



**VERSION ENRICHIE DU
STANDARD SUR LES SITES WEB
MULTILINGUES
(SGQRI 011)**

Version 1.0 du 11 décembre 2006

Table des matières

SECTION I : DISPOSITIONS GÉNÉRALES.....	1
S.-s. 1 – Objet du standard.....	1
S.-s. 2 – Champ d’application.....	1
S.-s. 3 – Définitions.....	2
SECTION II : SPÉCIFICATIONS.....	2
S.-s. 1 – Conditions de conformité au standard.....	2
S.-s. 2 – Accès au site.....	2
S.-s. 3 – Sélection de la langue de consultation.....	3
S.-s. 4 – Exigences techniques.....	5
SECTION III : DISPOSITIONS TRANSITOIRES ET FINALES.....	12
S.-s. 1 – Mesures transitoires.....	12
S.-s. 2 – Révision.....	13
S.-s. 3 – Date d’entrée en vigueur.....	13
RENSEIGNEMENTS COMPLÉMENTAIRES.....	14
R.C. 1 – Autres sigles et définitions.....	14
R.C. 2 – Références bibliographiques.....	15
R.C. 3 – Dérogation aux autres standards du gouvernement du Québec.....	17
R.C. 4 – Conformité au concept d’adaptabilité culturelle et linguistique.....	17
R.C. 5 – Composition du groupe de travail responsable de l’élaboration du standard.....	17

Remarque :

Élaboré par le ministère des Services gouvernementaux, le standard adopté par le Conseil du trésor le 11 décembre 2006 se trouve dans le *Recueil des politiques de gestion* du Conseil du trésor (RPG 11 2 2 13). Ce document porte exclusivement sur les éléments obligatoires pour les ministères et les organismes.

Afin de faciliter la mise en place de ce standard dans l’Administration gouvernementale, le ministère des Services gouvernementaux rend disponible une version enrichie, à la manière d’une version annotée, dans le *Recueil des éléments normatifs en matière de ressources informationnelles*. Ce document reprend le contenu du standard adopté par le Conseil du trésor en y ajoutant des renseignements supplémentaires pertinents et d’autres éléments qui ne sont pas de nature obligatoire.

Les renseignements supplémentaires dans cette version enrichie sont présentés dans des encadrés en grisé et peuvent porter, notamment sur une mise en contexte, un exemple, une recommandation, une remarque, une déclaration sur la conformité ou sur la dérogation de ce standard à d’autres standards du gouvernement du Québec, ou une déclaration sur la conformité au concept d’adaptabilité culturelle et linguistique. Ils indiquent finalement la composition du groupe de travail responsable de l’élaboration du standard.

VERSION ENRICHIE DU STANDARD SUR LES SITES WEB MULTILINGUES (SGQRI 011)

SECTION I : DISPOSITIONS GÉNÉRALES

S.-s. 1 – Objet du standard

1. Ce standard énonce les exigences minimales relatives à l'organisation et à la gestion d'un site Web multilingue.

Remarque :

Les dispositions du présent standard quant à l'organisation et la gestion des sites Web multilingues s'appliquent à tous les sites gouvernementaux publiés en plus d'une langue. Les sites Web publiés uniquement en français n'y sont pas soumis, mais il est fortement recommandé de s'inspirer de ces dispositions, puisqu'un site unilingue aujourd'hui pourrait devenir multilingue demain.

S.-s. 2 – Champ d'application

2. Ce standard s'applique aux ministères et aux organismes visés par l'article 64 de la Loi sur l'administration publique (L.R.Q., c. A-6.01).

Loi sur l'administration publique (L.R.Q., c. A-6.01) :

CHAPITRE I

OBJET ET APPLICATION

Composition.

3. Pour l'application de la présente loi, l'Administration gouvernementale est constituée :

- 1° des ministères du gouvernement ;
- 2° des organismes budgétaires, soit les organismes dont tout ou partie des dépenses sont prévues aux crédits qui apparaissent dans le budget de dépenses déposé à l'Assemblée nationale sous un titre autre qu'un crédit de transfert ;
- 3° des organismes dont le personnel est nommé suivant la Loi sur la fonction publique (chapitre F-3.1.1) ;
- 4° des organismes dont le gouvernement ou un ministre nomme la majorité des membres ou des administrateurs et dont au moins la moitié des dépenses sont assumées directement ou indirectement par le fonds consolidé du revenu.

Organisme.

Est considérée comme un organisme, une personne nommée ou désignée par le gouvernement ou par un ministre, avec le personnel qu'elle dirige, dans le cadre des fonctions qui lui sont attribuées par la loi, le gouvernement ou le ministre.

Applicabilité.

