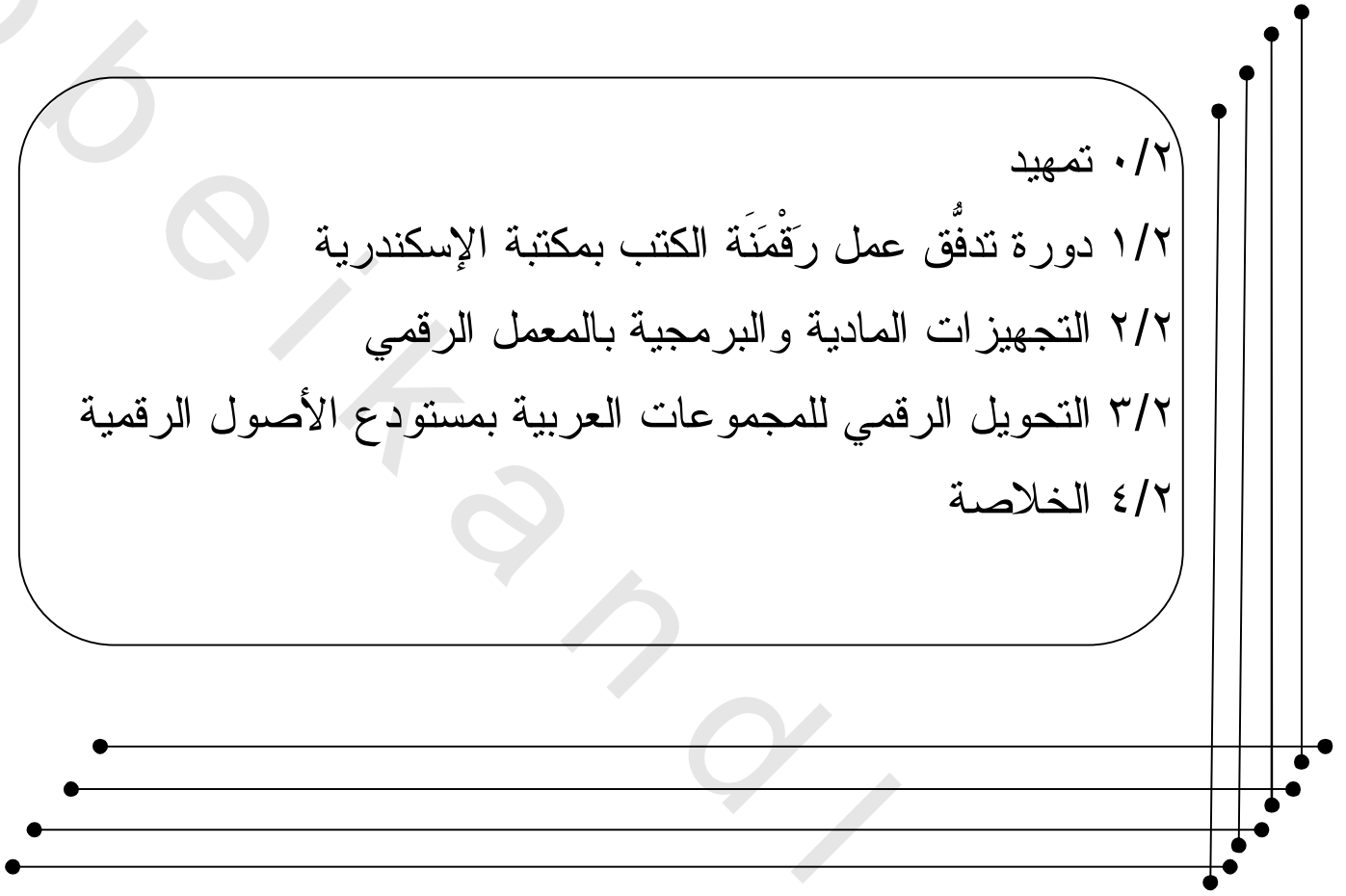
الفصل الثاني

التحويل الرقمي للكتب بمشروع مستودع الأصول الرقمية



٠/٢ تمهيد

١/٢ دورة تدفُّق عمل رَقْمَنَة الكتب بمكتبة الإسكندرية

٢/٢ التجهيزات المادية والبرمجية بالمعمل الرقمي

٣/٢ التحويل الرقمي للمجموعات العربية بمستودع الأصول الرقمية

٤/٢ الخلاصة

obeikandl.com

obeikandl.com

٢/٠ تمهيد

يتناول هذا الفصل دورة تدفُّق عملية رقمنة الكتب، بدءًا من استلام المعمل الرقمي للكتاب، إلى أن يتم نشره على شبكة الإنترنت، كما يتناول التجهيزات المادية والبرمجية بالمعمل الرقمي، مع بيان بالمشكلات والصعوبات التي تواجه عملية التحويل الرقمي للمجموعات العربية في مستودع الأصول الرقمية.

٢ /١ دورة تدفُّق عمل رَقْمَنَة الكتب بمكتبة الإسكندرية

تُعتبر عملية الرقمنة (Digitization) بمثابة البنية التحتية الفنية (Technical Infrastructure) لمشروعات التحويل الرقمي، وتشتمل تلك البنية الفنية للمشروعات الرقمية على ثلاث حلقات، وتُسَمَّى بسلسلة الرقمنة (Digitization Chain)، والتي تهدف إلى الربط بين التجهيزات المادية (مرحلة إنشاء المصدر المُرَقْمَن من الشكل المطبوع أو التناظري) والبرمجيات (مرحلة تنظيم المجموعات المُرَقْمَنَة وفقًا لتجانس محتواها أو مضمونها أو شكلها) والشبكات (مرحلة إتاحة المُخْرَجات والمنتجات الرقمية متعددة الأشكال لتلبية احتياجات المستفيدين المختلفة)، إلى جانب المعايير والبروتوكولات والسياسات والإجراءات الأخرى[١]. وسيتم توضيح تلك المراحل الثلاث بشيء من التفصيل، وذلك وفقًا لتجربة مكتبة الإسكندرية في رقمنة مجموعاتها من خلال مشروع (مستودع الأصول الرقمية).

وترتبط نماذج تدفق عمليات الرقمنة بنوعية المصادر التي يتم تحويلها، كذلك تتأثر بما إذا كانت ستتم داخل المؤسسة أم سيُعهد بها إلى شركة متخصصة، مما يعني اختلاف تدفق العمليات من مشروع لآخر[٢]. ولذلك فقد أنشأت مكتبة الإسكندرية معملاً للرقمنة، وكان يحتوي في البداية على جهاز واحد للمسح الضوئي، أما الآن فقد زاد عدد الأجهزة عن عشرة، إضافة إلى أربعة أجهزة أخرى للكتب الأكبر حجمًا، وأهم ما أُسس في المعمل هو نظام عمل متكامل workflow تتبعه عملية الرقمنة، وهي عملية تشمل الترتيب والمعالجة من العيوب الفنية، وغيرها من الإحداثيات الأخرى. وتختلف هذه الطريقة وفقًا لاختلاف المادة المراد رقمنتها، سواء كانت مخطوطًا أو كتابًا أو خريطة، وتتم بتحديد النتيجة المطلوبة، بعد إتمام العملية، ومن ثَمَّ استُخدم نظام العمل (مصنع الأصول الرقمية) الذي طوره المعمل بمفهوم البرمجيات المفتوحة المصدر (Open Source) ، ليكون استخدامه متاحًا للجميع[٣].

ويتحكم مصنع الأصول الرقمية (DAF) في دورة تدفُّق عمل رقمنة الكتب وغيرها من مصادر المعلومات، حيث يتولى الإدارة الآلية لعملية التحويل الرقمي بمراحلها المختلفة من: مسح، وتجهيز، وتعرُّف ضوئي على الحروف، ونشر، وإدارة صلاحيات المستخدمين أثناء التعامل مع محتويات المستودع، وملف الحركة ويشمل (توحيد خطة ترميز الملفات، والتحقق من صحة المجلدات والملفات الناتجة عن مراحل عملية التحويل الرقمي المختلفة)، والأرشفة، وترميز (تكويد) المواد الممسوحة ضوئيًّا استعدادًا لنشرها، وإدارة الكيانات الرقمية وحفظها واسترجاعها، وتقديم تقارير سير العمل، وإجراء الاستعلامات على الإنترنت،

بالإضافة إلى توفير المرونة اللازمة لإدارة عدد من المشروعات متنوعة المحتوى في الوقت نفسه، والتي تشمل الكتب، والمجلات، والصحف، والمخطوطات، والفيديو، والشرائح[٤].

يوفِّر دار مجموعة من الأدوات اللازمة لإدخال الميتاداتا، وتتبُّع سير عملية الرقمنة بكل مراحلها، والتحقق من مطابقة المواد الرقمية لأصولها؛ فمن الممكن أن تحدث اختلافات ما بين المواد الأصلية وصورها المرقمنة نتيجة لعملية تحويلها رقميًّا، وأرشفة المواد الرقمية بغرض حفظها واسترجاعها على المدى الطويل، وتناسب تلك الأدوات الفئات الثلاثة المستخدمة له، وهي: فئة مسئولي الرقمنة، واختصاصي المكتبات -الذين ينقسمون بدورهم إلى مفهرسين ومراجعين-، والمستخدم النهائي للنظام، وتتمثل هذه الأدوات فيما يلي:

أولاً- أداة الإدارة (Administration Tool)

هي أداة يتم استخدامها من قِبَل مصنع الأصول الرقمية لإنشاء الميتاداتا الوصفية لمصدر المعلومات المطلوب رقمنته، فإذا كان قد سبق فهرسته داخل النظام الآلي للمكتبة LIS، يتم استخدام رقم الباركود نفسه (الترميز العمودي) الخاص به للربط بين تسجيلته في مستودع دار وفهرس المكتبة، وإذا لم يُفهرس من قبل فإنها توفِّر الحد الأدنى من الميتاداتا التي تساعد في الوصول إليه وفي إنشاء معرِّف فريد لكل من المجلدات المستخرجة، كما تعمل هذه الأداة على تحديد معايير استخدام النظام وإنشاء تقارير معدلات الإنتاج.

ثانيًا- عميل الرقمنة (Digitization Client)

هي أداة يستخدمها مصنع الأصول الرقمية، ويتمثل دورها في إنشاء مجموعة المجلدات المنظمة لعملية الرقمنة، كما تُستخدم بعد الانتهاء من كل مرحلة من مراحل عملية الرقمنة، في: التحقق من صحة الملفات والتنبيه على وجود أخطاء، وتحديث قاعدة البيانات بتحديد المسئولين عن العمل، وتاريخ الانتهاء، وعدد المجلدات والملفات المستخرجة، ونقل المجلدات والملفات بعد تشفيرها تجنُّبًا لأي أخطاء أثناء عملية النقل ووضعها على قائمة انتظار المرحلة الخاصة بها في الخادم.

وينحصر دورها في كل مرحلة من المراحل الأربعة الرئيسة للرقمنة وهي: المسح الضوئي، والمعالجة، والتعرف الضوئي على الحروف، وإعادة المعالجة، فيما يلي:

{١} المسح الضوئي (Scanning)

ويشمل الخطوات التالية:

أ- تسليم مصدر المعلومات إلى المعمل الرقمي للقيام بعملية الرقمنة.

ب- وضع مصدر المعلومات في قائمة الانتظار لإجراء عملية المسح الضوئي له.

ج- استدعاء مصدر المعلومات من قائمة الانتظار.

د- إنشاء مجلد خاص بمصدر المعلومات لتخزين ملفاته الممسوحة ضوئيًّا به، بحيث يحتوي المجلد على ثلاثة مجلدات فرعية، بكل من هذه المجلدات شكل من أشكال الملفات التالية: نسخة المسح الضوئي الأصلية، والنسخة المعالجة، والنسخة المُكَوَّدة، وتختلف بِنْيَة المجلد ودرجة وضوح المسح الضوئي تبعًا لنوعية مصدر المعلومات سواء أكان نصًّا، أم صورة، أم تسجيلاً صوتيًّا، أم فيديو.

هـ- وضع المجلدات في قائمة انتظار المعالجة.

{٢} المعالجة (Processing)

أ- استدعاء المجلدات من قائمة انتظار المعالجة.

ب- استخدام مزيج من أدوات معالجة الصور الآلية واليدوية لتحسين نوعية الصور الممسوحة ضوئيًّا.

ج- نقل المجلدات لتوضع في قائمة انتظار التعرف الضوئي على الحروف للمواد النصية، وقائمة انتظار الحفظ الرقمي للأنواع الأخرى.

{٣} التعرف الضوئي على الحروف (Optical Character Recognition)

أ- استخراج النص من الصور الممسوحة ضوئيًّا لتمكين عملية التعرف الضوئي على الحروف، والتي تعمل على السماح ببحث النص.

ب- استخدام برنامج Fine Reader 6.0 المنتج بواسطة شركة ABBYY للتعرف الضوئي على الحروف اللاتينية، وبرنامج القارئ الآلي صخر للتعرف الضوئي على الحروف العربية Sakhr Automatic Reader، هذا وقد بَنَتْ مكتبة الإسكندرية مكتبة من الخطوط باستخدام العينات المأخوذة من المواد المختلفة قبل البدء في إجراء عملية التعرف.

