Standards for E-content Development

Al-Ayham Saleh

The Arab Centre for eContent Development

Table of Contents

Chapter 1
Introduction

Project Background

The eGovernment Authority of Bahrain (eGA) and the United Nations Development Programme (UNDP) have partnered to improve the quality of Arabic e-content. The Arab Centre for eContent Development was established to instill best practices and meet the need for a Knowledge Based Society (KBS) in Bahrain and the Arab region as a whole.

The Arab Centre for eContent Development has launched a project to identify standards for Arabic electronic content. The selected standards will be shared with the community to encourage developing standards compliant e-content projects.

This report documents the identified standards and specifies the recommended standards for Arabic e-content projects.

The report Team

Photo	**Al-Ayham Saleh**
	Author
	alayham@alayham.com
	Al-Ayham has been actively engaged in Arabic e-Content projects since 1999. He is currently a professional consultant and a quality Web site architect.
Photo	**Abdallah Noureddine**
	Reviewer

Methodology

The report used induction to identify and recommend the standards. The team started by researching all available regional and international standards and identifying the standards relevant to the scope of this report. The identified standards were then analyzed and a subset of them was accepted as Arabic e-content standards. Furthermore, a subset of the accepted e-content standards has been selected as recommended for use in Arabic e-content projects.

Selection Criteria

For a standard document to be included in this report, the following conditions must be met:

- The standard document must be directly related to electronic content. Technical standards (such as how to understand a standard document) and protocols (such as the TCP/IP family of protocols) have been excluded because they are not directly related to e-content.
- The standard document must be directly applicable in an e-content project. Generic standards that provide a base for building other higher level standards are excluded from this report. For example, the SGML and XML languages are excluded, and other high level standards based on them are included.

For a standard to be accepted, the following conditions must be met:

- A standard must be reviewed and accepted by the developer community. A standard document is not accepted simply because it has been issued by a government entity.

Standards must be adopted by a wide audience of e-content builders before consideration by the Centre.
- The standard document must provide valid justification for its norms. If the standard document lacks clear and unbiased justifications, it will not be accepted.
- The standard must not be deprecated by another accepted standard document.

For a standard document to be recommended, the following conditions must be met:

- A standard must be an accepted standard.
- A standard must provide a direct value to the e-content project.
- A standard must not contradict with another recommended standard.

Chapter 2
Understanding Standards

Definitions

The following terms are used in this document:

Content: information and experiences that may provide value for an end-user/audience in specific contexts[1]

E-content: Electronic or digital content delivered over a network such as the internet.

Standard: A set of normative specifications.

Standard Document: A document that describes a published standard.

E-content Standards: Standards dealing with electronic content.

Understanding standards

E-content standards are rules that most e-content project comply to. Standards compliance is voluntary, and an e-content project might or might not comply with a certain standard. The standard compliance decision is owned by the e-content project owner.

Standards are engineering or technical specifications that help projects establish uniformity. There are different kinds of standards:

- Mandatory: compliance is required;
- Voluntary: compliance is optional;
- De facto: may not be formally published but is a widely used and well established common practice;
- De jure: formal legal requirements.

This report focuses on the electronic standards published by formal standards bodies such as the World Wide Web Consortium (W3C), the Internet Engineering Task Force (IETF), and International Organization for Standardization (ISO). E-content standards compliance is voluntary and the decision to comply must be made by the project owner.

Compliance with e-content standards increases the value of the project. This increase in value depends on the nature of the project and on how the project can benefit by complying with the standards.

As e-content projects seek to increase their value through standards compliance and, to assist business owners with this critical aspect of their decision making, the team researched content-related standards including the most widely used and internationally adopted standards as well as regional and local standards and guidelines. Project owners are encouraged to review these recommendations, determine the benefits the project can gain from complying with these

---

[1]http://en.wikipedia.org/wiki/Content_%28media%29

standards, and identify the implementation methodology that will achieve optimum compliance and benefits.

The individual choices made by each and every project to meet the specific needs at certain times as it relates to standards compliance must be respected; however, we do believe that compliance with the recommended set of standards will bring benefits that the project cannot acquire without such standards compliance.

Standard Bodies

Standards are generally designed and published by internationally acknowledged bodies. As a result, the majority of the standards identified in this document were established by the following organizations: the World Wide Web Consortium (W3C); Internet Engineering Task Force (IETF); International Organization for Standards (ISO); and ECMA International.

The World Wide Web Consortium

The World Wide Web Consortium (W3C) is the main international standards organization for the World Wide Web (abbreviated WWW or W3)[2].