4. L'Assemblée nationale, toute personne nommée ou désignée par cette dernière pour exercer une fonction en relevant, avec le personnel qu'elle dirige, ainsi que la Commission de la représentation ne sont assujetties à la présente loi que dans la mesure prévue par une loi.

Il en est de même des tribunaux au sens de la Loi sur les tribunaux judiciaires (chapitre T-16), des organismes dont l'ensemble des membres sont juges de la Cour du Québec, du Conseil de la magistrature et du comité de la rémunération des juges de la Cour du Québec et des cours municipales.

CHAPITRE VI

GESTION DES RESSOURCES INFORMATIONNELLES

Application.

64. Le présent chapitre s'applique à l'Administration gouvernementale.

S.-s. 3 – Définitions

3. Dans le présent standard, on entend par :

- a) **page Web** : un document hypertexte diffusé dans le Web, qui forme un tout couvrant un sujet donné ;
- b) **répertoire** : un ensemble précis de caractères, défini de manière indépendante du codage ;
- c) **site Web** : un lieu où se trouve implanté un hôte Internet, identifié par une adresse Internet et où sont stockées des données accessibles par le Web ;
- d) **site Web multilingue** : un site Web qui comprend des documents dans une langue autre que le français.

SECTION II : SPÉCIFICATIONS

S.-s. 1 – Conditions de conformité au standard

4. Un site Web multilingue est conforme au présent standard si chacune de ses pages Web respecte les exigences des sous-sections qui suivent.

S.-s. 2 – Accès au site

5. Un site Web multilingue doit apparaître par défaut en français à un utilisateur qui y accède.
6. La méthode de sélection automatique de la langue doit être désactivée, sauf pour un service donné en langue anglaise ou en langue autochtone, qui fait l'objet d'un règlement ou d'une entente particulière ou qui est visé par une loi.

Remarques :

Déoulant de la Charte de la langue française, cette exigence a pour but d'exposer implicitement le citoyen à la langue officielle du Québec. Ainsi, tout choix de langue doit être explicite – donc manuel – après que l'on se soit assuré de servir implicitement l'utilisateur dans la langue officielle du Québec. La sélection automatique souffre, en effet, de quelques limitations qui peuvent sérieusement affecter un utilisateur non spécialiste. En voici quelques-unes :

- La présence de l'en-tête `Accept-Language` n'est pas garantie. Certains utilisateurs (ce qui peut-être le cas d'un technicien qui a installé le fureteur dans un endroit public) préfèrent ne pas exprimer leurs préférences automatiquement avec chaque requête et règlent leur fureteur en conséquence ;
- La liste des langues préférées peut être disjointe de la liste des langues disponibles. Dans ce cas, la page demandée devrait être servie en français, mais cela pourrait ne pas être le cas (l'anglais pourrait être choisi par le fureteur, par exemple, si un utilisateur a choisi le chinois et que cette langue n'est pas offerte dans un site Web de l'Administration gouvernementale) ;
- La liste des langues préférées peut ne pas correspondre aux préférences réelles de l'utilisateur, notamment lorsque celui-ci utilise un fureteur dans un endroit public (ex. : café Internet, bibliothèque, aéroport, etc.) et qu'il ne prend pas la peine d'effectuer le réglage de ses préférences linguistiques, ou qu'il ne sait pas que de tels réglages sont possibles.

Pour toutes ces raisons, il est nécessaire d'offrir des moyens de sélection manuelle de la langue dans les cas où il est approprié d'offrir dans un site Web une langue additionnelle en plus de la langue officielle. Il ne faut, en aucun cas, se servir des possibilités techniques permettant la sélection automatique.

S.-s. 3 – Sélection de la langue de consultation

7. Un site Web multilingue doit permettre la sélection explicite de la langue de consultation.
8. Cette sélection doit être offerte en tout temps, que ce soit sur la page Web d'accueil du site ou en descendant dans son arborescence.

Toutefois, un site Web qui est partiellement traduit doit seulement mentionner, à chaque référence à un document en une autre langue que le français, les langues dans lesquelles ce document est disponible.

Considérations techniques :

En principe, cette sélection peut être automatique ou manuelle. Techniquement parlant, lorsqu'elle est appropriée, la sélection automatique est effectuée par le serveur, à partir d'information technique contenue dans la requête http (*Hypertext Transfer Protocol*) de l'utilisateur. La sélection manuelle est toujours le résultat d'un choix délibéré de l'utilisateur, par exemple à la suite d'un clic sur un bouton de choix de langue.