{٤} إعادة المعالجة (Reprocessing)

يدعم النظام دورة تدفُّق خاصة بإعادة معالجة الكيانات المرقمنة، وتُستخدم هذه العملية في حالة الحاجة إلى تحسين جودة التعرف الضوئي على الحروف، أو تطبيق إجراءات جديدة لمعالجة الصور، أو لإنشاء صيغة جديدة قابلة للنشر من الكيان المُرَقْمَن، وتتكون هذه العملية مما يلي:

أ- بحث واسترجاع الملفات المراد معالجتها من الأرشيف.

ب- وضع الملفات المراد معالجتها في قائمة الانتظار لتأخذ دورها من جديد في عملية الرقمنة.

ج- إعادة الأرشفة للملفات التي تم تعديلها أو تغييرها، ويتم الكشف عن تلك التغييرات باستخدام برنامج الفحص مثل (Checksums) ، وهو برنامج يُستخدم لفحص التغييرات، ويتم حسابها قبل وبعد عملية إعادة المعالجة.

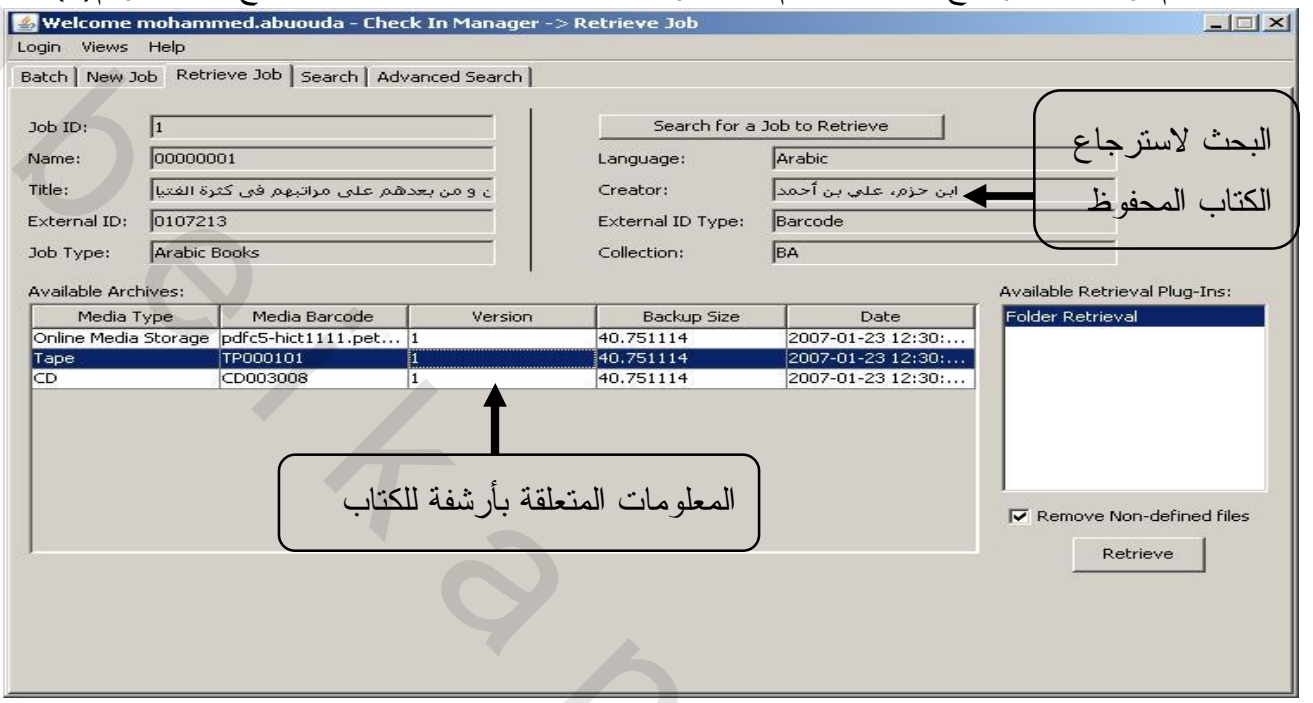
د- تسجيل معلومات الأرشفة للملف الجديد في قاعدة بيانات المستودع بعد ربطها بالإصدارة الأقدم منه للحفاظ على العلاقة بينهما.

ثالثًا- أداة الأرشفة (Archiving Tool)

يتم فيها تمثيل الكيان المرقمن بملف واحد أو أكثر بعدد مختلف من الأشكال و/أو درجات الوضوح، وتخزينه بطريقتين: الأولى على وسائط تخزين خارجية للحفظ طويل المدى باستخدام وسيطين هما: الأقراص المُلَيْزَرَة Cds، و/أو الشرائط Tapes، بعد وضع معرِّف فريد يُطبع ويُلصق عليهما لضمان الوصول إليه مستقبلاً، وعادة يتم الحفظ للأرشيف الأساس (النسخة الأصلية الممسوحة ضوئيًّا والإصدارات المعالجة

بدرجات وضوح عالية)، والثانية في نظام تخزين RAID storage system* على الخط المباشر لأغراض النشر، وذلك للنسخ المشتقة الأقل جودة (low resolution) بصيغ DJVU ، PDF ، JPG ، ويقوم النظام بالربط بين الإصدارات الأحدث والأقدم من الملف(٥).

ولدعم أرشفة واسترجاع البيانات، تقوم أداة الأرشفة بالوظائف التالية كما هو موضَّح بالشكل رقم(٣):



شكل رقم (٣) استرجاع الكتاب من خلال الباركود الخاص به

ووفقًا للشكل رقم (٣) تتمثل وظائف أداة الأرشفة في الخطوات التالية:

١- فحص الملفات والمجلدات بشكل ثابت ومطرد.

٢- إعداد مجلدات للأرشفة، وذلك من خلال ضغط المجلدات الفرعية والملفات وتجميعها في حِزَم تتلاءم مع السعة التخزينية للوسيط (CD or tape)، وإعداد عنوان يُطبع ويُلصق على الوسيط، وربط المجلدات المحفوظة بالوسيط في قاعدة البيانات.

---

* (RAID) المجموعات الوافرة من الأقراص المستقلة: اختصارًا لـ **Redundant Array of Independent Disks** وهي تكنولوجيا أنشاها مجموعة باحثين من جامعة باركلي بكاليفورنيا سنة ١٩٨٧ والذين درسوا إمكانية تعرف النظام على أكثر من قرص صلب كوحدة واحدة من طرف النظام، وهي تقنية تسمح بتخزين البيانات في عدة أقراص صلبة، وبالتالي تطوير فعالية وحدات التخزين على الخوادم، وعليه فهي تقدِّم إمكانيات ومزايا مهمة فيما يتعلق بالفعالية (سرعة تدفق البيانات)، والتكلفة، وخاصة تجاوز أعطاب نظام التخزين؛ ففي حالة حدوث عطب فإنه يمكن استرجاع البيانات -وحتى تغيير القرص المعطوب- دون فَقْد للبيانات، ودون توقيف الشبكة. للاطلاع أكثر على موضوع RAID، يمكن زيارة الرابط التالي:
http://www.clubic.com/article-13989-1-initiation-raid.html

٣- إعداد ملف بلغة الترميز القابلة للامتداد (XML)* للمجلد المحفوظ بالأرشيف، ويشتمل هذا الملف على جميع الميتاداتا المتعلقة بالتسجيلة المطابقة لمصدر المعلومات في قاعدة بيانات المستودع بلغة XML، ويتم حفظ ملف XML مع المجلد الذي يقوم بوصفه.

٤- إعداد اختبارية (Checksums) للملفات المحفوظة للكشف عن التغييرات التي طرأت عليها، وذلك في حالة تحميل او إعادة معالجة الملف.

٥- تتيح أداة البحث لمسئولي الرقمنة إمكانية استرجاع المجلد المحفوظ عن طريق تحديد مكانه، واسترجاع المجلدات الفرعية والملفات غير المضغوطة، ونسخ تلك الملفات والمجلدات غير المضغوطة إلى المكان الذي يحدده مسئولو المعمل الرقمي.

٦- تقوم الأداة بعمل تحذير عندما يتجاوز مستوى التخزين القيمة المحددة سلفاً لكل وسيط.

٧- تحديث قاعدة بيانات المستودع بتسجيل معلومات الأرشفة للملف الرقمي[٦].

رابعًا- أداة الترميز (التكويد) (Encoding Tool)

يتم فيها إنشاء كيان رقمي نهائي للنشر على الإنترنت، وذلك بإنشاء نسخ JPG بدرجات وضوح وجودة مختلفة مشتقة من النسخة الأصلية لتُناسِب السرعات المختلفة لاتصالات الإنترنت.

خامسًا- أداة الفهرسة (Cataloging Tool)

هي أداة تطبيقية على شبكة الإنترنت يتم استخدامها لإضافة وتحرير الميتاداتا في حافظة الأصول الرقمية (DAK)، مما يساعد في الوصول إلى التسجيلات التي أُنشِئَت في مرحلة الرقمنة، بالإضافة إلى إنشاء تسجيلات جديدة للكيانات المرقمنة والتحميل الصاعد للملفات المرتبطة بها، كما يتم مسبقًا تحميل المستودع بقوائم المفردات المقيدة مثل: قائمة رؤوس موضوعات مكتبة الكونجرس (LCSH) لرؤوس الموضوعات اللاتينية، وقائمة رؤوس الموضوعات العربية المحلية (QRMAK) للموضوعات العربية، وتسمح أداة الفهرسة بتصميم قوالب لأنواع محددة من المواد كالكتب والخرائط والشرائح، واستيراد بيانات الميتاداتا من المصادر الخارجية، والاستخراج الآلي لميتاداتا المحتوى الرقمي.

---

* تُعرف لغة XML بأنها: مجموعة محددة مسبقا من الواصفات سواء كانت (رموز و تيجان) أو طريقة لتعيين وتحديد تلك الواصفات والتي تستخدم في إدماج وتضمين أية معلومات خارجية داخل وثيقة نصية إلكترونية و عادة ما تشير إلى تحديد شكل تلك المعلومات أو لتيسير إجراءات التحليل التي تتم عليها. لمزيد من التفصيل، راجع: هبة عبد الستار مصيلحي . XML :هل تغير مستقبل المكتبات الرقمية .- cybrarians journal .-ع ١ (يونيو ٢٠٠٤) . - تاريخ الإتاحة [٢٠٠٩/١/٦] .-متاح في: <:www.cybrarians.info/journal/no1/xml.htm>

سادسًا- واجهة النشر (Publishing Interface)

هي واجهة ويب ذات صلة بحافظة الأصول الرقمية (DAK) توفر الوصول إلى الكيانات الرقمية المخزنة بالمستودع من خلال إمكانيات البحث والتصفح في:

١- محتويات المستودع بواسطة المجموعة، ونوع العمل، والموضوع، والمنشئ، والعنوان.

٢- المحتوى عن طريق حقل الميتاداتا المفهرس كالمؤلف، والعنوان، والموضوع، والكلمات المفتاحية.. إلخ.

٣- النص الكامل.

٤- مستويات مختلفة للعرض (التكبير) متاحة بالنسبة لبحث الصور.

٥- عرض موجز ومفصَّل للتسجيلات مع روابط إلى الكيانات الرقمية.

٦- عرض التسجيلات بصيغ MARC, DC or XML.

سابعًا: التكامل مع النظام الآلي للمكتبة (Integration with the ILS)

يتكامل مستودع الأصول الرقمية (دار) مع المصادر الخارجية مثل الفهرس الببليوجرافي، والمستودعات الخارجية، ونُظُم عرض الصور، كما يسمح بالتصدير والاستيراد منه وإلى المستودعات الأخرى، ويساعد هذا التكامل في إلغاء الحاجة إلى تكرار إدخال البيانات[٧].

وتتكامل النسخة الثالثة من دار مع النظام الآلي لمكتبة الإسكندرية Virtua Integrated Library System، وقد أُعِدَّت التسجيلات الببليوجرافية بفهرس المكتبة وفقًا لمارك٢١ والتزمت التسجيلات بالمواصفة التي أصدرتها المنظمة الدولية للتقييس ISO 2709 التي تُستخدم لنقل البيانات بين المكتبات؛ لتكون بمثابة المعيار الموحد للتجانس بين مختلف صيغ مارك، ولتصبح مقروءة من قبل النظم الآلية كافة مهما تعدَّدَت[٨].