W3C standards are issued as "recommendations". The World Wide Web Consortium Process Document[3] clarifies the maturity levels of W3C recommendations and other documents in their development process. A technical document in the W3C process starts as a Working Draft (WD). Some WDs advance to the W3C Recommendation Track, where there are three maturity levels[4]:

1. Candidate Recommendation (CR)
2. Proposed Recommendation (PR)
3. W3C Recommendation (REC)

The W3C internationalization (i18n) activity is the interest group that works to make the web usable in multiple languages.

The Internet Engineers Task Force (IETF)

The internet engineering task force (IETF) develops and promotes internet standards and protocols. All members and participants in the IETF activities are volunteers or funded by other parties.

An IETF approved standard has 3 maturity levels[5]:

1. Proposed Standard.
2. Draft Standard.
3. Internet Standard.

Internet standards are published in "Request for Comment" (RFC) documents. RFCs can update or obsolete other previous RFCs.

---

[2]http://en.wikipedia.org/wiki/W3c
[3]http://www.w3.org/2005/10/Process-20051014/
[4]http://www.w3.org/2005/10/Process-20051014/tr.html#q74
[5]http://tools.ietf.org/html/rfc2026#section-4.1

The International Organization for Standards (ISO)

ISO is composed of representatives of various national standards organizations. ISO standards are sometimes enforced by national laws or regulations in some countries.

ISO standards are developed in a lengthy six-stage process:

1. Proposal stage
2. Preparatory stage
3. Committee stage
4. Enquiry stage
5. Approval stage
6. Publication stage

ISO standards can be reviewed or withdrawn after publication.

Ecma International

Ecma International is a membership-based standards organization. Ecma International was established by the European Computer Manufacturer Association and then later became an internal standards organization.

Chapter 3
The Current Status of Arabic E-content Standards

Although many initiatives have been launched to improve the situation of Arabic e-content, the team was not able to locate information on initiatives to standardize Arabic e-content. This report focuses on the standards and not on existing initiatives.

During the research, the team identified only one document that attempted to standardize e-content projects for media sites.

The Syrian Standards for Official Media Websites

Status	Established by the Ministry of Information and issued under declaration No. 42 on 6 February 2008.
Approving Body	Ministry of Information, Syria
Reference	http://acnc.sy/site/acnc/index.php?path=cat-index/reference-documents/274&node-details=1
Description	The document states that these selected standards are built on international standards and local requirements. The selected standards present the preferred model for media Websites on the internet.
Advantages	• Includes operational standards, such as a sample project plan and a suggested team structure • Includes a set of standard features • References the W3C standards
Disadvantages	• Very high level standards in some areas. • Do not take technical standards into consideration. • Focus on official media Websites, not on Arabic content.
Arabic support	• The document is written in Arabic • The standard emphasizes support for Arabic language features.
Notes	• The standard does not justify its recommendations. For example, a recommendation such as "avoid using technologies like JavaScript, Flash, and Dreamweaver" does not provide any reason for avoiding such technologies. • The standard provides a good starting point for media Websites.
Recommendation	• We do not accept this standard as an Arabic content standard in its present form.

Chapter 4
Technology Standards

Document Format Standards

Generally, e-content is presented in documents. The format of e-content documents might follow a set of standards.

HTML 4.01

Status	W3C Recommendation 24 December 1999
Approving Body	W3C
Reference	http://www.w3.org/TR/1999/REC-html401-19991224/
Description	This specification defines the HyperText Markup Language (HTML), the publishing language of the World Wide Web. This specification defines HTML 4.01, which is a subversion of HTML 4. In addition to the text, multimedia, and hyperlink features of the previous versions of HTML (HTML 3.2 [HTML32] and HTML 2.0 [RFC1866]), HTML 4 supports more multimedia options, scripting languages, style sheets, better printing facilities, and documents that are more accessible to users with disabilities. HTML 4 also takes great strides towards the internationalization of documents, with the goal of making the Web truly World Wide. HTML 4 is an SGML application conforming to International Standard ISO 8879 -- Standard Generalized Markup Language [ISO8879].
Advantages	• Widely adopted, almost all user agents in the world can render html 4.01 documents correctly[6].
Disadvantages	• Allows mal-formed documents[7] • Not flexible: other standards provide more flexibility for content authors. • Does not fully support rich content: Browsers had to implement custom plug-ins to support some basic rich content tags such as <marquee> • Does not fully support the latest semantic standards such as RDFa.
Arabic support	• HTML 4.01 supports right to left documents using the "dir" directive. • HTML 4.01 supports multiple Arabic charsets, including the widely used windows-1256 Arabic charset. • There is no Arabic translation of this standard.
Notes	• There is a considerable Arabic e-content presented in charset windows-1256 using HTML4.01.
Recommendation	• We accept HTML 4.01 as an Arabic Web standard. • We do not recommend using HTML4.01 as a document format for new content. • We recommend migrating HTML4.01 documents to the XHTML1.1 format