Au plan technique, la sélection automatique est principalement fondée sur l'en-tête `Accept-Language` prévue par le standard de l'*Internet Engineering Task Force* (IETF) désigné sous l'appellation HTTP 1.1 (*RFC 2616 Hypertext Transfer Protocol*). Cet en-tête peut être présent ou non dans une requête ; lorsqu'il est présent, il contient une liste de codes de langue provenant

du standard de l'IETF désigné sous l'appellation RFC 3066 (*Tags for the Identification of Languages*) exprimant les préférences de l'utilisateur. Tous les fureteurs Web modernes permettent à l'utilisateur de régler ces préférences et de les transmettre dans chaque requête HTTP. En théorie, le serveur doit examiner ces préférences et fournir la page demandée dans la langue choisie parmi les langues disponibles pour la page demandée. En l'absence de l'en-tête `Accept-Language`, le serveur peut aussi examiner l'en-tête `User-Agent`, qui contient parfois des indications de langue (du fureteur ou du système d'exploitation).

- La sélection de la langue se fait au moyen d'un élément d'interface textuel qui identifie chaque langue dans cette langue et que l'utilisateur peut activer manuellement pour indiquer son choix.

Remarque :

La langue française sera ainsi désignée par « français », l'anglais par « English », le japonais par « 日本語 » et ainsi de suite. Lorsque des contraintes liées au codage des caractères empêchent d'écrire le nom d'une langue dans cette langue (par exemple « 日本語 » ne peut apparaître dans une page codée selon la norme ISO/CEI 8859-15 : 1999 (*Traitement de l'information – Jeux de caractères graphiques codés sur un seul octet, Partie 15 : Alphabet latin n° 9*), il sera possible de se replier vers une translittération en un alphabet repris par le codage de la page (par exemple « nihongo » pour le japonais en alphabet latin) ou d'utiliser des images montrant correctement le nom de chaque langue dans cette langue, comme dans l'exemple ci-dessous.



Cet élément peut être accompagné d'éléments d'interface visuelle que l'utilisateur peut activer.

Exemples d'éléments d'interface visuelle :

bouton, hyperlien, etc.

Une image de drapeau national ne doit pas être utilisée comme élément d'interface pour la sélection de la langue.

Remarque :

Peu importe la situation, une image de drapeau national ne désigne pas correctement la langue.

C'est notamment le cas lorsqu'une langue internationale comporte des variantes locales dans les pays où cette langue est parlée. Par exemple : l'anglais parlé aux États-Unis diffère de l'anglais parlé au Royaume-Uni.

10. Une image utilisée pour désigner une langue que l'on ne peut décrire à l'aide d'un jeu de caractères conforme au Standard sur les jeux de caractères codés (SGQRI 003), doit porter un attribut `alt` expliquant le contenu de l'image à ceux qui ne peuvent la voir. Dans ce cas, l'attribut `alt` doit contenir le nom de chaque langue dans cette langue, translittéré au besoin.

Remarque :

L'attribut `alt` est parfois utilisé à tort dans certains sites Web pour faire apparaître une infobulle. Pour ce faire, l'utilisation de l'attribut `text` (visible dans tous les navigateurs) est préférable. L'attribut `alt` doit être réservé pour l'accessibilité des pages Web aux personnes handicapées (notamment en braille).

Lorsqu'un site Web soutient la notion de personnalisation (stockage dans le site Web de préférences pour chaque utilisateur) ou la notion de session, la sélection manuelle d'une langue devrait être enregistrée pour la durée de la session, de manière à éviter le besoin de répéter la sélection de langue.

URL génériques et spécifiques :

Bien que cette méthode de sélection automatique de langue soit déconseillée dans les sites Web de l'Administration gouvernementale, elle est tout de même discutée dans cette section. En présence de sélection automatique de la langue¹, la signification de l'adresse d'une page Web (un URL + signification) peut changer. Lorsqu'une page est soumise à la sélection automatique, l'URL ne désigne plus exactement une ressource Web bien précise, mais un ensemble de ressources considérées comme équivalentes en tous points, sauf selon l'axe de sélection automatique (par exemple la langue). Il faut parler alors d'un URL générique. Un exemple d'URL générique est <http://www.exemple.org> désignant la page d'accueil du site Web multilingue <http://www.exemple.org> si ce site Web met en œuvre la sélection automatique de langue.

Un URL spécifique, au contraire, est un URL qui ne désigne qu'une seule ressource bien précise, sans qu'aucune sélection automatique ne soit possible. La page d'accueil en français de <http://www.exemple.org> pourrait par exemple être accessible en <http://www.exemple.org/index.fr.html>. Un des principes architecturaux du Web est que toute ressource importante est adressable par un URL. Par conséquent, les réalisateurs de sites Web mettant en œuvre la sélection automatique doivent s'assurer que toutes les pages où cette sélection est fonctionnelle sont aussi accessibles par des URL spécifiques. Ces URL peuvent d'ailleurs être utiles à la sélection manuelle de langue et sont nécessaires pour pouvoir dire, par exemple, « Il y a une coquille à la troisième ligne du premier paragraphe de telle page », la coquille ne se trouvant fort probablement que dans une seule version de la page.