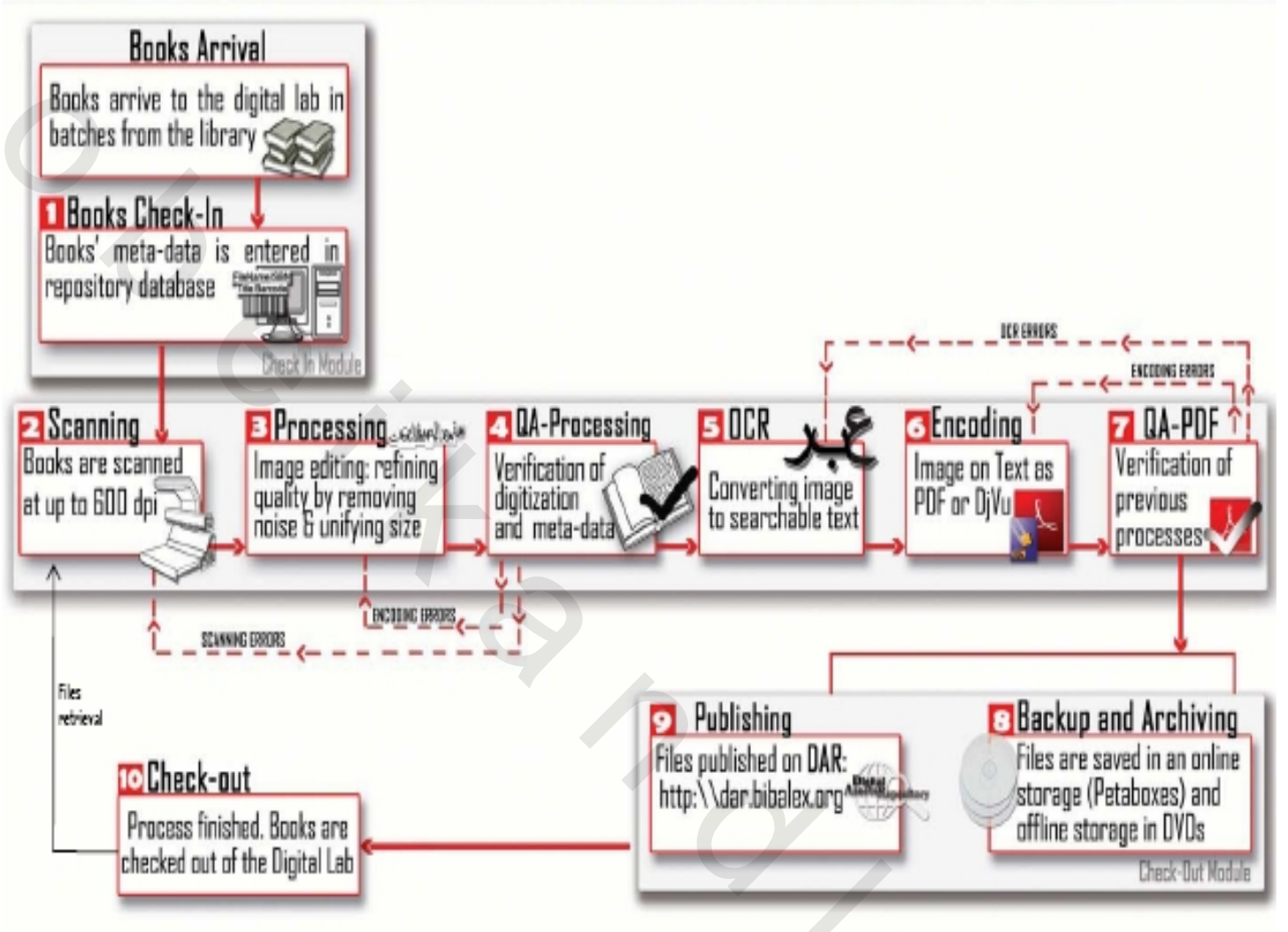
ثامنًا:التحقق من هوية المستخدم والترخيص بالاستخدام (Authentication and Authorization)

يُستخدم لتحديد تراخيص ومعرفات الوصول، مثل: كلمة السر، واسم المستفيد، كما يسمح النظام للمستفيد غير المسجَّل بالدخول والقيام بعمليات البحث.

تاسعًا: بوابة مبادرة الأرشيفات المفتوحة (OAI Gateway)

تُطَبِّق هذه البوابة بروتوكول الأرشيفات مفتوحة المصدر الموضوعة من قِبَل مبادرة الأرشيف المفتوح، والتي تعمل على توفير الوصول إلى محتويات المستودع عبر بِنْيَة منظَّمة، حيث تقوم بتَلَقِّي الطلب في صيغة XML وتقوم بترجمته في شكل استعلام بقاعدة البيانات المماثلة، ثم تسترجع النتائج وتترجمها مرة أخرى إلى صيغة XML وترسلها في صورة استجابة نهائية للطلب وتستجيب البوابة لستة أنواع من الطلبات اللازمة للامتثال لبروتوكول تجميع الميتاداتا التابع لمبادرة الأرشيفات المفتوح Open Archives Initiative-Protocol for Metadata Harvesting (OAI- PMH)، وهي: المعرِّف (Identify)، وقائمة تنسيقات الميتاداتا (List Metadata Formats)، وقائمة المجموعات (List Sets)، وقائمة المعرِّفات (List Identifiers)، وقائمة السجلات (List Records)، والحصول على سجل (Get Record)[٩].

- ويوضح الشكل رقم (٤) بالتفصيل دورة تدفق عمل رقمنة الكتب التي تقوم بها مكتبة الإسكندرية منذ أن يستلم المعمل الرقمي الكتاب إلى أن يتم نشره على شبكة الإنترنت.



شكل رقم (٤) دورة تدفق عمل رقمنة الكتب بمكتبة الإسكندرية
BA Digitization Workflow

وبتتبُّع الخطوات من البداية سنجد أن عملية رقمنة الكتب تتم وفقًا لما يلي:

(١)- يستلم المعمل الرقمي الكتب، ويتم إدخال الميتاداتا للكتب في قاعدة بيانات المعمل الرقمي.

وتُعَد مرحلة وضع الميتاداتا للكيانات الرقمية من أهم مراحل المشروع؛ فبدون ميتاداتا (تحدد مصادر وأماكن وجود ومحتويات وتفاصيل إنشاء الكتب المرقمنة، وتسهِّل استكشافها وتوفر محددات رقمية تساعد على التمييز بين كتاب مرقمن وآخر وتعمل على ضمان إتاحته مستقبلاً، وتوثِّيقه وإتباعه بمعلومات عن مستويات حقوق النشر والاستنساخ) يصبح المشروع بلا فائدة، كما أنها تساعد على إمكانية التشغيل البيني الذي يسمح بتبادل البيانات بصرف النظر عن اختلاف الأجهزة أو بيئة البرمجيات أو واجهات الاستخدام[١٠].

ويدعم المشروع معايير الميتاداتا التالية:

أولاً: الميتاداتا الوصفية للكيانات الرقمية (Digital objects descriptive metadata): يعتمد نموذج البيانات الخاص بوصف مقتنيات المكتبة على نظامَي:

- § معيار الفئات البؤرية للمصادر المرئية (VRA Core Categories) لوصف الصور والوسائط المتعددة (المواد المرئية).
- § معيار MARC 21 لوصف المواد النصية.

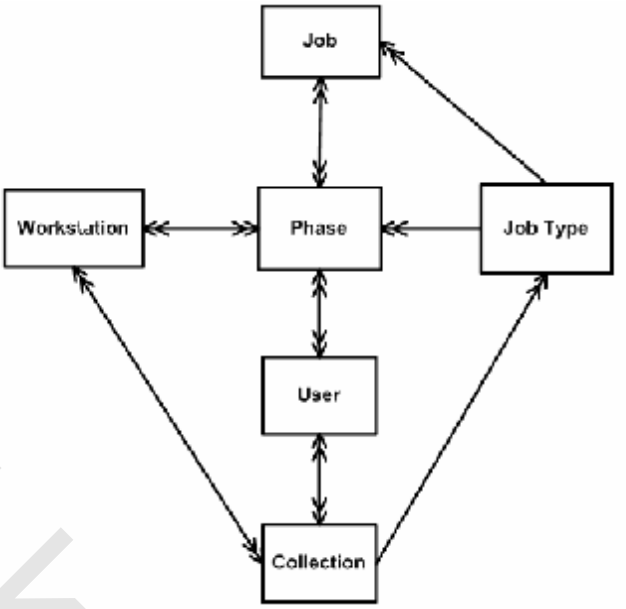
وتم استخدام معياري MODS و METS بالإصدارة الثالثة من دار DAR 3.0، فبالنسبة لمخطط وصف كائن الميتاداتا Metadata Object Description Schema(MODS) فهو مخطط ميتاداتا وصفي مشتق من معيار مارك ٢١ ويهدف إلى القيام بتعيين البيانات من واقع التسجيلات الببليوجرافية المبنية على أساس شكل اتصال مارك فى طبعته الحادية والعشرين، وبأسلوب يماثل ذلك الذى يتم على أساسه إعداد تسجيلات الوصف الخاصة بالمصادر الأصلية، وتُستخدم اللغة بدلاً من الأرقام المستخدَمة في حقول مارك ٢١، وتُستخدم لغة الـ XML في التعبير عن هذا المخطط، وتُعتبر عناصر الـ MODS أكثر ثراءً من عناصر الدبلن كور، كما أن عناصره أكثر توافقًا مع البيانات الببليوجرافية المكتبية بالمقارنة مع الدبلن كور، كما انه أبسط في التطبيق من شكل مارك ٢١ الببليوجرافي، وباستخدامه للغة الـ XML فإن الـ MODS يوفر تقنيات وإضافات أكثر من تلك المُعَدَّة بمارك، وقد ساهمت مميزاته العديدة في استخدامه في إعداد التسجيلات الوسيطة في العديد من عمليات استيراد وتصدير البيانات الببليوجرافية(١١).

أما بالنسبة لمعيار تحويل وتكويد الميتاداتا (METS) Metadata Encoding and Transmission Standard، فأنشئ هذا المعيار لتلبية الحاجة إلى البنية المعيارية للبيانات لوصف مواد المكتبة الرقمية المعقدة، والـ METS عبارة عن مخطط تم بناؤه على لغة الـ XML، ويهدف إلى إنشاء وثائق XML تُعَبِّر عن بنية الكيانات الموجودة بالمكتبات الرقمية، والميتاداتا الوصفية والإدارية المرتبطة بهذه الكيانات، بالإضافة إلى أسماء ومواقع الملفات التي تتكون منها هذه الكيانات الرقمية.

وتُعَد الميتاداتا ضرورية لإدارة ناجحة واستخدام مواد رقمية أكثر تكثيفًا ومختلفة عن الميتاداتا المستخدمة لإدارة مجموعة الأعمال المطبوعة والمواد المحسوسة الأخرى، ونحتاج إلى الميتاداتا البنائية لتأكيد أن الملفات المرقمنة مستقلة (مثل صفحات مختلفة من كتاب رقمي) ومبنية ومهيكَلة بالشكل الملائم الصحيح، ونحتاج إلى الميتاداتا الفنية لنحصل على معلومات عن عملية الرقمنة؛ وذلك لتقرير وتحديد مدى دقة انعكاس الأصل الذي توفره النسخة الرقمية، كما أن هناك حاجة للميتاداتا الفنية الأخرى لأغراض وعمليات داخلية؛ وذلك من أجل إنعاش وتجديد وترحيل البيانات، مما يؤكد مدى قدرة المصادر القيمة على التحمل(١٢).

ولقد تم التخطيط لاستخدام معياري MODS و METS بالأخص للتكامل مع نظم المستودعات المستقبلية ومع مستودع (Fedora) Flexible Extensible Digital Object Repository Architecture.

ويمكن من خلال نموذج البيانات وصف البيانات المرئية والنصية في إطار متجانس، ويوضح الشكل رقم (٥) نموذج البيانات الخاص بوصف الكيانات الرقمية.



شكل رقم (٥) نموذج بيانات وصف الكيانات الرقمية

ويتكون نموذج البيانات من ستة أنواع من الكيانات (Entities) المرتبطة بإدارة تدفق عملية الرقمنة، وهي بالترتيب التالي:

١- العمل (Job): مصدر المعلومات المراد رقمنته، سواء كان كتابًا أو صورة رقمية.

٢- نوع العمل (Job type): أشكال مصادر المعلومات المختلفة كالكتب، والصور، والخرائط، والمخطوطات... إلخ.

٣- مراحل عملية الرقمنة (Phase): وتتمثل المراحل المختلفة لعملية الرقمنة في المسح الضوئي، والمعالجة، والتعرف الضوئي على الحروف، والتكويد، والنشر على شبكة الإنترنت، وتأمين الجودة، والأرشفة الرقمية.

٤- المستخدم (User): وتتمثل في الفئات المستخدمة للنظام، وهم: مسئولو الرقمنة، واختصاصي المكتبات الذين ينقسمون بدورهم إلى مفهرِسين ومراجِعين، ومدير للنظام.

٥- المجموعة الرقمية (Collection): وهي تجميع منطقي لمجموعة من المواد (الكيانات) الرقمية، ومن أمثلتها: مجموعة جمال عبد الناصر، وأنور السادات... إلخ.

٦- محطة العمل (Workstation): وتتمثل في أجهزة الحاسب الآلي المستخدمة لأداء المراحل المختلفة لعملية الرقمنة[١٣].

ويتم تخصيص الميتاداتا الوصفية للكيانات الرقمية إما يدويًّا بواسطة اختصاصي المكتبات وإما يتم استرجاعها من النظام الآلي للمكتبة، فيتم إدخال الميتاداتا للكتب من خلال واجهة تعامل متاحة على الويب مرتبطة بالنظام الآلي للمكتبة (Virtua Integrated Library System) .

ثانيًا: ميتاداتا العرض (Metadata Presentation): في بداية المشروع كان يتم عرض الميتاداتا في الأشكال التالية:

- لغة الترميز القابلة للامتداد XML.
- شكل مارك MARC format.
- معيار دبلن كور Dublin Core* (١٤).