---

[6]http://www.webdevout.net/browser-support-html
[7] Refer to the definition of well-formed documents in the XML specification http://www.w3.org/TR/REC-xml/#sec-well-formed

XHTML™ 1.0 Second Edition

Status	W3C Recommendation 26 January 2000, revised 1 August 2002
Approving Body	W3C
Reference	http://www.w3.org/TR/2002/REC-xhtml1-20020801/
Description	This specification defines the Second Edition of XHTML 1.0, a reformulation of HTML 4 as an XML 1.0 application, and three DTDs corresponding to the ones defined by HTML 4. The semantics of the elements and their attributes are defined in the W3C Recommendation for HTML 4. These semantics provide the foundation for future extensibility of XHTML. Compatibility with existing HTML user agents is possible by following a small set of guidelines.
Advantages	• Extensible and flexible document format standard. • Backward compatible with HTML4.01 • Wide browser support.[8]
Disadvantages	• The use of utf-8 or utf-16 is strongly recommended. Some processors do not support other charsets. • This standard is not modular by nature, although it can be extended.
Arabic support	• XHTML 1.0 maintained the internationalization features of HTML 4.01, including the directionality and the language attributes. • There is no Arabic translation of this standard.
Notes	• The differences between XHTML 1.0 and XHTML 1.1 are minor.
Recommendation	• We accept XHTML 1.0 as an Arabic Web standard. • We don't recommend using XHTML 1.0 as a document format for new content.

XHTML™ 1.1 - Module-based XHTML

Status	W3C Recommendation 31 May 2001
Approving Body	W3C
Reference	http://www.w3.org/TR/2001/REC-xhtml11-20010531/
Description	This Recommendation defines a new XHTML document type that is based upon the module framework and modules defined in Modularization of XHTML [XHTMLMOD]. The purpose of this document type is to serve as the basis for future extended XHTML 'family' document types, and to provide a consistent, forward-looking document type cleanly separated from the deprecated, legacy functionality of HTML 4 [HTML4] that was brought forward into the XHTML 1.0 [XHTML1] document types. This document type is essentially a reformulation of XHTML 1.0 Strict using XHTML Modules. This means that many facilities available in other XHTML Family document types (e.g., XHTML Frames) are not available in this document type. These other facilities are available through modules defined in Modularization of XHTML, and document authors are free to define document types based upon XHTML 1.1 that use these facilities (see [XHTMLMOD] for information on creating new document types).
Advantages	• Highly modular and extensible. • Modernized, with extensive research on developing it. • Rich semantic support
Disadvantages	• Does not support deprecated elements in XHTML 1.0 or HTML4.0. So migration may not be as easy as it was to XHTML 1.0.

---

[8]http://www.webdevout.net/browser-support-html

Arabic support	• XHTML 1.1 removed the deprecated HTML 4.01 "lang" attribute in favor of the modular "xml:lang" attribute. • There is no Arabic translation of this standard.
Notes	• XHTML 1.1 has many useful extensions. Some of them are valuable to new Arabic e-content projects.
Recommendation	• We accept XHTML 1.1 as an Arabic Web standard. • We recommend using XHTML 1.1 in building new Arabic content.

The HTML5 Working Draft

The HTML 5 specification is still in the working draft status. Many user agents lack support for HTML5 tags and features. If in the future, HTML5 becomes a standard specification, we will evaluate HTML5 and make our recommendation in relation to Arabic content standards.

The latest HTML5 working draft is available at http://www.w3.org/TR/2010/WD-html5-20100624/

Client Side Scripting Standards

Document Object Model (DOM) Level 2 Core Specification

Status	W3C Recommendation 13 November 2000
Approving Body	W3C
Reference	http://www.w3.org/TR/2000/REC-DOM-Level-2-Core-20001113/
Description	This specification defines the Document Object Model Level 2 Core, a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content and structure of documents. The Document Object Model Level 2 Core builds on the Document Object Model Level 1 Core. The DOM Level 2 Core is made of a set of core interfaces to create and manipulate the structure and contents of a document. The Core also contains specialized interfaces dedicated to XML.
Advantages	• Most of the standard is supported across most user agents[9]. • Many standardized extensions are supported[10].
Disadvantages	• Some internal algorithms are not easy to understand and implement. • Nested documents are not easy to manage. • Does not support entity declarations making it incompatible with some XML documents.
Arabic support	• Right to left text manipulation is supported. • There is no Arabic translation of this standard.
Notes	Users may test their user agent support for DOM levels by accessing the following page: http://www.w3.org/2003/02/06-dom-support.html
Recommendation	• We accept DOM2 Core as an Arabic content standard. • We recommend using DOM2 as the Document Object Model for scripting new documents.