S.-s. 4 – Exigences techniques

11. L'organisation d'un site Web statique multilingue doit refléter la dimension linguistique selon l'une des méthodes suivantes :

¹ Les notions discutées ici sont aussi valables par rapport aux autres types de sélection automatique que permet le standard de l'IETF HTTP 1.1 (RFC 2616) : sélection de format (HTML [« HyperText Markup Language »], PDF [« Portable Document Format »], XML [« Extensible Markup Language »], etc.), de jeu de caractères et de codage de transport (compression, chiffrement, etc.). En pratique, seule la sélection de langue est couramment utilisée.

Remarque technique :

Un site Web statique est généralement organisé en sections, sous-sections, etc. sur la base des sujets traités dans les différentes pages Web. Cette organisation transparaît dans l'URL permettant d'accéder à une page (par exemple <http://www.exemple.org/section/sous-section/page.html>). Dans nombre de sites Web, les fichiers sont à leur tour rangés dans des répertoires et sous-répertoires qui reflètent cette même structure.

- la structure par sujet est conservée intacte et la langue est indiquée par une extension au nom du document ;

Exemple :

<http://www.exemple.org/section/sous-section/page.fr.html> (notez le codet de langue « fr » avant l'extension finale « .html »). Pour plus d'information sur les codets de langues, consulter le standard sur l'identification des langues (SGQRI 046-04).

Remarque technique :

Le délimiteur TRAIT BAS (« _ ») peut être utilisé à la place du POINT (« . ») avant le codet de langue dans un nom de fichier, notamment si ce dernier caractère pose un problème technique dans un environnement donné.

- la structure est reproduite autant de fois qu'il y a de langues, avec des répertoires de premier niveau indiquant la langue.

Exemple :

L'exemple précédent devient alors <http://www.exemple.org/fr/section/sous-section/page.html>.

Remarque :

La première méthode est préférable lors de la conversion d'un site Web unilingue en un site Web multilingue, puisqu'elle permet de conserver les URL du site Web original, lesquels sont parfois répertoriés par de nombreux moteurs de recherche et conservés dans d'innombrables fichiers de signets.

Une troisième méthode, à éviter, est assez fréquente lors de la conversion d'un site Web unilingue en un site Web multilingue ; il s'agit de reproduire la structure existante en sous-répertoires pour les langues ajoutées tout en conservant la structure intacte pour la langue d'origine (nous obtenons alors, pour un site Web en français traduit en espagnol, <http://www.exemple.org/section/sous-section/page.html> pour une page en français et <http://www.exemple.org/es/section/sous-section/page.html> pour sa version espagnole). Cette méthode mène à une structure bancale et nuisible à la mise en œuvre de la réutilisation des actifs comme les images et à l'entretien du site Web.

Quelle que soit la méthode choisie, il est conseillé d'utiliser des codets alphabétiques à deux caractères tels que précisés dans le standard sur l'identification des langues (SGQRI 046-04)² qui est conforme à la norme de l'ISO désignée sous l'appellation ISO 639-1 (*Codes pour la représentation des noms de langue : Code alpha-2*). Les dispositions du présent standard sont aussi conformes au standard de l'IETF désigné sous l'appellation RFC 3066 (qui cite aussi la norme internationale ISO 639-1). Cela a pour but de garantir la cohérence avec les codets de

² Au besoin comme marqueur de langue dans les noms de fichiers ou de répertoires, le standard SGQRI 046-04 suggère aussi une méthode pour identifier le pays où la langue est parlée.

langue inscrits dans les fichiers et avec ceux utilisés dans les protocoles Internet, par exemple pour la sélection automatique de langue en HTTP.

Pour une langue donnée, le standard RFC 3066 préfère le codet à deux caractères de la norme internationale ISO 639-1 sauf lorsqu'il est inexistant et qu'il y en a un à trois caractères disponible dans la norme de l'ISO désignée sous l'appellation ISO 639-2 (*Codes pour la représentation des noms de langue : Code alpha-3*).

Il y a une différence normative entre, d'une part, l'usage des codets des normes de l'ISO désignées sous les appellations ISO 3166-1, ISO 639-1 et ISO 639-2 (citées dans le standard SGQRI 046-4) et, d'autre part, celui du standard RFC 3066. Alors que les normes internationales de l'ISO donnent une signification à la casse, le standard RFC 3066 n'en donne aucune, et cela peut être source de confusion. Par exemple, dans la norme internationale ISO 639-1, le codet « fr » signifie « langue française », alors que dans la norme internationale ISO 3166-1, le codet « FR » signifie « France ».