ثم في المراحل التالية من تطوير المشروع، وعلى الرغم من أن النظام الآلي المستخدَم في مكتبة الإسكندرية يتيح عرض الميتاداتا الخاصة بالكتاب -والمتمثلة في لغة XML، أو صيغة مارك، أو دبلن كور- فإن المسئولين قرروا غلق هذا الجزء من النظام؛ للحفاظ على حقوق الملكية الفكرية للمكتبة، وبالتالي لا يجوز للمستفيد الاطلاع عليها.

ويستخدم DAR ستة أنواع من الميتاداتا لوصف الأصول الرقمية:

(أ) الميتاداتا الوصفية (Descriptive Metadata): وتشمل البيانات الخاصة بوصف مصدر المعلومات، مثل: نوع الوعاء (كتاب، خريطة، دورية، تسجيل صوتي، فيديو... إلخ)، والعنوان، والمنشئ، والتاريخ، هذا بالإضافة إلى الكلمات الدالة، وبيانات الوصف المادي، والأبعاد، وغيرها من الميتاداتا التي تُحدَّد وفقًا لنوع المصنف، وتشمل حقولاً مثل (الترقيم الدولي الموحد للكتب ISBN، واللغة، والناشر -وذلك بالنسبة للكتب-، والطريقة Technique، والمادة Material وذلك بالنسبة لمصنف فني وصور).

(ب) ميتاداتا المحتوى الرقمي (Digital Content Metadata): وتحتوي على وصف الكيان الرقمي من حيث دقة المسح، وأبعاد الصورة، ومدة الفيديو؛ حيث يَدْعَم المستودع أشكالاً مختلفة من الملفات مثل: JPG, TIFF, JPG 2000, PDF, DJVU, OCR Text وتجدر الإشارة إلى مرونة دخول أشكال ملفات جديدة للنظام مع توافر الأدوات الملائمة للتعامل معها.

(ج) ميتاداتا الحفظ (Archiving Metadata): وتتضمن: (مكان حفظ الكيان الرقمي -حيث يُحفظ الكيان على وسائط تخزين كالأقراص الضوئية والشرائط المغناطيسية-، ورقم المعرف الفريد للوعاء -والذي من الممكن أن يُلحق به-، كما يحدد موقع مصدر المعلومات داخل المؤسسة التي تملكه).

(د) ميتاداتا النشر (Publishing Metadata): وتشمل [تكويد الكيان الرقمي استعدادًا للنشر، ومسار الملف على الخادم، وتاريخ النشر، وتصنيف الفئات المستهدفة (طلبة، وباحثين)] وتحفظ الكيانات الرقمية المُكَوَّدَة للنشر على الخط المباشر online.

(ه) ميتاداتا حق الوصول (Access Right Metadata): وتضم قيود وصلاحيات ومستويات الوصول إلى محتويات المستودع -سواء بالعرض الكامل للوعاء ككل أو لأجزاء منه فقط-.

---

* معيار دبلن كور هو مجموعة مبسطة من عناصر ما وراء البيانات المستخدمة في المكتبات الرقمية، وتتمثل وظيفتها الرئيسة في وصف الكيانات الرقمية لأغراض إدارة المجموعات، وتبادل البيانات.

obeikandi.com

وتدعم النسخة الثالثة من مستودع الأصول الرقمية تطبيق أربعة مستويات من حق الوصول إلى محتويات المستودع على النحو التالي:

- المستوى الأول: عرض الميتاداتا فقط.

- المستوى الثاني: عرض الميتاداتا ومصغرات Thumbnail ويُستخدم هذا المستوى للكيانات غير النصية كالصور التي تمتلك حق المؤلف على نسخ ذات درجة وضوح عالية.

- المستوى الثالث: عرض الميتاداتا ومقتطفات من المحتوى، فبالنسبة للكتب التي تخضع لحق المؤلف يتم عرض الملخص أو بعض الصفحات المختارة من الكتاب المُرَقْمَن.

- المستوى الرابع: عرض الميتاداتا وإتاحة الوصول الكامل للمحتوى، وذلك بالنسبة للكتب التي دخلت في الملك العام، فلم تَعُد يطبَّق عليها قيود قانون حق المؤلف. وقد يضاف المزيد من القيود على المستويات في المستقبل؛ لاستيعاب أنواع أخرى من المواد.

(و) ميتاداتا التحقق من هوية المستخدم والترخيص بالاستخدام Authentication and Authorization Metadata وتشمل:

- تحديد بيانات المستفيد User Identification Data مثل اسم المستخدم وكلمة السر الخاصة به كفرد أو مجموعة.
- التراخيص Permission وتتحدد بمستوى الكيان الرقمي الواحد، أو مستوى مجموعة من الكيانات الرقمية المختلفة ككل[١٥] .

- ويوضح الشكل رقم (٦) نموذج عرض للميتاداتا المستخدَمة لوصف كتاب في مصنع الأصول الرقمية:

شكل رقم (٦) نموذج عرض للميتاداتا المستخدَمة لوصف كتاب في مصنع الأصول الرقمية

وفيما يلي شرح لجميع العناصر المُدْرَجة بالشكل رقم (٦):

(أ) مُعَرِّف العمل (Job ID): مُعَرِّف فريد للكتاب.

تُشير كلمة عمل (Job) إلى أشكال مختلفة من مصادر المعلومات، مثل الكتب والخرائط والمخطوطات وغيرها من المواد التي ستخضع لعملية الرقمنة، بدءًا من المسح الرقمي وانتهاءً بالنسخ الاحتياطي، ويُقصد بالعمل في المثال الحالي [كتاب].

(ب) عنوان العمل (Job Title): عنوان الكتاب.

(ج)تاريخ الاستحقاق (Due Date): يوضح هذا الجزء التاريخ المتوقع فيه إنهاء المهمة المحددة ومن الممكن تطبيق تعديل على التاريخ المُدْخَل.

(د)معلومات (Info): يتم تخصيص ثلاثة حقول للمعلومات العامة يمكن تسجيلها عن الكتاب كعدد المجلدات أو أسماء المترجمين، أو أية معلومات عامة أخرى يُمكن إضافتها من جانب المفهرِس.

(ه) اللغة (Language): يُستخدم هذا الحقل لتحديد لغة الكتاب، ويتم اختيار اللغة من خلال القائمة المُنْسَدِلَة (drop-down menu) وهي الموجودة في مربعات الاختيار، والتي تتيح للمستخدم الاختيار من بين عدة خيارات موجودة بها.

(و)الأولوية (Priority): يُستخدم هذا الحقل لتحديد مستوى الأولوية المعتمد لإنجاز المهمة، والذي يتراوح بين ١ إلى ١٠٠، وكلما زادت القيمة المدرجة دلت على أهمية إنجاز العمل.

(ز)البيانات الكبيرة جدًّا (Large Object Binary (Lob: يُستخدم هذا الحقل لحفظ البيانات الكبيرة جدًّا، حيث يشتمل على بيانات الميتاداتا الموجودة كلها على النظام الآلي المتكامل للمكتبة، بما فيها تسجيلة مارك للكتاب، ودبلن كور، MODS XML أو أي نوع آخر من الميتاداتا التي يحتاج المستفيد أن تُحفظ مع (الكيان الرقمي)، والتي يقتضي استخدامها في مرحلة النشر. ولذلك يُستخدم هذا الجزء لتعويض أي قصور قد يحدث نتيجة لعدم كفاية حقول الميتاداتا.

(ح) حزمة من الكتب (Batch): وتتمثل في دفعات الكتب الجديدة التي يستلمها المعمل الرقمي، ويتألف اسمها من اسم المجموعة الرقمية التي ينتمي إليها الكتاب، بالإضافة إلى اختيار اسم مناسب لحزمة الكتب.

(ط) الحاسب الخادم (Server): يسمح نظام إدارة تدفُّق عمل الرقمنة (Digitization Workflow Management System) بحيازة أشكال مصادر المعلومات المتنوعة على عدة حاسبات خادمة، ولكن ينبغي أن يكون كل شكل من مصادر المعلومات كالكتب كاملاً على خادم واحد وغير قابل للتقسيم.

(ي) حالة العمل (Job condition): يمكن اختيار الحالة المادية للكتاب التي تتراوح بين (سيئة – جيدة – ممتازة) من خلال القائمة المنسدلة، فإنه يستلزم دراسة وفحصًا مسبقًا للنصوص من ناحية حالة الورق وأشكال وأحجام وأنواع الحروف المستخدمة قبل البدء في عملية الرقمنة.

(ك) نوع العمل (Job type): يشير إلى الأشكال المتنوعة من مصادر المعلومات، ويتم تحديد هذا البيان مسبقًا ولا يمكن تعديله في هذا القسم.

(ل)مُعرفات خارجية (External Ids): يعرض هذا الجزء المعرفات الخارجية المرتبطة بالعمل.

فعند الضغط على (View External IDs) يتم عرض المعرفات الخارجية مثل: رقم الباركود، والرقم الببليوجرافى Bib ID فعند تسجيل الوعاء يأخذ رقمًا تلقائيًّا يُسَمَّى الرقم الببليوجرافي، وذلك يمنع إضافة معرف جديد لعمل موجود بالفعل(١٦).

(٢)- يتم المسح الضوئي للكتب عن طريق أجهزة المسح الضوئي لتتحول في شكل صورة رقمية، وفي هذه المرحلة يتم إنشاء الملفات الرقمية من المواد التناظرية.

(٣)- تنتج عن عملية المسح الضوئي في الخطوة السابقة صور رقمية على الحاسب الآلي المتصل بوحدة المسح الضوئي، ومن الممكن إجراء أية معالجات على هذه الصور من أجل تنقية الصور وتعديل الصور، مثل تغيير حجم الملف المنقول إلى الحجم المناسب، ثم إزالة أية فراغات زائدة، ومراعاة حجم الهوامش مع تعديل الزوايا وضبط التباين والسطوع، وهذا يتم من خلال برنامج تحرير الصور، وقد تم استخدام برنامج Scanfix لمعالجة الصور الممسوحة ضوئيًّا؛ وذلك لتحسين تلك الصور لخدمة أغراض البحث والدراسة.

(٤)- تأمين الجودة*، وتتضمن الآتي: التأكد من أن جميع المصادر تم مسحها، واختبار ومقارنة الملفات بالأصول، وفحص أسماء الملفات وأحجامها وأنواعها ودرجة وضوحها، وتصحيح أخطاء المسح الضوئي، وكذلك أخطاء المعالجة التي تم إجراؤها على الصور الرقمية، ومن هذا المنطلق فإن المعايير التي يعتمد عليها قياس الجودة في عملية الرقمنة تشمل:

(أ) تطابُق الصفحات الرقمية مع النسخة الأصلية من حيث العدد، والترتيب، والأبعاد.

(ب) تطابُق النسخة الرقمية مع النسخة الأصلية من حيث وضوح المحتوى -فلا يجب أن تكون الصورة باهتة أو داكنة عن النسخة الأصلية- مما قد يؤثِّر سلبًا على إمكانية قراءة النسخة الرقمية لكل من القارئ أو برنامج التعرف الضوئي لاستخراج النص.

(ج) من حيث الشكل تجب مراعاة عدم وجود ميل أو دوران أو قص بالصور الرقمية بعد عملية معالجة الصور.

---

* ينبغي التفرقة بين مصطلح ضبط الجودة Quality Control (QC) ومصطلح تأمين الجودة Quality Assurance (QA) وخاصة أنه يتم استخدامهما بشكل تبادلي، وتتضمن عملية ضبط الجودة الإجراءات والممارسات التي يتم وضعها لضمان الاتساق والثبات والسلامة والاعتمادية لعملية الرقمنة، أما تأمين الجودة فيشير إلى الإجراءات التي عن طريقها يتم فحص جودة المنتج النهائي.

obeikandl.com

(د) بالنسبة لعملية التعرف الضوئي على النصوص؛ يُراعى مراجعة دِقَّة العملية بحيث يمكن التعرف على أكبر قدر ممكن من نص الكتاب، مما يخدم القارئ في البحث أو استنساخ مقاطع من الكتاب. ولتحقيق ذلك يُتبع الآتي:

- Ø يتم استخدام برامج التعرف الضوئي على عينة من الكتاب.
- Ø يتم عمل حصر للأخطاء الخاصة بالتعرف، إذا زادت النسبة عن حد معين يتم محاولة تلقين البرنامج على التعرف على الخط المستخدَم في هذا الكتاب بالذات. تتم التجربة على عدد من الصفحات كعينة من الكتاب.
- Ø تتم إعادة عملية التعرف الضوئي للكتاب مرة أخرى بعد التدريب.
- Ø يتم عمل مراقبة جودة لنسبة الأخطاء مرة أخرى.
- Ø في حالة الكتب القديمة ذات الخطوط الزخرفية لا يمكن استخدام برامج للتعرف على الحروف؛ وإنما يتم ذلك يدويًّا باستخدام برامج أخرى لتحديد الكلمات على الصفحة، وتستهلك هذه العملية وقتًا طويلاً وتُستخدم في الحالات الخاصة.