[9] http://en.wikipedia.org/wiki/Comparison_of_layout_engines_%28Document_Object_Model%29
[10] http://www.webdevout.net/browser-support-dom

Document Object Model (DOM) Level 3 Core Specification

Status	W3C Recommendation 07 April 2004
Approving Body	W3C
Reference	http://www.w3.org/TR/2004/REC-DOM-Level-3-Core-20040407/
Description	This specification defines the Document Object Model Core Level 3, a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents. The Document Object Model Core Level 3 builds on the Document Object Model Core Level 2 [DOM Level 2 Core]. This version enhances DOM Level 2 Core by completing the mapping between DOM and the XML Information Set [XML Information Set], including the support for XML Base [XML Base], adding the ability to attach user information to DOM Nodes or to bootstrap a DOM implementation, providing mechanisms to resolve namespace prefixes or to manipulate "ID" attributes, giving to type information, etc.
Advantages	• Improved algorithms over DOM2
Disadvantages	• Only partial support in some user agents[11].
Arabic support	• Right to left text manipulation is supported. • There is no Arabic translation of this standard.
Notes	Users may test their user agent support for DOM levels by accessing the following page: http://www.w3.org/2003/02/06-dom-support.html
Recommendation	• We accept DOM3 Core as an Arabic content standard. • We do not recommend using DOM3 features until they are supported in major mainstream user agents.

ECMAScript 262

Status	ISO/IEC 16262
Approving Body	ISO
Reference	http://www.ecma-international.org/publications/standards/Ecma-262.htm
Description	This Standard defines the ECMAScript scripting language. ECMAScript is the base language of many scripting implementations, such as Mozilla's JavaScript Engine and Microsoft's JScript Engine.
Advantages	• Most of the ECMAScript standard is supported by most browsers.
Disadvantages	• ECMAScript is the specification, not the language. Existing implementations for ECMAScript might not provide 100% ECMAScript compliance.
Arabic support	• ECMAScript supposes that each character is encoded using 16 bits. Many Arabic pages are still in single byte encodings like windows-1256. ECMAScript implementations such as JavaScript or JScript might perform string normalization in some cases but there is no guarantee that single byte strings are fully supported by ECMAScript implementations. • There is no Arabic translation of this standard.
Notes	JavaScript is a good ECMAScript implementation.
Recommendation	• We accept ECMAScript as an Arabic Web standard. • We recommend using ECMAScript compliant scripts when scripting documents.

---

[11] http://en.wikipedia.org/wiki/Comparison_of_layout_engines_%28Document_Object_Model%29

Protocol Standards

The Platform for Privacy Preferences (P3P1.0) Specification

Status	W3C Recommendation 16 April 2002
Approving Body	W3C
Reference	http://www.w3.org/TR/2002/REC-P3P-20020416/
Description	This is the specification of the Platform for Privacy Preferences (P3P). This document, along with its normative references, includes all the specification necessary for the implementation of interoperable P3P applications.
Advantages	• Standardize the declaration of some privacy practice policies
Disadvantages	• Does not guarantee that the privacy policy is implemented.
Arabic support	• P3P policies use an XML namespace; therefore, support for Arabic text inside the policies is available. • There is no Arabic translation of this standard.
Notes	• This standard is supported by most browsers, but not widely used.
Recommendation	• We accept the P3P1.0 as an Arabic Web standard. • We recommend P3P1.0 compliance in all new content.

Simple Knowledge Organization System (SKOS)

Status	W3C Recommendation 18 August 2009
Approving Body	W3C
Reference	http://www.w3.org/TR/2009/REC-skos-reference-20090818/
Description	This document defines the Simple Knowledge Organization System (SKOS), a common data model for sharing and linking knowledge organization systems via the Web. Many knowledge organization systems, such as thesauri, taxonomies, classification schemes and subject heading systems, share a similar structure, and are used in similar applications. SKOS captures much of this similarity and makes it explicit, to enable data and technology sharing across diverse applications. The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organization systems. It may be used on its own, or in combination with formal knowledge representation languages such as the Web Ontology Language (OWL).
Advantages	• Facilitates sharing structured knowledge.
Disadvantages	• The modeling required to develop useful web ontologies may be demanding in terms of expertise, effort, and cost.
Arabic support	• There is no Arabic translation of this standard.
Notes	•
Recommendation	• We accept the SKOS as an Arabic Web standard.