12. La confection d'un site Web dynamique multilingue doit produire des résultats fonctionnellement équivalents à ceux d'un site Web statique.

Remarque :

Il existe toutefois une tendance qui empêche dans ces cas un utilisateur de pouvoir conserver des signets à une page dynamique dans sa langue si la programmation n'a pas prévu les paramètres appropriés dans les URL.

Dans la mesure du possible, il faut permettre à un utilisateur de pouvoir revenir rapidement sur une page dynamique, et ce, dans la langue choisie.

Note importante concernant l'interopérabilité du codage des caractères :

Un jeu de caractères codés utilisé pour les sites Web multilingues, et minimalement conforme au standard SGQRI 003, soutient d'emblée les vingt-six langues suivantes : albanais, allemand, anglais, basque, breton, catalan, danois, espagnol, estonien, féroïen, finnois, français, frison, gaélique écossais, gaélique irlandais (nouvelle orthographe), galicien, groenlandais, islandais, italien, latin, luxembourgeois, néerlandais, norvégien, portugais, romanche, suédois. L'intégrité des caractères de ces langues est assurée dans les échanges entre les sites gouvernementaux. Au-delà, cette interopérabilité n'est pas garantie.

Recommandations pour les autres langues :

Le codage des caractères des pages d'un site Web multilingue dépend de la langue de chaque page et des exigences imposées au site Web. Pour tous les sites Web, la principale exigence est généralement qu'ils soient consultables par le plus grand nombre d'utilisateurs potentiels. Puisque les utilisateurs emploient divers fureteurs (plus ou moins récents, de différentes marques), cette contrainte revient à exiger la compatibilité avec le plus grand nombre de fureteurs possible pour chaque langue. Une compatibilité raisonnable est obtenue en choisissant le codage de caractères défini au tableau suivant pour différentes langues.

Langues d'Europe occidentale (allemand, anglais, espagnol, français, italien, néerlandais, suédois, etc.) [voir l'article précédent pour plus de précision]	ISO/CEI 8859-15
Langues d'Europe centrale (croate, hongrois, polonais, roumain, slovaque, tchèque, etc.)	ISO/CEI 8859-2

Russe	KOI8-R
Arabe	ISO/CEI 8859-6
Japonais	Shift-JIS
Chinois, orthographe traditionnelle (Hong Kong, Taïwan)	Big5
Chinois, orthographe simplifiée (RPC)	GB2312
Coréen	EUC-KR (aussi appelé KS C 5601)
Grec	ISO/CEI 8859-7
Turc	ISO/CEI 8859-9

Il peut arriver que des exigences particulières d'un site Web rendent nécessaire le choix d'un autre codage. Ce sera le cas si, par exemple, la technologie du serveur exige un seul codage pour tout le site Web ou si plus d'une langue doit être présente dans certaines pages, de telle façon qu'aucun des codages traditionnels identifié ci-dessus ne soit satisfaisant. En de telles circonstances, le choix d'un codage de la norme internationale ISO/CEI 10646 (*Jeu universel de caractères codés sur plusieurs octets [JUC]*) est le plus souvent approprié. En vertu du standard de l'IETF désigné sous l'appellation RFC 2277 (*SMTP Service Extension for 8bit-MIMEtransport*), ce codage doit alors être UTF-8 (UTF : *UCS Transformation Format*, format transformé du jeu universel de caractères).

Le codage des caractères doit se faire normalement au moyen d'un des jeux de caractères mentionnés ci-dessus. Il faut éviter autant que possible l'utilisation des références de caractères (par exemple é pour « é ») et des appels d'entités (par exemple ´ pour ce même « é »). Ces mécanismes ajoutent lourdeur et complexité au codage des pages, nuisent au fonctionnement de certains logiciels et constituent *stricto sensu* une violation de la séparation en couches des fonctionnalités du Web. Il convient donc de les utiliser que de façon exceptionnelle, par exemple pour coder quelques caractères japonais dans une page rédigée en ISO/CEI 8859-15.

Il est impératif que l'identité du codage des caractères soit toujours correctement transmise avec les pages Web. Cette transmission peut se faire de deux manières non exclusives :

- 1) le codage peut être indiqué dans les pages elles-mêmes ;
- 2) l'identité du codage peut être transmise dans les en-têtes HTTP.