ولا يطبَّق تأمين الجودة على الكيانات الرقمية فقط ولكن أيضًا على الميتاداتا المصاحبة لها للتأكُّد من الالتزام بقواعد الإدخال العامة، واستخدام خطة الميتاداتا بطريقة سليمة، والتأكُّد من الاستخدام الصحيح للمفردات المحكمة، وضمان إدخال كل العناصر الإجبارية بخطة الميتاداتا [١٧].

(٥)- تتم عملية التمييز الضوئي للحروف (OCR) لتتحول الصورة الرقمية إلى مستند قابل للبحث.

وجدير بالذكر أن (ISIS) قام بإجراء مجموعة من الأبحاث بالتعاون مع منتجي برامج (OCR) العربية في مصر والعالم العربي، من أجل تحقيق جودة عالية وإنتاج أكبر كَمٍّ من المواد الرقمية العربية.

(٦)- التشفير (التكويد) Encoding:

وتتلخص فكرته في عمل مسح وقراءة بتقنية OCR للكتب، ووضع النص في خلفية الصور Image on Text في طبقة أخرى، بنفس مواقع الحروف والكلمات، ولا يظهر للقارئ إلا الصورة، ولكن عند النسخ والبحث يكون التعامل مع النص المخفي، ويدعم الملفات في صيغة DJVU and PDF [١٨].

ويتم إعداد نسختين من الكتاب بصيغة PDF، الأولى ذات درجة وضوح عالية تصل إلى (١٥٠ dpi) وذلك للنشر داخل جدران مكتبة الإسكندرية من خلال الشبكة الداخلية Intranet، والثانية ذات درجة وضوح منخفضة تصل إلى (٧٢ dpi) للنشر على شبكة الإنترنت[١٩].

(٧)- تأمين الجودة على النصوص الرقمية المُكَوَّدَة للتأكد من استخدام لغة التكويد بطريقة سليمة، وتُعتبر عملية تأمين الجودة من أهم مراحل مشروعات التحويل الرقمي، وذلك يرجع إلى قياس مدى جودة كل مرحلة على حِدَة للحصول على منتج مرقمن فعال، ولذلك فقد قام مشروع دار بتأمين الجودة خلال المراحل المختلفة للمشروع وتُحفظ بيانات تأمين الجودة لكل مرحلة من المراحل المختلفة لعملية الرقمنة في ملفات؛ وذلك لتعزيز المراقبة المستمرة لجودة الكيانات الرقمية خلال تدفق العمل.

(٨)- الأرشفة الرقمية، وفيها يتم تخزين الكتاب الرقمي على وسائط قابلة للنقل كالأقراص الضوئية والشرائط المغناطيسية، فالمعلومات الرقمية معرَّضة للفقدان على المدى البعيد سواء لأسباب متعلقة بتقادُم ملفات المعلومات نفسها، أو لتقادُم وسيط التخزين، أو لتقادُم الأجهزة المطلوبة لقراءة تلك المعلومات، فضلاً عن أن البرامج التي تم استخدامها لإنشاء ملفات المعلومات قد لا تكون متاحة بعد فترة من الزمن، الأمر الذي يجعل قراءة الملفات المكتوبة باستخدام تلك البرامج أمرًا غير ممكن في حالات كثيرة. ولتحقيق استمرارية إتاحة محتويات المكتبة الرقمية ينبغي وضع استراتيجيات للحفظ، لضمان قابلية تلك المحتويات للاستخدام والإتاحة على المدى البعيد، حتى في ظل التغييرات التقنية المستقبلية سواء للبرامج أو الوسائط[٢٠].

ومن الجوانب التي ينبغي مراعاتها في المشروع الرقمي منذ بداياته لتقليص احتمالات فقدان المعلومات الرقمية، نذكر الآتي:

أولاً :وضع الاستراتيجية الملائمة للحفظ الرقمي

ينبغي في المشاريع الرقمية أن تُعِد استراتيجية للحفظ الرقمي قبل القيام بتنفيذ المشروع؛ لأن عملية التنفيذ ينبغي أن تتم اعتمادًا على المعايير المناسبة لاستراتيجية الحفظ المحدَّدَة[٢١].

وقد أوضحت الدراسة أن مشروع (مستودع الأصول الرقمية) قد اعتمد في صياغة وتشكيل استراتيجيات الحفظ المستخدمة على نظام المعلومات الأرشيفي المفتوح، والذي يُعرف اختصارًا بنمط (OAIS) المرجعي:

Open archival information system (OAIS) reference model (ISO 14727:2003) وهو من أبرز أدوات وآليات الحفظ الرقمي التي طوَّرتها اللجنة الاستشارية لنظم بيانات الفضاء بوكالة ناسا، وحصلت على الموافقة والتصديق عليها من منظمة التوحيد القياسي، وهو عبارة عن إطار مرجعي يصف ويحدد المبادئ ومنهجية استراتيجيات الحفظ طويل المدى التي تكفل الحفظ والإتاحة المستمرة للمواد الرقمية، ويضع حلولاً لجميع العقبات المتوقعة في عملية الحفظ الرقمي في ظل التغيرات المستمرة في البيئة الرقمية، كما تحدد العلاقة بين منتج المعلومات الرقمية والأرشيف والمستفيد[٢٢]، وقد استُخدمت استراتيجيات الحفظ الرقمي التالية:

♦ التنشيط (Refreshing) وتتضمن عملية التنشيط النقل الدوري للملف من وسيط تخزين مادي واحد إلى الآخر لتجنُّب التلف أو تقادُم الوسيط، وهنا يمكن نقل الملفات من وسيط ما إلى وسيط آخر من النوع نفسه لتجنُّب تدهور الوسائط، أو نقل الملفات من وسيط غير مستخدم أو متقادم إلى وسيط أحدث، فمثلاً قامت مكتبة الإسكندرية بنقل الملفات من القرص الضوئي إلى أقراص DVD لتجنُّب تقادُم الوسائط.

♦ التهجير (Migration) ويتضمن اتجاه الهجرة نقلَ أشكال الملفات الحالية إلى أشكال ملفات جديدة، والتي أصبحت فعالة، وهذا قد يحدث عندما يتاح معيار جديد أو عندما يتقادم البرنامج الحالي.

ومثالاً على ذلك أنه في مشروع (دار) تم نقل شكل صورة DJVU إلى شكل TIFF* [٢٣]. وقد تبنَّى المشروع تلك الاستراتيجيات لضمان إتاحة وقراءة الكيانات الرقمية في المستقبل.

وتبقى الإشارة إلى أن المؤسسات ونظم المعلومات تَعْمُد إلى أسلوب النسخ الاحتياطي Back up للمحافظة على بياناتها؛ حيث تحاول بذلك التغلُّب على احتمالات فقدان البيانات وضياعها في حالة وجود نسخة واحدة فقط منها، وبالتالي يتم إعداد نسخ أخرى يتم استخدامها في حالة فقدان البيانات لأي سبب من الأسباب، سواء أكان لمشكلة في البرنامج، أم الأجهزة أم خطأ من الموظفين، أو حدوث كوارث طبيعية، وبذلك فإن وجود نسخ احتياطية للبيانات محفوظة في أماكن آمنة؛ غالبًا في موقع آخر غير الموجودة فيه النسخة الأصلية، يساعد في حماية تلك البيانات وتقليل احتمالات ضياعها[٢٤].

وقد حرصت مكتبة الإسكندرية على عمل نسخ احتياطية للمعلومات لحمايتها من الفقدان، حيث قامت بحفظ نسختين من الملف على هيئة ملف مضغوط ZIP؛ وذلك لأن الملفات المضغوطة تشغل مساحة تخزين أقل ويمكن نقلها إلى كمبيوتر آخر بسرعة أكبر من مثيلتها من الملفات غير المضغوطة. وتوزَّع النسختان واحدة على CD والأخرى على شريط tape، وتتمثل مواصفات نسخة الحفظ فيما يلي:

○ نوع امتداد الملف TIFF. ○ دِقَّة النسخة (Resolution)600 dpi.

ثانيًا: اختيار وسيط التخزين الملائم

يعتمد المشروع في عملية الحفظ الرقمي على الأقراص المدمجة (CD-Rom) وأقراص الفيديو الرقمية (DVD) والتي تتميز بالسعة العالية للتخزين، غير أنها تفتقر للمعيارية[٢٥].

ثالثًا: الالتزام بالمعيارية في أشكال الملفات

ينبغي الحرص في المشروعات الرقمية على الابتعاد عن أشكال الملفات غير المعيارية؛ لأنها تكون عُرْضَة للتغيير والزوال مع احتمالات توقُّف الشركات التي تُصدرها أو تطوير تلك الشركات لأشكال جديدة، وتقوم هذه الفكرة على افتراض أن منتجات البرامج التي تساير معايير واسعة الانتشار في السوق، تكون أقل عُرْضَة للزوال من البرامج الأخرى المتاحة في السوق.

وبالتأكيد فإن إدارة تلك الملفات ستكون أسهل على المدى البعيد إذا تطابقت تلك الملفات مع الأشكال المعيارية[٢٦]. وذلك ما تم تحقيقه بالمشروع؛ فيتمثل ناتج عملية رقمنة الكتب في أحد أشكال الملفات المعيارية للصور وهو TIFF.

رابعًا: إنشاء الميتاداتا اللازمة لتمثيل المعلومات

إن الحفظ الجيد للمصادر الرقمية يتطلب توافر نظام فعال للميتاداتا لوصف تلك المصادر، وإدارة الوصول إليها وضمان حفظها، وعناصر الميتاداتا مطلوبة لدعم اختيار استراتيجية الحفظ المتَّبَعة في

---

* (Tagged Image File Format (Tiff هو شكل ملفات على درجة عالية من الوضوح والجودة، غير أنه يشغل حيِّزًا كبيرًا؛ لذا فإنه يناسب الحفظ الأرشيفي للنسخ الرئيسة للصور الرقمية، ومن الصعب نقل الملفات في هذا الشكل عبر الشبكة إلا إذا كانت ذات سرعة عالية.

المشروع الرقمي؛ حيث إنها تتيح معلومات مفصلة لوصف أشكال الملفات المستخدمة، وبيئة البرامج والأجهزة، وقد تتضمن أيضًا معلومات حول إدارة الحقوق وضبط إتاحة الوصول.

ولا تُعتبر مرحلة الحفظ الرقمي المرحلة النهائية للمشروع، ولكن أي قرار فني يتصل بالرقمنة يجب أن يضع في اعتباره استمرارية المواد الرقمية، فيجب مثلاً اختيار أحدث وسائط التخزين واستخدام أشكال الملفات وتقنيات الضغط المعيارية المتاحة بشكل واسع(٢٧). وقد اعتمد (دار) على صيغة TIFF للحفظ بعد استخدام تقنية ضغط CCITT Group 4 Compression، والتي ثبت نجاحها مع الكتب غير الملونة.

(٩)- نشر الكتاب الرقمي على شبكة الإنترنت، وتجدر الإشارة إلى أن النسخة الأولى من (مستودع الأصول الرقمية) تَعْرِض الكتب الرقمية في صيغتيPDF & DJVU، ثم في النسختين الثانية والثالثة تم عرض الكتب اعتمادًا على تقنية تركيب الصورة على النص.

(١٠)- خروج الكتب من المعمل الرقمي بعد الانتهاء من جميع المراحل سالفة الذكر، وإعادة جميع المصادر التي خضعت لعملية الرقمنة إلى أماكنها الأصلية داخل المكتبة.

٢/٢ التجهيزات المادية والبرمجية بالمعمل الرقمي.

١/٢/٢ المعمل الرقمي:

قام المعهد الدولي للدراسات المعلوماتية بإنشاء معمل رقمي مجهَّز بأحدث التكنولوجيات، يضطلع بمهمة تقديم الخدمات الرقمية المتخصصة. ويعمل المعمل الرقمي على رقمنة الوسائط المختلفة (الشرائح على اختلاف أشكالها، نيجاتيف الصور، الكتب، المخطوطات، الصور، الخرائط). وقد تم تجهيز المعمل بالأدوات اللازمة للمسح الضوئي والفهرسة والأرشفة والإدارة، مما يسمح بالتحكم آليًّا في العمل بأكمله.

وقد قامت جامعة كارنيجي ميلون Carnegie Mellon University في عام ٢٠٠٣ بإهداء المكتبة خمسة أجهزة للمسح الضوئي من نوع (Zeutchel) OS-5000 لتعزيز إمكانيات المعمل الرقمي، وذلك بعد أن لاحظت التقدم الكبير الذي حدث به، ويعتمد الكثير من مشروعات المعهد الدولي للدراسات المعلوماتية بشكل كبير على المعمل الرقمي لتحقيق أهدافه.

تحتاج كل خطوة من خطوات الترقيم الموضحة في الشكل رقم (٥) إلى مجموعة من التجهيزات المادية والبرمجية، والتي يتوقف اختيارها على طبيعة وأهداف المشروع، وعلى نوعية المواد التي يتم ترقيمها، وتجدر الإشارة إلى أن عملية الرقمنة تتم داخل مكتبة الإسكندرية؛ حيث تمتلك معملاً رقميًّ مجهزًا بالكفاءات البشرية المتخصصة في تقنية المعلومات وتطبيقاتها، وكذلك الإمكانيات والخبرات اللازمة لمعالجة المجموعات المرغوب في رقمنتها، وعدد من المتخصصين في الحاسب الآلي لتشغيل وإدارة نظام المكتبة الإلكتروني وتقديم خدمات الدعم والصيانة والتدريب، ويقدَّر عدد العاملين بالمعمل الرقمي بنحو ١٢٠ فردًا(٢٨).