The Sitemap Protocol

Status	Sitemap 0.90 is a de facto standard with industry-wide adoption.
Approving Body	Google, Yahoo!, and Microsoft
Reference	http://sitemaps.org/

Description	Sitemaps are an easy way for webmasters to inform search engines about pages on their sites that are available for crawling. In its simplest form, a Sitemap is an XML file that lists URLs for a site along with additional metadata about each URL (when it was last updated, how often it usually changes, and how important it is, relative to other URLs in the site) so that search engines can more intelligently crawl the site.
	Web crawlers usually discover pages from links within the site and from other sites. Sitemaps supplement this data to allow crawlers that support Sitemaps to pick up all URLs in the Sitemap and learn about those URLs using the associated metadata. Using the Sitemap protocol does not guarantee that web pages are included in search engines, but provides hints for web crawlers to do a better job of crawling a site.
Advantages	• De facto standard used by major search engines
Disadvantages	• Not approved by an international standard body
Arabic support	• The standard is based on XML; therefore, it has the same Arabic support features of XML. • There is no Arabic translation of this standard.
Notes	• XML sitemaps are different from the general use sitemaps, used typically in small websites.
Recommendation	• We accept the sitemaps.org as an Arabic web standard. • We recommend including a sitemaps.org compliant sitemap on every e-content portal.

The Really Simple Syndication (RSS) 2.0.11 Protocol

Status	Version 2.0.11 of the Really Simple Syndication (RSS) 2.0 specification, published by the RSS Advisory Board on March 30, 2009
Approving Body	RSS Advisory Board
Reference	http://www.rssboard.org/rss-specification
Description	RSS is a Web content syndication format. RSS is a dialect of XML. All RSS files must conform to the XML 1.0 specification, as published on the World Wide Web Consortium (W3C) Website.
Advantages	• Wide support by many user agents
Disadvantages	• Does not allow specifying the content type in the feed.
Arabic support	• Does not support mixed languages in feed items; each feed should have its own language. • There is no Arabic translation of this standard.
Notes	• RSS can be used to syndicate feeds and/or to aggregate feeds. RSS is not recommended for syndicating feeds.
Recommendation	• We accept the RSS 2.0.11 as an Arabic internet standard. • We do not recommend using RSS 2.0.11.We recommend using other standards.

The Atom Syndication Format

Status	RFC4287 December 2005
Approving Body	The Internet Society
Reference	http://tools.ietf.org/html/rfc4287

Description	Atom is an XML-based document format that describes lists of related information known as "feeds". Feeds are composed of a number of items, known as "entries", each with an extensible set of attached metadata. For example, each entry has a title. The primary use case that Atom addresses is the syndication of Web content such as weblogs and news headlines to Web sites as well as directly to user agents.
Advantages	• Wide support across major user agents
Disadvantages	• Atom is a very strict standard because it is based on XML. Therefore, it is easy to make mistakes while building the feeds.
Arabic support	• Supports the xml:lang attribute on every element. • There is no Arabic translation of this standard.

Notes	• Although RSS is the most used protocol for syndicating content, ATOM is more appropriate for syndicating multi-language content.
	• Not to be confused with the Atom Publishing Protocol (RFC5023)
	• Atom 1.0 has many useful extensions:
	    o RFC 4685 - Atom Threading Extensions
	    o RFC 4946 - Atom License Extension
	    o RFC 5005 - Feed Paging and Archiving

Recommendation	• We accept the ATOM 1.0 protocol as an Arabic internet standard. • We recommend using the ATOM 1.0 protocol for syndicating Arabic content.

Chapter 5
Content Design Standards

Cascading Style Sheets Level 2 (CSS2)

Status	CSS 2.0 is a W3C Recommendation 12-May-1998 (revised 11 April 2008) CSS 2.1 is a W3C Candidate Recommendation 08 September 2009
Approving Body	W3C
Reference	CSS 2.0: http://www.w3.org/TR/2008/REC-CSS2-20080411/ CSS 2.1: http://www.w3.org/TR/CSS21/
Description	This specification defines Cascading Style Sheets, level 2 (CSS2). CSS2 is a style sheet language that allows authors and users to attach style (e.g., fonts, spacing, and aural cues) to structured documents (e.g., HTML documents and XML applications). By separating the presentation style of documents from the content of documents, CSS2 simplifies Web authoring and site maintenance.
Advantages	• Wide support across most user agents[12]
Disadvantages	• Lacks some advanced styling features such as dynamic sizes.
Arabic support	• Supports right to left text. • There is no Arabic translation of this standard.
Notes	• The CSS Working Group is developing CSS Level 2 Revision 1, which corrects many errors and omissions in this document as well as making a few other changes as documented in the changes section. The CSS Working Group encourages authors and implementers to reference CSS 2.1 (or its successor) instead of this document and when features common to CSS2 and CSS 2.1 are defined differently to follow the definitions in CSS 2.1. • CSS3 is still under development, and it has not been formulated as a complete draft recommendation yet.
Recommendation	• We accept CSS2.1 as an Arabic content standard. • We recommend using CSS2.1 in styling new content.