Les langages Web comme HTML et XML offrent des mécanismes qui permettent de désigner le codage des caractères à l'intérieur des documents. En HTML selon le standard de l'IETF désigné sous l'appellation RFC 2070 (*Internationalization of the Hypertext Markup Language*) et selon le standard du World Wide Web Consortium (W3C) désigné sous l'appellation HTML 4.01 (*HTML 4.01 Specification*), un élément `meta` est utilisé à cette fin :

```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-15">
```

Cet élément doit être placé dans l'en-tête du document, idéalement immédiatement après la balise `head`. La plupart des éditeurs HTML modernes prennent en charge ce mécanisme. En XML, la désignation du codage se fait au moyen d'une *déclaration de codage* :

```
<?xml version="1.0" encoding="ISO-8859-15"?>
```

Il n'existe aucun standard pour programmer les serveurs Web de manière à transmettre l'identification de codage de caractères dans les en-têtes HTTP ; tout dépend du serveur utilisé. Seul l'emplacement final de cet identificateur est standardisé (cf. standard RFC 2616) : il s'agit

d'ajouter à l'en-tête `Content-Type`, un paramètre appelé `charset`, dont la valeur est l'identificateur de codage. Par exemple :

```
Content-Type: text/html; charset=ISO-8859-15
```

Lorsque l'identification de codage est transmise à la fois dans l'en-tête HTTP et dans le document lui-même, il est impératif que ces deux indications soient identiques.

Dans les exemples ci-dessus, l'identifiant de codage de caractères est celui de la norme internationale ISO/CEI 8859-15. Dans la pratique, cet identifiant doit être l'identifiant préféré (paramètre `charset`) du registre des jeux de caractères de l'IANA (*Internet Assigned Numbers Authority*) correspondant au codage utilisé pour la page. Le registre de l'IANA est accessible depuis <http://www.iana.org>.

Remarque :

Dans la mesure où l'information dynamique recueillie lors d'une session Web doit être compatible (intégrité de l'information et interopérabilité garanties) avec toutes les plates-formes fonctionnelles du gouvernement du Québec, il faut s'assurer de se conformer au standard sur les jeux de caractères codés (SGQRI 003).

Pratiques recommandées accompagnant le présent standard :

La réalisation d'un site Web multilingue augmente la nécessité d'adopter certaines pratiques recommandées. Ces pratiques sont toujours de bon conseil, mais le bénéfice de leur utilisation est accru dans un cadre multilingue. Un aperçu des pratiques recommandées est décrit dans les sections qui suivent. Le lecteur peut aussi se référer au standard sur le format des documents diffusés dans le Web (SGQRI 013).

Pratique recommandée sur l'indication de langue :

Il est conseillé de toujours inclure dans chaque page Web une indication de la langue dans laquelle la page est rédigée. En HTML, cette indication se fait au moyen d'un attribut `lang` porté par l'élément `html` :

```
<html lang="fr">
  <meta http-equiv=...
  ...
```

La valeur de l'attribut `lang` doit être un code de langue du standard RFC 3066. Ce standard permet de préciser un dialecte ou une variante de langue, par exemple `fr-CA` pour le français canadien. Il faut éviter ce surcroît de précision (c'est-à-dire l'indication de pays, `CA` dans l'exemple), nuisible à certains processus, sauf lorsque pleinement justifié par le contenu linguistique de la page.

Lorsqu'une partie importante du texte d'une page est dans une autre langue, il peut être utile d'indiquer cette langue au moyen d'un attribut `lang` porté par un élément englobant le texte en langue étrangère (un paragraphe par exemple). L'élément `span` peut être utilisé lorsque aucun autre élément HTML n'est approprié.

Note. Cet attribut est notamment utile aux moteurs de recherche multilingues, qui peuvent y puiser la signification de certains mots clefs ; la phrase « Jean put dire comment on tape » est un excellent exemple de phrase bien formée en deux langues distinctes (anglais et français), qui a deux sens différents dans ces deux

langues (si la phrase est en anglais, elle signifie « Jeanne a placé un commentaire affreux sur ruban », un sens très différent du sens français). L'identification de la langue permet d'attacher un sens aux mots et de retrouver certains sujets en différentes langues à l'aide d'un mot clef précisé en une langue autre.

Pratique recommandée sur les images :

Il est conseillé de toujours réduire au minimum la quantité et la taille des images dans les pages Web, pour des raisons de bande passante. L'économie de bande passante devient plus importante pour une version linguistique destinée à un public lointain, et encore plus si ce public réside dans un pays où l'infrastructure n'a pas un niveau de développement équivalent à celui du Québec.