ويتضح فيما يلي أهم التجهيزات المادية والبرمجية التي تم الاعتماد عليها بالمشروع، وتنقسم الأجهزة والتقنيات المرتبطة بعملية الرقمنة إلى قسمين رئيسين هما:

أولاً/ التجهيزات المادية:

من أهم المتطلبات الأساسية لمشروعات الرقمنة المعدات والأجهزة اللازمة للقيام بعملية الرقمنة، ومن هذه الأجهزة ما يلي:

١- الماسحات الضوئية للكتب (Book Scanners)

بدأ استخدام هذه الفئة من الماسحات الضوئية الإلكترونية بشكل تجاري منذ عدة سنوات، وهي موجَّهة إلى رقمنة مصادر المعلومات المجلدة –المجلدات-، ويُطْلَق عليها كذلك "ماسحات الكتاب المفتوح"، وهي تناسب لدرجة كبيرة احتياجات المكتبات؛ وذلك نتيجة أن السطح الزجاجي للجهاز المخصص للرقمنة يضمن معالجة النصوص ذات الأحجام الكبيرة، وفي أثناء عملية الرقمنة يكون الوعاء مفتوح والنص المراد رقمنته متجهًا إلى أعلى وأداة التعرف -القراءة- الضوئي المستخدمة في الرقمنة تُوجَد أعلاه[٢٩].

- ويشتمل المعمل الرقمي على خمس ماسحات ضوئية من نوع **Minolta PS7000** كما هو موضَّح بالشكل رقم (٧)، ويتمتع الماسح الضوئي بالعديد من المواصفات الموضحة بالجدول رقم (١):

شكل رقم (٧) صورة الماسح الضوئي Minolta PS7000

جدول رقم (١) مواصفات الماسح الضوئي Minolta PS7000

مواصفات الماسح الضوئي	
النوع	ماسح ضوئي مسطح – ثابت
الحجم الأقصى للوسيط الممكن رقمنته	A2 (420 x 594 mm)
نوع المدخلات	أحادي اللون
عمق مستويات التدرج الرمادي	٨ بت (٢٥٦ مستوى رماديًّا)
درجة وضوح المسح الضوئي Resolution	٦٠٠/٤٠٠/٣٠٠/٢٠٠ نقطة في البوصة 200/300/400/600 DPI
سرعة الرقمنة	٦ صفحات / الدقيقة.
مطابقة المعايير	TWAIN

يسمح الماسح الضوئي من نوع Minolta PS7000 بتصوير أحجام مختلفة وبنوعيات متعددة بالألوان أو بالأبيض والأسود، ويتصل الماسح الضوئي بالحاسوب من خلال منفذ USB، أما من حيث

اتصاله بالبرامج فإنه يتم من خلال برامج تشغيل محركات Drivers، وأشهرها برنامج TWAIN وهو (معيار قياس صمم ليسمح لبرنامج الصور الذي تتعامل معه بالتواصل مع الماسح الضوئي).

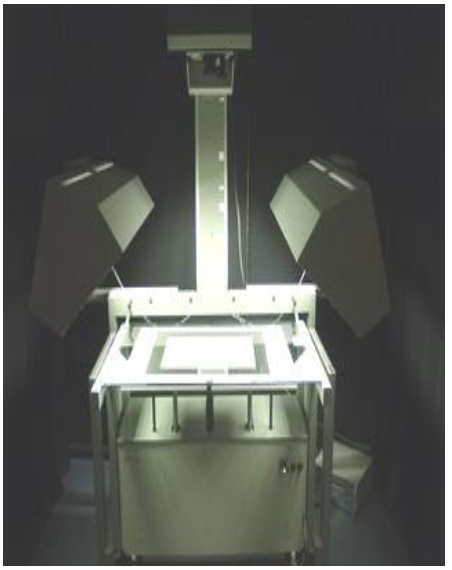
- نوع آخر من الماسحات الضوئية من نوع **Cruse 200SL 850** كما هو موضح بالشكل رقم (٨)، ويتمتع بالمواصفات الموضحة بالجدول رقم (٢):

شكل رقم (٨) صورة الماسح الضوئي Cruse 200SL 850

جدول رقم (٢) مواصفات الماسح الضوئي Cruse 200SL 850

مواصفات الماسح الضوئي	
النوع	Large format colored scanner
الحجم الأقصى للوسيط الممكن رقمنته	2×A0 (1189 mm × 1682 mm)
درجة الوضوح المسح الضوئي Resolution	من ٤٠٠ إلى ٨٠٠ نقطة في البوصة 400 – 800 DPI
عمق اللون	48 bit

- نوع ثالث **Zeutchel Omniscan 9000** كما هو موضح بالشكل رقم (٩)، ويتمتع بالمواصفات التي يوضحها الجدول رقم (٣):



شكل رقم (٩) صورة الماسح الضوئي Zeutchel Omiscan 9000

جدول رقم (٣) مواصفات الماسح الضوئي Zeutchel Omniscan 9000

مواصفات الماسح الضوئي	
النوع	Colored Scanner
الحجم الأقصى للوسيط الممكن رقمنته	A0 (841 mm × 1189 mm)
درجة الوضوح المسح الضوئي Resolution	٨٠٠ نقطة في البوصة 800 DPI
عمق اللون	36 bit

يسمح الماسح الضوئي Zeutchel Omniscan 9000 بنظام المسح الضوئي الملون للكتب والصحف والوثائق الكبيرة مثل {الرسومات والخرائط}[٣٠].

٢- الحاسبات الإلكترونية.

تُعَد الحاسبات جزءًا لا يتجزأ من عملية الرقمنة، وهي تُستخدم في عمليات الإدخال وإدارة وتخزين الصور الرقمية، كما تُستخدم في ضبط إعدادات الماسح الضوئي ومراقبة جودة الصورة الرقمية، كما يتم عن طريقها إدخال الميتاداتا للمواد الممسوحة ضوئيًّا. وتُستخدم الحاسبات الآلية في تخزين الملفات وفهرستها وعمل النسخ الاحتياطية، وعند اختيار الحاسب الآلي لاستخدامه في عملية الرقمنة يتم التركيز على اختيار الأجهزة القادرة على تلبية مطالب الرقمنة بناءً عليه، من حيث سعتها التخزينية والسرعة، والمعالجة وغيرها. ويجب استشارة المتخصصين في تكنولوجيا المعلومات عند شراء الحاسبات الآلية، ويتوقف عدد أجهزة الحاسبات المطلوبة للرقمنة على عدد وحجم المواد المراد رقمنتها، وتتحدد السعة التخزينية للحاسبات وفقًا لنوعية المواد المراد رقمنتها[٣١]. ويشتمل المعمل الرقمي بمكتبة الإسكندرية على عدد من أجهزة الحاسوب الحديثة، وتتمتع بإمكانيات متطورة من حيث السرعة العالية والسعة التخزينية، وترتبط تلك الحاسبات بشبكة المكتبة المحلية، وفيما يلي عرض لإمكانيات الأجهزة الموجودة بالمعمل الرقمي:

- أنظمة التشغيل الموجودة: Microsoft Windows XP Professional - Linux - Microsoft Windows Server 2003.
- نوع المعالج Core 2 duo/Core 2 Quad.
- حجم ذاكرة الوصول العشوائي (RAM) ١-٤ جيجا بايت.
- حجم القرص الصلب HDD 400 GB.
- اتصال الإنترنت: تُقَدَّر سرعة الاتصال بشبكة المعلومات الدولية (الإنترنت) داخل مكتبة الإسكندرية بمقدار 155 Mbps. حيث قامت المكتبة بتحديث سرعة الإنترنت بها من 10 Mbps إلى 155 Mbps (STM1) وذلك بالتعاون مع الشركة المصرية للاتصالات، وبهذا التحديث أصبحت مكتبة الإسكندرية في مصاف عدد محدود من المنظمات والمؤسسات الثقافية في مصر التي تستخدم هذا النوع من الشبكات STM1، وأتاح هذا التحديث سرعات عالية جدًّا تُسَهِّل الاطلاع على المعلومات الوفيرة التي تتيحها مكتبة الإسكندرية على المستويين المحلي والدولي؛ حيث يدعم نوعية الخدمات التي توفرها المكتبة لزوارها وكذلك للمُطَّلِعين من خلال شبكة المعلومات الدولية، فقد أصبح بالإمكان الاطلاع على المواد الإلكترونية والمجموعات الرقمية بمكتبة الإسكندرية بسرعة وسهولة، سواء من داخل المكتبة أو عن طريق شبكة المعلومات الدولية[٣٢].

٣- أجهزة التخزين ووسائط التخزين Storage.

اعتمدت مكتبة الإسكندرية على وسائط التخزين التالية: وسائط تخزين خارج الخط كالأقراص المدمجة (CD-Rom)، وأقراص الفيديو الرقمية (DVD)، والشرائط المغناطيسية (Tapes)، ووسائط التخزين على الخط المباشر (Online Media Storage)، وتُستخدم في حالة أرشفة كل الكتب المنتهية في مخزن على الخط المباشر. وقد وضعت المكتبة ألية لمنع الحذف أو التعديل غير المصرَّح به من المعلومات المخزَّنة[٣٣].

ثانيًا/ التجهيزات البرمجية:

تحتاج عملية الرقمنة إلى بعض البرمجيات المهمة التي ينبغي توفرها واستخدامها حتى تتم عملية الرقمنة، وتُستخدم عدة برمجيات لتحقيق الغرض المراد، وهي:

١- برمجيات المسح الضوئي Scanning.

٢- برمجيات معالجة الصور Image Manipulation.

٣- برمجيات التعرف الضوئي على الحروف OCR.

١- برمجيات المسح الضوئي Scanning:

يتضمن هذا العنصر برمجيات التحكم في أجهزة الماسحات الضوئية الإلكترونية، فتحتوي الماسحات الضوئية على برامج لالتقاط الصور تُستخدم للتحكم في جودة الصورة، ويجب أن تتوافق البرامج التي تُستخدم في المسح الرقمي للمواد مع أجهزة الحاسب والماسح الضوئي، وهذه البرامج يتم اختيارها من قِبَل الشركة المصمِّمة للماسح، ومعرفة درجة توافقها مع أجهزة المسح الضوئي، وقد استخدمت مكتبة الإسكندرية برنامج مسح ACDSee 4.0 بدرجة وضوح ٣٠٠ نقطة في البوصة للمواد باللغة العربية؛ حيث ثبت أنها الدرجة الأنسب لنجاح عملية التعرف الضوئي، في حين أن المسح الضوئي بدرجة ٦٠٠ نقطة في البوصة أو تخفيضه إلى درجة ٣٠٠ نقطة في البوصة بعد المسح يؤدي إلى ضَعْف عملية التعرف عليها، ويعتمد خيار "نص" أو "صورة" على طبيعة الصفحة الأصلية، ويكون ملف الصور الممسوحة ضوئيًّا في صيغة Tiff[٣٤].

٢- برمجيات معالجة الصور Image Manipulation:

توجد برامج إضافية بالإضافة إلى برمجيات المسح الضوئي تتيح مزيدًا من التحكم في الصورة وإعطاء مميزات أفضل كإمكانية تصغير وتكبير الصورة وتحريكها في الاتجاهات كافة، وزيادة درجة تباين الصورة وسطوعها، وقد استخدمت مكتبة الإسكندرية نوعين من البرمجيات يدويًّا برنامج Photoshop وآليًّا برنامج Scanfix وبرنامج CMU ويحقق المزج بين كلا الأسلوبين كفاءة أفضل وتحسين جودة الصور ووضوحها[٣٥].

٣- برمجيات OCR المستخدمة بالمعمل الرقمي بمكتبة الإسكندرية:

 # معايير اختيار برمجيات الـ OCR (Optical Character Recognition):

يعتمد اختيار برمجيات التعرف الضوئي للحروف -في الأساس- على مدى صلاحيته للاستخدام والتطبيق واشتماله على قواميس متخصصة تتعلق بموضوع النصوص التي تتم معالجتها بواسطة مثل هذه البرمجيات. تتضمن كذلك معايير الاختيار السبلَ المُتَّبَعة في التعرف على الحروف والسرعة في "قراءة" الحروف، وأنواع وأحجام وأشكال الحروف التي يمكن التعرف عليها، واللغات التي تحتويها البرمجيات ويمكن التعامل معها، وأشكال وأنماط عرض النصوص وإتاحتها بعد إتمام عملية التعرف عليها، وضرورة اشتمال البرمجيات على قواميس متخصصة غنية وثرية، وأن تشتمل على بعض المهام المتقدمة، منها على

سبيل المثال إمكانية التدقيق الإملائي وبرامج التحرير والنشر، وأخيرًا يُعتبر عامل السعر من العوامل التي يمكن أن تؤخذ في الاعتبار[٣٦].

ونشير في هذا الإطار إلى نوعين من البرمجيات المستخدَمة في التعرف الضوئي على الحروف، واللذين يُستخدمان بشكل مستمر في رقمنة المجموعات العربية بمكتبة الإسكندرية، وهما:

(أ) برنامج القارئ الآلي من صخر Sakhr Automatic reader.