Portable Network Graphics (PNG) Specification (Second Edition)

Status	W3C Recommendation 10 November 2003 ISO/IEC 15948:2003 (E)
Approving Body	W3C and ISO
Reference	http://www.w3.org/TR/2003/REC-PNG-20031110/

Commented [u1]: Needs to be revised to reflect current W3C recommendations.

---

[12]http://www.webdevout.net/browser-support-css

Description	This document describes PNG (Portable Network Graphics), an extensible file format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF and can also replace many common uses of TIFF. Indexed-color, grayscale, and true color images are supported as well as optional alpha channel. Sample depths range from 1 to 16 bits. PNG is designed to work well in online viewing applications, such as the World Wide Web allowing it to be fully streamed with a progressive display option. PNG is robust, providing both full file integrity checking and simple detection of common transmission errors. In addition, PNG can store gamma and chromaticity data for improved color matching on heterogeneous platforms.
Advantages	• Patent-free file format. • Supports transparency. • Can replace the patented gif file format.
Disadvantages	• Larger file size when compared to gif or JPG. • Does not support animation (like GIF).
Arabic support	• Some PNG design software can produce PNG images with Arabic text. • There is no Arabic translation of this standard.
Notes	• PNG is a lossless file format, so the file size may be large, but image details are retained after saving.
Recommendation	• We accept the PNG as an Arabic content standard. • We recommend using the PNG for developing new Arabic e-content.

Scalable Vector Graphics (SVG) 1.1 Specification

Status	W3C Recommendation 14 January 2003, edited in place 30 April 2009
Approving Body	W3C
Reference	http://www.w3.org/TR/2003/REC-SVG11-20030114/
Description	This specification defines the features and syntax for Scalable Vector Graphics (SVG) Version 1.1, a modularized language for describing two-dimensional vector and mixed vector/raster graphics in XML.
Advantages	• Smaller file size • Easier to manipulate on a Web page. • Can be created dynamically.
Disadvantages	• Limited toolsets for creating and manipulating SVG • SVG implementation requires a learning curve. • SVG cannot represent bitmaps, but it can embed them.
Arabic support	• There is no Arabic translation of this standard.
Notes	• SVG is a vector graphic, which makes it resolution independent. • SVG is a very good choice to represent dynamic charts on the web.
Recommendation	• We accept SVG as an Arabic content standard.

Chapter 6
Content Standards

International Content standards

Resource Description Framework–in–attributes (RDFa) in XHTML

Status	W3C Recommendation 14 October 2008
Approving Body	W3C
Reference	http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014
Description	RDFa is a specification for attributes to express structured data in any markup language. This document specifies how to use RDFa with XHTML. The rendered, hypertext data of XHTML is reused by the RDFa markup, so that publishers do not need to repeat significant data in the document content. The underlying abstract representation is RDF [RDF-PRIMER], which lets publishers build their own vocabulary, extend others, and evolve their vocabulary with maximal interoperability over time. The expressed structure is closely tied to the data, so that rendered data can be copied and pasted along with its relevant structure. The rules for interpreting the data are generic, so that there is no need for different rules for different formats; this allows authors and publishers of data to define their own formats without having to update software, register formats via a central authority, or worry that two formats may interfere with each other.
Advantages	• Allows embedding standardized metadata. • Uses RDF for data representation. RDF is a generally adopted standard.
Disadvantages	• Designed for coding. Might cause problems with editing software, especially with Copy-Paste operations.
Arabic support	• There is no Arabic translation of this standard.
Notes	• The RDFa Primer document provides a straight forward explanation on how to use RDFa in XHTML. http://www.w3.org/TR/2008/NOTE-xhtml-rdfa-primer-20081014/ • Refer to http://rdfa.info for more information on RDFa, including an RDFa checker.
Recommendation	• We accept RDFa as an Arabic web standard. • We recommend using RDFa when building new content • We recommend migrating old Arabic content to include RDFa features.

The Dublin Core Metadata Element Set

Status	ISO 15836:2009
Approving Body	ISO
Reference	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52142 http://dublincore.org/specifications/

Description	The Dublin Core Metadata Element Set, commonly referred to as Dublin Core, provides 15 core properties[13] used to include semantic metadata information in XHTML documents.
Advantages	• The Dublin Core Metadata Element Set is a very simple way to start embedding metadata in e-content. • The Dublin Core Metadata Element Set supports embedding metadata in legacy HTML documents. • The Dublin Core Metadata Element Set allows extensions.
Disadvantages	• Defined properties are limited. Complex use cases must extend the Dublin Core properties.
Arabic support	• There is no Arabic translation of this standard.
Notes	• Refer to http://dublincore.org/ for more information on Dublin Core.
Recommendation	• We accept the Dublin Core as an Arabic content standard. • We recommend using the Dublin Core and RDFa in building new content. • We recommend adding Dublin Core and RDFa features to existing content.

