – تسمح ملفات XML بحفظ السجلات الرقمية كما هي بدون تغيير في عناصرها الأساسية أو خصائصها الوظيفية , مما يحافظ على صحتها وتكاملها .
3 – الواصفات الخاصة بلغة XML - التيجان والرموز , متاحة مجانا فيستطيع أي شخص أن ينشئ ملف مكتوب بلغة XML بدون الاعتماد على ملكية الصانع أو منتج تكنولوجيا المعلومات .
4 – لا تتأثر الملفات المكتوبة بلغة XML بتغير التكنولوجيا وبالتالي لا تحتاج إلى إجراء عملية هجرة Migration لمحتوى الملف كل فترة مما يوفر تكلفة كبيرة للمنظمات . (Heslop,Helen&Davis,simon,2002).

وقد وضعت المنظمة الدولية التوحيد القياسي ISO ولجنة الاتصالات الدولية IEC واتحاد الاتصالات الدوليةITU والعديد من المنظمات الإقليمية والوطنية للتوحيد القياسي عدد من المعايير التي يجب عن تطيق عند إنشاء الأشكال الرقمية بصفة عامة :

1- أن تكون أشكال مفتوحة المصدر وموثقة بحيث يمكن مشاهدتها وقراءتها عبر أي برنامج أو نظام
2- أن تكون ثابتة ومستقرة بحيث يستخدم في إنشاءها واصفات معيارية (رموز وتيجان) حتى لا نضطر إلى تغيرها كل فترة .
3- أن يتم دعمها واستخدامها من قبل كل البرامج والأنظمة الآلية .

<u>وهناك عدد من المعايير التي يجب أن تتوفر في الشكل الرقمي التي يتم اختياره لأغراض الأرشفة وهى :</u>

- أن تكون موحدة ومعيارية بحيث تظل ثابتة ومستقرة ولا تعتمد

على منتج واحد .

- أن تكون منتشرة على نطاق واسع بما يكفى لإتاحتها في الأسواق .
- أن تكون قابلة للتنفيذ وتعمل على كافة التطبيقات ونظم التشغيل وبروتوكولات الشبكات .
- أن توفر هذه الأشكال الرقمية بناء جيد لتخزين المعلومات .
- أن توفر تقنية لاكتشاف الأخطاء وتصحيحها عند تخزين البيانات.
- تخزين المعلومات دون فقد أي منها . - أن يكون الشكل الرقمي سهل الاستخدام .
- إمكانية القيام بعمل حقول خاصة تستخدم في تخزين ما وراء البيانات الواصفة Metadata .
- القدرة على الاحتفاظ بالخصائص الأساسية للسجل الالكتروني على مر الزمن .

تحويل الأشكال الرقمية إلى أشكال أرشيفية تعتمد على لغة XML

يتم تحويل الأشكال الرقمية إلى أشكال أرشيفية عن طريق عملية تسمى بالتطبيع أو Normalization وهى عبارة عن تحويل الشكل الرقمي الأصلي من مصدره إلى شكل أرشيفي يعتمد على لغة XML وهى عملية تتم بطريقة آلية باستخدام برنامج يسمى ب Normalizer , وتتم عملية التحويل عن طريق نقل الشكل الرقمي الأصلي من مصدره إلى ملف مكتوب بلغة XML , وبالتالي يصبح لدينا شكل رقمي أصلى وهذا يتم نقله إلى المستودع الرقمي الرئيس . وشكل رقمي منقول ويحمل كل خصائص الشكل الرقمي الأصلي , وهذا الشكل المنقول هو ما يتم استخدامه وقراءته وعمل نسخ منه للإطلاع عليها .

وعند إجراء عملية التحويل إلى الأشكال الأرشيفية أن نتجنب ما يلي :

- لا يجوز بحال أن ننقل الأشكال الرقمية المشفرة التي لا يمكن فتحها إلا بكلمة سر Password
- لا يمكن تحويل الملفات المضغوطة إلى ملفات أرشيفية لان ضغط الملفات يؤدى في كثير من الأحيان إلى فقد معلومات من السجل الرقمي .

وهناك توصية خاصة بعدم ضغط الملفات عند الرغبة في حفظها للمدد زمنية طويلة , فعند فك هذه الملفات باستخدام احد برامج فك الملفات

المضغوطة مثل Zip Rar فإننا قد نفقد بعض البايتات مما يجعل السجلات الرقمية غير قابلة للقراءة وعدم القدرة على مشاهدتها . ونحن لا نلجأ إلى ضغط الملفات لأنه يقلل مساحتها , وإذا كان لا مفر من ضغط الملفات فعلى الارشيفى أن يختار طريقة للضغط تقلل من فقد البيانات واختيار طريقة موثقة ومعيارية .

توصيات عامة عند استخدام الأشكال الرقمية

أولا :يجب العمل على الحد من الأشكال الرقمية الكثيرة التي تستخدم في المنظمة خاصة التي لا تتناسب مع الحفظ الارشيفى
ثانيا : من الأفضل استخدام أشكال رقمية تتناسب مع الحفظ الارشيفى من بداية إنشائها في الإدارات .
ثالثا :عدم حفظ السجلات الرقمية في أشكال غير موثقة أو في ملفات مغلقة لا يمكن الدخول عليها .
رابعا تجنب طريقة ضغط الملفات مثل ZIP--TIFF / LIW—JPG6 .
خامسا :عدم التخلص من الملفات الأصلية إلا بعد إجراء عملية النقل إلى الأشكال الأرشيفية والتأكد من صحة ذلك
سادسا :التأكد من تطابق خصائص وسمات السجلات الرقمية مع المعايير الدولية المعمول بها في هذا الشأن .