أصدرت شركة صخر العديد من البرمجيات، كان آخرها Sakhr Arabic OCR الإصدار الثامن، والذي يستطيع إدراك علامات التشكيل العربية؛ ويستخدم تِقَنِيَّتَيْن للتعرف الضوئي أُولاهما: تقنية التعرف العامOmni Technology التي تعتمد على أبحاث متقدمة في مجال الذكاء الاصطناعي، للمساعدة في التعرف على الوثائق مباشرة دون الحاجة إلى تدريب، والثانية تقنية التعليم أو التدريب Training technology لزيادة نسبة الدقة في التعرف على الحروف، كما يعتمد على أبحاث صخر اللغوية في اللسانيات الحاسوبية Computational Linguistics التي تقوم باستخدام قواعد اللغة أثناء التمييز، ويقوم برنامج القارئ الآلي من صخر بتمييز النص العربي الممسوح ضوئيًّا أو خط اليد العربي، آخذًا في الاعتبار الميزات المتعددة للخطوط العربية وخصائص اللغة العربية، كما يستخدم الدونجل "Dongle" لحماية حقوق التأليف والنشر[٣٧].

(ب) برنامج نوفو دينامكس NovoDynamics' VERUS™ Professional.

صُمم البرنامج في معمل آن أربور بجامعة ميتشجان "Ann Arbor, "Michigan، ويُستخدم لتحويل صور الوثائق الورقية أو المستندات إلى شكل قابل للقراءة على الحاسوب، سواء كانت متهالكة الحالة أو منخفضة الجودة أو صفراء الأوراق أو غير مألوفة الخطوط كما في حالة النصوص القديمة، وقد وقَّعت شركة نوفو دينامكس العقد مع مكتبة الإسكندرية وبدأت المشروع رسميًّا في عام ٢٠٠٦م.

- ويتميز البرنامج بالعديد من الخصائص أهمها:

١- تنقية الصورة قبل مسحها، مما يقلل نسبة الخطأ.

٢- إمكانية الحفظ كملف PDF قابل للبحث، حيث يتيح البرنامج إمكانية حفظ المستندات المصوَّرة كملف PDF قابل للبحث، وتوصيله إلى تطبيقك المفضل لتوفير مساحة تخزين أكبر، ويقوم البرنامج بدمج النص الناتج عن عملية الـــ OCR مع صورة الوثيقة في ملف الـــ PDF، وتمكن هذه الخاصية المستخدم من البحث الحر عن كلمات لاسترجاع النص المطلوب عبر البرنامج المعروف Adobe Reader.

٣- التعرف على عدة أنواع من الخطوط العربية، كما يدعم الخطوط الفارسية / الأردو / الدارية / البيشتية، مع إمكانية دعم النصوص العربية المختلطة مع الإنجليزية والفرنسية والإسبانية الموجودة في الوثائق العربية؛ للوصول إلى نتائج بدِقَّة فائقة، ثم تأتي خاصية Searchable PDF Output حيث يتم وضع النص في خلفية الصورة، وتوجد تلك الخاصية في النسخة الاحترافية منه فقط دون النسخة العادية،

وتتلخص فكرتها في عمل مسح وقراءة بتقنية OCR لملف PDF، ويوضع النص في خلفية الصور في طبقة أخرى، بنفس مواقع الحروف والكلمات، ولا يظهر للقارئ إلا الصورة، ولكن عند النسخ والبحث يكون التعامل مع النص المخفي.

٤- القابلية للتدريب والتعليم، فلا يحتاج لتقنية التعليم كبرنامج صخر، بل يخزن معلوماته وتحليلاته بنفسه عن كل حرف آليًّا في نظام XML، فيتعرف آليًّا على الأحرف، ويبثها في ملف قابل للبحث بالعربية، وربما تفُوق نتائجه برنامج القارئ الآلي من صخر.

٥- يمتلك واجهة استخدام عملية (إنجليزي أو عربي)، ولا يتميز البرنامج بالدِّقَّة المتناهية فقط؛ ولكن يتميز أيضًا بسهوله الاستخدام.

٦- يوفر البرنامج واجهة تطبيق برمجية (API) تعمل مع بيئة .Net و ava و + + C البرمجية، وذلك لتبسيط التكامل مع التطبيقات الخارجية[٣٨].

§ بينما يستخدم برنامج **Tricky** للكتب القديمة، كما تم إعداد برنامج **OCR Corrector** ويُستخدم في حالة الكتب القديمة جدًّا، والتي تحتاج إلى مزيد من التدقيق؛ حيث تحاكي إلى حد كبير دِقَّة الإنسان البشري في المراجعة.

◘ وتكون صيغ الملفات بعد إتمام عملية التعرف عليها في الأشكال التالية:

- ملفات نص (TXT).
- ملفات لغة ترميز النصوص الفائقة HTML.
- ملفات صيغة النص الغنية (.RTF) Rich Text Format.
- ملفات نص القارئ الآلي (.FRF,.ART) Native.

٣/٢ التحويل الرقمي للمجموعات العربية بمستودع الأصول الرقمية.

تواجه عملية التحويل الرقمي للمجموعات العربية بمستودع الأصول الرقمية بمكتبة الإسكندرية بعض التحديات، ويُقصد بالتحديات هنا التثبُّت من مدى توافر المقومات الأساسية، أو من مدى وضوح الرؤية في التعامل مع المتوافر منها، وفي الفقرات التالية يتم استعراض المعوقات والإشكاليات الخاصة بها، ومحاولة استشراف الحلول المقترحة.

١/٣/٢ إشكاليات التحويل الرقمي:

على الرغم من أهمية عملية الرقمنة والمميزات التي تمنحها، فإنها عادة ما تصطدم بكثير من التحديات، سواء كانت تحديات مالية خاصة بالميزانية والاعتمادات المخصصة، أو بالمسائل الفنية المتعلقة

بتبني أفضل المقاييس وأشكال ملفات مصادر المعلومات الناتجة عن الرقمنة، أو القضايا المرتبطة بالبنية التقنية لمشروع الرقمنة، والاتفاقيات الخاصة بتخطي الإشكاليات المرتبطة بحقوق المؤلفين والناشرين.

وتُعتبر هذه التحديات جوهرية، ولها تأثيرها المباشر في إعداد سياسة رقمنة مصادر المعلومات، وتبني معايير اختيار مصادر المعلومات التي تتم رقمنتها وأساليب حفظها واختزانها. وبناءً عليه تستعرض الفقرات التالية أبرز هذه الإشكاليات:

١- ترتبط هذه الإشكالية بالمسائل المتعلقة بحماية حقوق الناشرين والحقوق الفكرية للمؤلفين، وحتى يمكن أن تتفادى المكتبات الدخول في منازعات قضائية لجأت في البداية إلى قصر عمليات الرقمنة على مصادر المعلومات التي لا تخضع لحقوق المؤلفين والناشرين، لذلك كان من الضروري على مكتبة الإسكندرية التفكير في الحقوق المتعلقة بإتاحة مصادر المعلومات وعرضها للجمهور العام، وعدد المستفيدين الذين يمكنهم الاطلاع على مصادر المعلومات، وتحديد مستوى جودة النصوص المرغوب في استعراضها من خلال منظومة الشبكات، وتقع تلك المسؤولية على عاتق الإدارة القانونية المعنية بالملكية الفكرية بالمكتبة.

٢- عدم تبني مشروع (دار) سياسة اختيار مكتوبة للكتب التي ستخضع لعملية الرقمنة، فقد تم اتخاذ قرار الرقمنة مباشرة من خلال النص الأصلي، ويتم اتخاذ مثل هذا القرار في ضوء الخصائص الفيزيائية والقيمة التي يحظى بها مصدر المعلومات[٣٩].

٣- ترتبط هذه الإشكالية بحال بعض الأعمال؛ حيث التلف أو نقصان بعض الصفحات، أو عدم توافر بعض المصادر فعليًّا برصيد المكتبة، علاوة على المعالجة الخاصة التي تتطلبها الرسوم والصور والخرائط.

٤- بالنسبة للمجموعات العربية (نظرًا لاختلاف طبيعة اللغة العربية عن اللغة الإنجليزية، وهذا الاختلاف له تأثيره المباشر على القراءة الآلية، فكان أول تحدٍّ يواجه القراءة الآلية في اللغة العربية هو اتصال النص العربي، حيث إن الكتابة العربية لا يوجد بها خط مفكك، ويجب أن تكون الكتابة متصلة).

وهكذا فإن المشكلة الأساسية التي يجب حلها هي تقسيم الكلمات إلى حروف، ولكن يختلف الأمر بالنسبة للغة الإنجليزية، حيث يتم التركيز بشكل أكبر على عملية تمييز الحروف ذات الخطوط والأحجام المختلفة.

وهناك بعض المشاكل الأخرى التي تَظهر في خطوط كثيرة في اللغة العربية، مثل تركيب الحروف فوق بعضها، بالإضافة إلى أنه قد تختلف الحروف في الشكل تبعًا لموقعها في الكلمة، مما يؤدي إلى توسيع دائرة التمييز مقارنة باللغة الإنجليزية واللغات الأخرى[٤٠]، ولذلك تطرح عملية التعرف الضوئي على الحروف (OCR) باللغة العربية تحديات فريدة من نوعها؛ فتختص تقنية التعرف الضوئي على الحروف بمجموعة من العمليات التي تتعلق بالتعرف على نص معين وقراءته، وبالتالي تحويله إلى ملف نصي، ويتضمن ذلك الملف النصي بيانات ومعلومات مكودة في شكل معياري (ASCII) American Standard Code for Information Interchange، والذي يحتلّ غالبًا مساحة أقل من تلك التي يحتاج إليها ملف الصور، ويجب

–في جميع الحالات- تصحيح الملف النصي الذي يتم إنشاؤه بواسطة برمجيات الـــOCR بطريقة يدوية؛ وذلك بهدف ضمان نص بدون أي أخطاء، ويحتاج هذا العمل إلى فترات طويلة ومجهود للقيام به، مما يخلق نوعًا من التأثر السلبي نحو الانجذاب لتطبيق تقنيات الـــOCR مقارنة بالإدخال اليدوي للنصوص[٤١].

٥- تسعى الكثير من مشروعات التحويل الرقمي إلى إتاحة الكيانات الرقمية لقطاع عريض من المستفيدين من خلال تيسير التعامل مع تلك الكيانات بأكثر من شكل من أشكال الملفات المستخدمة للتخزين، ومن ثَمَّ يجب أن يوضع في الاعتبار -عند تصميم المستودعات- أشكال التخزين وعرض الكيانات بأسلوب يتناسب مع احتياجات الفئات الخاصة مثل ضعاف وفاقدي البصر، وقد وضع الاتحاد العالمي للويب معايير لبناء وعرض الكيانات الخاصة بتلك الفئات، ومن أبرز توصياته أن توضح تسجيلة الميتاداتا أسلوب عرض يتناسب مع طبيعة الفئات الخاصة، وتوجد العديد من البرمجيات التي تُساعد على تحقيق ذلك، إلا أنه للأسف معظم المعايير تفتقر إلى هذا الجانب، كما أن مشروع (مستودع الأصول الرقمية) لم يأخذ هذا الموضوع على محمل الجد حتى الآن[٤٢].

وبعد إبراز أهم المعوقات والتحديات المتعلقة بعملية التحويل الرقمي للمجموعات العربية بمشروع (دار)، يمكن التحقق من أن هذه الإشكاليات تمثل تحديًا حقيقيًّا يجب تخطيه في المرحلة المقبلة من تطوير المشروع.

٢/٣/٢ الآفاق المستقبلية:

بعد هذا العرض للمحاور الرئيسة المتعلقة ببرمجيات الـــOCR، يبدو من الأهمية تتبع الآفاق المستقبلية الخاصة بتلك التقنيات، لذا يمثِّل هذا الهدف نقطة المعالجة الرئيسة في الفقرات التالية.

يكمن الغرض المنشود من وراء برمجيات التعرف الضوئي على الحروف في تحويل الصورة النصية إلى ملف نصي يمكن قراءته ومعالجته بواسطة الحاسب الآلي، وذلك مع تجنُّب نسبة الأخطاء قدر المستطاع.

وتقوم مكتبة الإسكندرية في الوقت الراهن بإجراء المزيد من الدراسات والبحوث المستقبلية للوصول إلى عملية تعرف ضوئي أكثر كفاءة وفعالية، خاصة فيما يتعلق بالبناء الهيكلي المنطقي للنص، ويمكن القول إن الحصول على نص مقروء ضوئيًّا بواسطة برمجيات OCR بدون أي نسبة أخطاء ما زال هدفًا ومطلبًا يُرجى الوصول إليه، وتحقيقه لا يعتمد فقط على كفاءة برمجيات التعرف الضوئي على الحروف؛ ولكن أيضًا يقتضي في الواقع بعض الاحتياطات التي ينبغي مراعاتها أثناء القيام بهذه العملية، ومنها:

١- التهيئة والإعداد المسبَق للنص المراد رقمنته من ضبط الإضاءة وتباين الألوان، مع ضرورة أن تكون الصورة كذلك سهلة المعالجة.