Web Content Accessibility Guidelines (WCAG) 2.0

Status	W3C Recommendation 11 December 2008
Approving Body	W3C
Reference	http://www.w3.org/TR/2008/REC-WCAG20-20081211/
Description	Web Content Accessibility Guidelines (WCAG) 2.0 covers a wide range of recommendations for making Web content more accessible. Following these guidelines will make content accessible to a wider range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech disabilities, photosensitivity and combinations of these. Following these guidelines will also often make your Web content more usable to users in general.
Advantages	• WCAG was built to be technology independent, and to focus on the accessibility results. • WCAG covers usability best practices, not just technology aspects.
Disadvantages	• WCAG is not completely measurable. Some tests require qualified auditing, and are subjective. • WCAG is general in nature and applies to various technologies, but in certain instances, it is very difficult for specific technologies to comply with the guidelines.
Arabic support	• There is no Arabic translation of this standard.
Notes	• The standard is accompanied with many supporting documents, the major supporting documents are highlighted in this diagram:

---

[13] The properties are: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifiers, Source, Language, Relation, Coverage, Rights.

Recommendation	• We accept WCAG 2.0 as an Arabic web standard. • We recommend compliance with the WCAG2.0 standard for new and existing content.

Web Ontology Language (OWL) 2

Status	W3C Recommendation 27 October 2009
Approving Body	W3C
Reference	http://www.w3.org/TR/2009/REC-owl2-overview-20091027/
Description	The Web Ontology Language (OWL) 2, informally OWL 2, is an ontology language for the Semantic Web with formally defined meaning. OWL 2 ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL 2 ontologies can be used along with information written in RDF, and OWL 2 ontologies themselves are primarily exchanged as RDF documents.
Advantages	• Provides a standard for sharing knowledge in multiple languages or domains
Disadvantages	• New standard, not widely supported by user agents.
Arabic support	• There is no Arabic translation of this standard.
Notes	• Ontologies are essential to represent and share knowledge on the internet.
Recommendation	• We accept the OWL standard as an Arabic content standard. • We recommend using OWL to build Arabic ontologies that will allow sharing Arabic knowledge.

UWEM 1.0 Accessibility Standard

Status	UWEM 1.0 Industry standard published on 14 March 2007 by AnySurfer.be
Approving Body	AnySurfer
Reference	http://www.anysurfer.be/fr/apropos-de-anysurfer/un-site-accessible/lois-et-standards/anysurfer-et-uwem-1.0/index.html
Description	Includes 146 tests. 54 tests are mapped to WCAG 1.0 priority 1 and 92 are mapped to Priority 2.
Advantages	• Made especially for impaired people.
Disadvantages	• Very strict standard (example: ASCII art is not allowed, flash is not allowed for navigation)
Arabic support	• Allows multi-language pages, but imposes identification of the principal language of a page. • There is no Arabic translation of this standard.

Notes	• Any surfer is specific to the internet technology, which makes compliance checking easier than WCAG.
Recommendation	• We accept this standard as a strict accessibility standard.
	• We do not recommend using this standard unless the site is especially designed for impaired people.

SPARQL Protocol for RDF

Status	W3C Recommendation 15 January 2008
Approving Body	W3C
Reference	http://www.w3.org/TR/2008/REC-rdf-sparql-protocol-20080115/
Description	The SPARQL Protocol and RDF Query Language (SPARQL) is a query language and protocol for RDF. This document specifies the SPARQL Protocol; it uses WSDL 2.0 to describe a means for conveying SPARQL queries to an SPARQL query processing service and returning the query results to the entity that requested them.
Advantages	• Provides a standard for querying knowledge databases for knowledge items.
Disadvantages	• Not widely known, and applications are still limited
Arabic support	• There is no Arabic translation of this standard.
Notes	• SPARQL is a relatively new technology and has not been widely adopted yet.
Recommendation	• We accept the SPARQL standard as an Arabic content standard.
	• We recommend building Arabic knowledge bases that support the SPARQL protocol.