Note. Il est difficile de donner des recommandations absolues, mais citons quelques exemples concrets : certains sites Web recommandent (exemple : annonces classées) de ne pas déposer d'images de poids supérieur à 21 K octets pour des images simples. À l'autre bout du spectre, des sites Web spécialisés en photographie recommandent de ne pas déposer de photos de format supérieur à 800 x 600 pixels et d'utiliser une compression raisonnable d'image. Le format BMP (*Bitmap Graphics*), qui n'offre aucune compression, devrait être évité. Voir aussi <http://www.piv.gouv.qc.ca> pour certaines indications ponctuelles applicables aux sites Web généraux de l'Administration gouvernementale.

Les images contenant du texte sont davantage déconseillées. Dans un contexte unilingue, elles nuisent à l'accessibilité pour les personnes handicapées et à l'indexation par les moteurs de recherche. En contexte multilingue, elles augmentent considérablement les délais et les coûts de traduction d'un site Web. Il est possible de pallier quelque peu ce désavantage en créant les images à l'aide d'un logiciel qui gère les images en couches, en isolant le texte dans une couche ne contenant rien d'autre et en fournissant aux traducteurs ces images en couche ainsi que la désignation des fontes utilisées de même que tout autre détail utile.

Note. Un bouton qui ne contient que des éléments textuels modifiables n'est pas considéré ici comme une image, mais comme un élément de formatage textuel.

Il faut aussi se méfier des images contenant des références culturelles ou basées sur des jeux de mots. Très peu d'éléments culturels sont universels ; il est donc probable que de telles images doivent être refaites dans une traduction du site Web.

Exemple : L'image d'une boîte aux lettres rurale nord-américaine comme icône représentant le courriel est totalement incomprise dans des pays autres que le Canada ou les États-Unis.

Note. Il faudra aussi porter une attention particulière à l'usage des couleurs. Il existe plusieurs formes de daltonisme et il est difficile de donner des indications absolues sur l'usage ou le non-usage de couleurs particulières (certaines personnes sont totalement insensibles aux couleurs) ; toutefois il faudrait prévoir un mécanisme pour que les daltoniens puissent se retrouver, tout comme il existe une convention pour les feux de circulation, qui respectent toujours un ordre dans la présentation des couleurs.

Pratique recommandée sur les feuilles de style :

Les feuilles de style décrites dans le standard du W3C désigné sous l'appellation CSS (*Cascading Style Sheets*) présentent de nombreux avantages, que ce soit en contexte unilingue ou multilingue. Malheureusement, leur prise en charge par les fureteurs anciens comme Netscape 4.x est imparfaite et incomplète. Il faut donc en user avec parcimonie, bien que cette prise en charge incomplète permet de profiter de certaines possibilités sans remettre en cause trop sérieusement la compatibilité.

La désignation des fontes (polices de caractères) est un aspect des feuilles de style décrites

dans le standard CSS qui est, d'une part, parmi les mieux soutenus, même par les fureteurs anciens et qui est, d'autre part, d'un intérêt certain en contexte multilingue.

Le standard CSS permet de désigner une famille de fontes sous un nom générique, sans faire référence au nom d'une fonte particulière qui ne saurait ni être présente sur tous les systèmes des utilisateurs, ni être appropriée pour toutes les langues. Pour choisir une fonte sans empattements (du style Arial ou Helvetica) comme fonte de base pour une page, il faut ajouter la mention suivante à la feuille de style :

```
body { font-family: sansserif; }
```

S'il existe des raisons de désigner une fonte particulière, l'utilisation de feuilles de style est tout de même avantageuse. D'une part, la désignation d'une fonte spécifique peut être accompagnée d'un nom générique, permettant une position de repli aux fureteurs ne possédant pas la fonte spécifique. D'autre part, si la fonte spécifique doit être changée à la traduction, il est beaucoup plus facile de le faire en un seul endroit (la feuille de style) plutôt qu'en de multiples endroits dans toutes les pages. Les frais et délais de traduction s'en trouveront donc réduits, et la maintenance facilitée.

Les feuilles de style permettent aussi de préciser une couleur de fond ou une image d'arrière-plan pour un élément quelconque³. Cette fonctionnalité permet souvent d'éviter l'utilisation d'images contenant du texte : l'image est posée à l'arrière-plan par la feuille de style et le texte saisi comme contenu de l'élément, qui peut alors être traduit efficacement sans toucher à l'image.

En résumé, la balise HTML `font` est à éviter, et à être remplacée par les feuilles de style décrites dans le standard CSS dans toutes les pages HTML nouvelles ou révisées. Il faut préférer l'utilisation de feuilles de style externes, communes à tout un site Web ou à de grandes sections de site Web, pour maximiser les avantages à retirer du standard CSS.