٢- إمكانية اللجوء لمساعدة برمجيات OCR في اختيار المقاييس والخصائص بهدف التأقلم الجيد وفقًا لطبيعة المحتوى المراد معالجته، مع الأخذ في الاعتبار جودة النص من الناحية الطباعية والورق، ولغة النص، وطبيعة التنسيق المستخدم، إلى غير ذلك[٤٣].

٤/٢ الخلاصة

تناول هذا الفصل عملية التحويل الرقمي للكتب بمشروع مستودع الأصول الرقمية والإشكاليات المرتبطة بها، وتبين من خلاله ما يأتي:

- اعتماد (مستودع الأصول الرقمية) في تأسيسه وإدارته على برنامج محلي تم تطويره بواسطة مكتبة الإسكندرية، وهو برنامج DAR.

- الاهتمام بتوفير القوى البشرية المؤهلة في مجال الرقمنة، والعمل على توفير التقنيات المناسبة التي تدعم عملية التحويل الرقمي، وكان ذلك عاملاً مساعدًا في إتمام مشاريع الرقمنة بجودة عالية داخل مكتبة الإسكندرية.

- صُمم دار ليتوافق مع الإطار المرجعي OAIS؛ حيث يُعَد الحفظ طويل المدى للمواد الرقمية بالمستودعات أمرًا بالغ الأهمية، وذلك لأن المواد الرقمية التي تودع بالمستودع في الوقت الحالي تكون وفقًا لأحدث المعايير التكنولوجية، ولكن بعد مرور عدة سنوات ستصبح هناك مشكلة تتمثل في تقادُم نُظم التشغيل وأشكال الملفات وإصدارات البرامج، ولذلك تعمل مكتبة الإسكندرية على اتباع الاستراتيجيات والخطط التي تكفل حفظ وإتاحة الكيانات الرقمية لفترات طويلة المدى، كما تراعي مكتبة الإسكندرية أنها قد تحتاج في المستقبل إلى الانتقال التكنولوجي من أشكال التخزين والتكويد المستخدمة إلى اصدارات أكثر حداثة أو إلى أشكال أخرى.

- يهدف (دار) من خلال عملية الرقمنة إلى إتاحة مصادر المعلومات لأكبر عدد ممكن من المستفيدين، وإتاحة الوصول لمصادر المعلومات التي كان يصعب الوصول إليها أو التي كان استخدامها مقيَّدًا كالكتب النادرة، وذلك من منطلق دعم المحتوى العربي على الإنترنت وخدمة البحث العلمي والباحثين في شتى التخصصات.

- لم يُرَاعِ القائمون على تصميم المستودع استخدام أشكال ملفات للتخزين وعرض المواد الرقمية بأسلوب يتناسب مع احتياجات الفئات الخاصة مثل ضعاف وفاقدي البصر.

مصادر الفصل الثاني:

[1] هناء شكري مصطفى عصفور. التوثيق المرقمن للتراث الحضاري المصري: دراسة لتجربة مركز توثيق التراث الحضاري والطبيعي/ إشراف شريف كامل شاهين، عبد الله حسين.- القاهرة: جامعة القاهرة، ٢٠٠٩.- أطروحة (ماجستير)– جامعة القاهرة- كلية الآداب- قسم المكتبات والوثائق والمعلومات.- ص ص ٤٤- ٤٥.

[2] عماد عيسى صالح. مشروعات المكتبة الرقمية في مصر: دراسة تطبيقية للمتطلبات الفنية والوطنية / إشراف محمد فتحي عبد الهادي، زين الدين عبد الهادي.- القاهرة: جامعة حلوان، ٢٠٠٤.- أطروحة (دكتوراه) – جامعة حلوان – كلية الآداب – قسم المكتبات والمعلومات.- ص ١٥٥.

[3] المحتوى العربي الرقمي ضرورة استراتيجية وحتمية لنكون أمة فعالة في منظومة اقتصاد المعرفة. تاريخ الخبر ٢٠١١/١/٨. تاريخ الإتاحة [٢٠١١/٤/٢٥].- متاح في:
http://www.ict-knowledge.com/Article.aspx?ArticleId=3480

[4] About DAF.- [Cited 16/2/2010].- Available at:
http://wiki.bibalex.org/DAFWiki/index.php/Main_Page

[5] نجلاء أحمد يس أحمد. رقمنة الدوريات العربية في المكتبات المصرية: الاختيار والتحويل والتسويق / إشراف سعد محمد الهجرسي، شريف كامل شاهين.- القاهرة: جامعة القاهرة، ٢٠١٠.- أطروحة (دكتوراه) – جامعة القاهرة- كلية الآداب- قسم المكتبات والوثائق والمعلومات.- ص ٧٨، ص ص ٢٢٧-٢٣٠.

[6] DAF Team. DAFv2 Hands on Lab[power point], Workshop on Digitizing Arabic Text, 3-7 December 2007.- [Cited 16/2/2010].- Available at: http://wiki.bibalex.org/DAFWiki/images/8/8f/DAFv2HandsOnLab.ppt#383,76

[7] نجلاء أحمد يس أحمد. مصدر سابق.- ص ٧٨، ص ص ٢٢٧-٢٣٠.

[8] Kokabi, M. (1995), "The internationalization of MARC, part 1: the emergence and divergence of MARC", Library Review, Vol. 44 No. 4, pp.21-35.

[9] نجلاء أحمد يس أحمد. مصدر سابق.- ص٢٣٠.

[10] نجلاء أحمد يس أحمد.- ص٧٨، ص ٢٢٣.

[11] عمر حسن عبد الرحمن. التشغيل البيني للميتاداتا.- cybrarians journal.- ع٢١ (ديسمبر ٢٠٠٩).- تاريخ الإتاحة [٢٠١١/٣/٢٧].- متاح في:
http://www.journal.cybrarians.info/index.php?option=com...id=8...

[12] National Information Standards Organization (NISO). Understanding Metadata.- Bethesda: NISO PRESS, 2004.- p.11.- [Cited 1/7/2011].- Available at: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.

[13] Yakout, M., Adly, N., Nagi, M. "Digitization Workflow Management System for Massive Digitization Projects." In the proceedings of the 2nd International Conference on Universal Digital Library ICUDL, edited by Ismail Serageldin and Raj Reddy (Alexandria, Egypt: Bibliotheca Alexandrina,2006).- [Cited 6/7/2011].- Available at: http://www.bibalex.org/isis/UploadedFiles/Publications/Massive_Digit_Workflow_Mgmt_Sys.pdf

[14] Adly, Noha. Bibliotheca Alexandrina: A Digital Revival.- EDUCAUSE Review.- Vol. 44, No. 6.-(November/December 2009).p.8-9.- [Cited 1/7/2011].- Available at: http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume44/BibliothecaAlexandrinaADigital/185233 =EF=BB=BF

[15] Saleh, Iman, Adly, Noha and Nagi , Magdy. DAR: A Digital Assets Repository for Library Collections, 9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005), 18-23 September 2005.- pp. 6-8.- [Cited 16/2/2011].- Available at: http://www.bibalex.org/isis/UploadedFiles/Publications/DAR_1.pdf

[16] Bibliotheca Alexandrina. ICT Department—ISIS Department. DAF Manual.pp.18, 87-89.- [Cited 16/5/2011].- Available at: http://wiki.bibalex.org/DAFWiki/images/4/42/UserManual.pdf

[17] Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH). (2002).The NINCH Guide to good Practice in the Digital Representation and Management of Cultural Heritage Materials. Version 1.0 of the First Edition ed.)National Initiative for a Networked Cultural Heritage.- [Cited 30/4/2011].- Available at: http://www.nyu.edu/its/humanities/ninchguide/VIII/

[18] Adly, Noha. Million Book Project at Bibliotheca Alexandrina[power point], 2006.- [Cited 28/3/2011].- Available at:

http://www.bibalex.org/icudl06/presentation/(Noha_Adly)_Million_Book_Project_ICUDL06.ppt

[19] Saleh, Iman, Adly, Noha and Nagi , Magdy. Ibid.- p13.

[20] سامح زينهم عبد الجواد. المكتبات والأرشيفات الرقمية: التخطيط والبناء والإدارة.- القاهرة: س.ز.عبد الجواد، ٢٠٠٧.- مج٢.- ص ٤٠٨-٤٢٢.

[21] فاتن سعيد بامفلح. المكتبات الرقمية: بين التخطيط والتنفيذ. الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٨.- ص ٢٧٦.

[22] Breeding, Marshall**.** Digital preservation: building digital collections that will outlast current technologies.- Information Today, Vol. 19, No.5 (May 2002).- [Cited 23/7/2011].- Available at:
http://www.librarytechnology.org/ltg-displaytext.pl?RC=9718

[23] سامح زينهم عبد الجواد. خطط ومعايير ميتاداتا الحفظ الرقمي: دراسة تحليلية (١).- مجلة المكتبات والمعلومات العربية.- س٣٠، ع١(يناير ٢٠١٠).- ص ص ١٧٢-١٧٣.

[24] آرمز، وليم. المكتبات الرقمية؛ ترجمة جبريل بن حسن العريشي، هاشم فرحات سيد.- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٦. ص ٥٠١.

[25] Saleh, Iman, Adly, Noha and Nagi , Magdy. Ibid.- p.9.- [Cited 16/3/2010].- Available at:
http://www.bibalex.org/isis/UploadedFiles/Publications/DAR_1.pdf

[26] فاتن سعيد بامفلح. مصدر سابق.- ص٢٨٦-٢٨٨، ص٢٩٠.

[27] سامح زينهم عبد الجواد. المكتبات والأرشيفات الرقمية: التخطيط والبناء والإدارة.- مج١.- مصدر سابق.- ص ١٠٠.

[28] مكتبة الإسكندرية. التقرير السنوي (يوليو ٢٠٠٦ – يونية ٢٠٠٧). [ملف إلكتروني].- تاريخ الإتاحة.- [٢٠١٠/٣/١٥]. متاح في:
http://www.bibalex.org/English/Publication/Attachments/Files/AnnualReport_ar__2006_2007_pdf.pdf

[29] أحمد فرج أحمد. استراتيجيات رقمنة مصادر المعلومات: معايير الاختيار؛ الإشكاليات؛ الآفاق المستقبلية [الإنترنت]. النسخة ٣. Knol. ٢٠٠٩ مايو ١٢. متوفرة من خلال:
http://knol.google.com/k/dr-ahmed-farag/استراتيجيات-رقمنة-مصادر-المعلومات/٥e6arz1cii4o/23

[30] International School of Information Science. ISIS Projects[powerpoint slide].- [Cited 5/7/2011].- Available at:
http://www.bibalex.org/isis/uploadedimages/presentations/005PS_Students_Mar_2005/play.htm

[31] أشرف عبد المحسن الشريف. المتطلبات المادية والبرامجية لرقمنة السجلات.- المعلوماتية.- ع٣١.- تاريخ الإتاحة [٢٠١١/٧/٥].- متاح في: http://informatics.gov.sa/details.php?id=338

[32] Bibliotheca Alexandrina. ICT Department - ISIS Department. DAF Manual.-p.28.- [Cited 5/8/2011].- Available at: http://wiki.bibalex.org/DAFWiki/images/4/42/UserManual.pdf

[33] أشرف عبد المحسن الشريف. مصدر سابق.

[34] نجلاء أحمد يس أحمد. مصدر سابق.- ص ٢٢٦.

[35] Noha adly. Information Technology at Bibliotheca Alexandrina [presentation], 2005.- [Cited 16/4/2010].- Available at: http://www.bibalex.org/isis/UploadedImages/Presentations/000ISIS_ICT_Dec_2003/play.htm

[36] أحمد فرج أحمد. استراتيجيات رقمنة مصادر المعلومات.- مصدر سابق.

[37] صخر. القارئ الآلي: دليل الاستخدام.-تاريخ الإتاحة [٢٠١١/٧/٢].- متاح في: .http://www.sakhr.com/R_Item/datasheets/OCR.pdf

[38] Ann Arbor, MI. NovoDynamics™ and Bibliotheca Alexandrina Sign Research Partnership Agreement.-[Cited. 8/2/2010].- Available at: http://www.novodynamics.com/press/pr_bibliotecha.htm

[39] أحمد فرج أحمد. دراسات في تحليل وتصميم مصادر المعلومات الرقمية.- مجلة مكتبة الملك فهد الوطنية.- ٢٠٠٩.- ص ٣٥-٣٦.- تاريخ الإتاحة [٢٠١١/٧/١٢].- متاح في: http://www.kfnl.gov.sa/idarat/alnsher%20el/drasatfet7leel/w/%D8%A7%D9%84%D8%AF%D8%B1%D8%A7%D8%B3%D8%A9%20%D8%A7%D9%84%D8%A3%D9%88%D9%84%D9%89.doc

[40] صخر.- مصدر سابق.

[41] أحمد فرج أحمد. استراتيجيات رقمنة مصادر المعلومات.- مصدر سابق.

[42] Chopey, M.. Planning and Implementing a Metadata-Driven Digital Repositories.- Cataloguing & Classification Quarterly, Vol. 40 No.3/4,2005.- pp.255-275.

[43] أحمد فرج أحمد. دراسات في تحليل وتصميم مصادر المعلومات الرقمية.- مصدر سابق.- ص ١٠٤-١٠٥.