Arabic content standards

Internationalization Tag Set (ITS) Version 1.0

Status	W3C Recommendation 03 April 2007
Approving Body	W3C
Reference	http://www.w3.org/TR/2007/REC-its-20070403/
Description	This document defines data categories and their implementation as a set of elements and attributes called the Internationalization Tag Set (ITS). ITS is designed to be used with schemas to support the internationalization and localization of schemas and documents. An implementation is provided for three schema languages: XML DTD, XML Schema and RELAX NG.
Advantages	• Standardizes writing Arabic and non-Arabic text in a single document
	• Standardizes the source-translation relationship in e-content.
Disadvantages	• Does not solve all translation/localization issues in XML.
Arabic support	• Supports bi-directional text
	• Supports accents for Arabic text.
	• There is no Arabic translation of this standard.
Notes	• There is a working draft on ITS best practices http://www.w3.org/TR/2007/WD-xml-i18n-bp-20070427/
Recommendation	• We accept ITS 1.0 as an Arabic web standard.
	• We recommend using ITS as a standard for building international content.

Character Model for the World Wide Web 1.0: Fundamentals

Status	W3C Recommendation 15 February 2005
Approving Body	W3C
Reference	http://www.w3.org/TR/2005/REC-charmod-20050215/
Description	This Architectural Specification provides authors of specifications, software developers, and content developers with a common reference for interoperable text manipulation on the World Wide Web, building on the Universal Character Set, defined jointly by the Unicode Standard and ISO/IEC 10646. Topics addressed include use of the terms 'character', 'encoding' and 'string', a reference processing model, choice and identification of character encodings, character escaping, and string indexing.
Advantages	• Set the common standard between developers and content authors on how to specify the display of characters on the web
Disadvantages	• Not widely supported by user agents and developers.
Arabic support	• Supports rendering right to left characters • Supports the visual selection of Arabic text[14] • There is no Arabic translation of this standard.
Notes	• This standard is very important for building truly multi-lingual user agents.
Recommendation	• We accept the Character Model for the World Wide Web 1.0: Fundamentals as an Arabic web standard. • We recommend using the Character Model for the World Wide Web 1.0: Fundamentals for building software that support logical and visual Arabic text features.

---

[14]http://www.w3.org/TR/2005/REC-charmod-20050215/#sec-VisualRenderingUnits

Chapter 7
Summary and Recommendations

Standards Summary

List of Standards that are not Accepted
1. The Syrian Standards for Official Media Websites
2. The HTML5 Working Draft

List of Standards that are Accepted, but not Recommended
1. HTML 4.01
2. XHTML™ 1.0
3. Document Object Model (DOM) Level 3 Core Specification
4. The Really Simple Syndication (RSS) 2.0.11 Protocol
5. UWEM 1.0 Accessibility Standard

List of Recommended Standards
1. XHTML™ 1.1 - Module-based XHTML
2. Document Object Model (DOM) Level 2 Core Specification
3. ECMAScript 262
4. The Platform for Privacy Preferences (P3P1.0) Specification
5. Simple Knowledge Organization System
6. The Sitemap Protocol
7. The Atom Syndication Format
8. Cascading Style Sheets Level 2 (CSS2)
9. Portable Network Graphics (PNG) Specification (Second Edition)
10. Scalable Vector Graphics (SVG) 1.1 Specification
11. Resource Description Framework–in–attributes (RDFa) in XHTML
12. The Dublin Core Metadata Element Set
13. Web Content Accessibility Guidelines (WCAG) 2.0
14. Web Ontology Language
15. SPARQL Protocol for RDF
16. Internationalization Tag Set (ITS) Version 1.0
17. Character Model for the World Wide Web 1.0: Fundamentals

Recommendations Going Forward

With this document, the team set a baseline by selecting a set of international standards that, when implemented together, will greatly improve the quality of Arabic e-content and increase collaboration efforts within the Arabic e-content development community.

To increase the value of the centre and further assist the e-content community, we recommend the following initiatives:

Establish a Contact Channel with the W3C Internationalization Activity

The W3C Internationalization (I18n) Activity works to ensure universal access to the World Wide Web. The centre can consider opening a communication channel with one of the groups under the

W3C I18n Activity to collaborate and support their work and simultaneously benefit from their results.

Establish an Online Standards Knowledge Base

A standards knowledge base can be developed to include the standards identified above as well as other standards that did not fall in the scope of this report. Registered members can search for standards, discuss them and exchange ideas or best practices with other members. In addition, members can suggest other standards for inclusion in the knowledge base.

Offer Specific Services for E-content Developers

The centre can start offering some services for e-content developers. These services can be completely automated. Here are a few examples:

1. Markup and code validation.
2. DOM Support testing.
3. Usability testing.

Invest in Building Arabic Fonts for the Web

There is a real need for truly scalable Arabic fonts that can be used in e-content development. Currently, e-content projects do not have enough options when it comes to producing readable text on web pages. For Arabic e-content to thrive online, this need must be addressed and new fonts must be designed for the Web.

Conclusion

This report identified 17 recommended standards for building Arabic e-content. Based on the research performed while preparing this report, this effort appears to be the first in its field. The results of this research can hugely benefit e-content projects and should be shared with the community in the most appropriate way.