Pratique recommandée sur la validation :

Il est important de ne publier dans le Web que des documents valides, qui satisfont les standards applicables, comme HTML et CSS. La validité favorise la compatibilité avec le plus grand nombre de fureteurs et la cohérence de l'affichage, mais aussi la compatibilité avec les outils de maintenance actuels ou futurs du site Web. Dans le cas d'un site Web multilingue, la validité favorise aussi la compatibilité avec les outils utilisés pour la traduction, que ce soit de simples éditeurs HTML capables de traiter les langues cibles ou des outils d'aide à la traduction (mémoires de traduction, gestionnaires de terminologie) servant à améliorer l'efficacité du processus. La validité est donc un facteur de réduction de coûts dans le cas des sites Web multilingues de même que de compatibilité et de pérennité dans tous les cas.

Il est préférable d'assurer la validité en n'utilisant que des outils d'édition qui la garantissent. Il est toutefois possible de la vérifier a posteriori au moyen d'outils séparés, tels que les validateurs du W3C pour les standards HTML (<http://validator.w3.org>) et CSS (<http://jigsaw.w3.org/css-validator>). Il est aussi souvent possible de réparer automatiquement des documents invalides, par exemple en utilisant Tidy du W3C (<http://www.w3.org/People/Raggett/tidy>).

Pratique recommandée sur les changements de dimension :

³ En pratique, la compatibilité avec les fureteurs plus vieux impose de limiter cet usage aux éléments de type *bloc* (comme les paragraphes) et aux cellules de tableaux.

Une des constantes en traduction est que le texte traduit change de dimension. Selon la langue source et la langue cible, le texte pourra être plus long ou plus court. Même dans un cas où le texte raccourcit, certains segments peuvent augmenter de taille, par exemple lorsqu'un mot ou une brève expression dans la langue source doit être traduit par une périphrase dans la langue cible. Ceci vaut aussi bien pour le corps de texte que les titres, les menus et le texte inscrit dans des images.

Il convient donc de rédiger les pages de façon à éviter ces changements de dimension. Les dimensions absolues (en pixels ou en nombres de caractères), les dispositions dépendant de la taille du texte, les assemblages d'images à refaire lors d'un changement de taille, etc. sont du nombre. Jusqu'à un certain point, ces lignes directrices favorisent aussi la compatibilité en ce que les pages ainsi composées pourront être affichées de façon cohérente sans égard à la taille de la fenêtre de rendu (le lecteur trouvera à ce sujet des renseignements pratiques, notamment sur <http://www.piv.gouv.qc.ca>).

Pratique recommandée sur les scripts, les applets et les servlets :

Une des règles d'or de l'internationalisation de logiciels est de séparer le plus complètement possible le programme objet des ressources. Les ressources comprennent toutes les chaînes de caractères de l'interface utilisateur ainsi que les images, icônes, dimensionnements et autres données statiques.

Cette règle s'applique aussi aux sites Web. Il faut en tenir compte lorsque les pages contiennent ou font appel à des codes (les pages purement en HTML/CSS n'étant constituées que de ressources, il n'y a alors rien à isoler).

La programmation de ressources peut se trouver soit dans des scripts dans les pages Web (scripts exécutés par le serveur ou transmis au fureteur pour exécution côté client), soit dans des applets exécutés par le fureteur, soit dans des servlets exécutés par le serveur. Dans tous ces cas, il convient d'isoler au maximum le programme objet lui-même de toute chaîne de texte, des images et des données associées, comme les dimensions. Il est préférable de composer les scripts dans des fichiers séparés qui seront inclus par référence dans les pages Web de sorte à éviter que ces scripts ne génèrent du texte (un script étant un programme, coder du texte dans un script revient à coder *en dur* un message dans une langue donnée, ce qui est à proscrire) ; tout texte doit provenir soit de la page qui inclut le script, soit de fichiers séparés ne contenant que des ressources. Les applets et les servlets devraient faire appel à des fichiers de ressources pour tout contenu linguistique.

Malgré que l'internationalisation d'un programme objet soit possible, il arrivera que ce programme objet doive être adapté pour tenir compte de différences culturelles (par exemple format d'affichage de la date) ou techniques (par exemple changement de jeu de caractères). Cette adaptation étant inévitable, il convient de minimiser les coûts et délais qui en découlent. D'une part, il s'agit de réduire au strict minimum l'utilisation des scripts, applets et servlets, et d'autre part, tout programme objet devra être correctement internationalisé dès sa conception.

SECTION III : DISPOSITIONS TRANSITOIRES ET FINALES

S.-s. 1 – Mesures transitoires

SGQRI 003 – Conception d'un site Web multilingue

Version 1.0 du 11 décembre 2006

Nom du fichier : S0110010 VE Sites Web multiling 20061211.2.doc

13. Un site Web multilingue en ligne avant la date d'entrée en vigueur du présent standard doit être conforme aux exigences de la section II au plus tard trois ans après cette date.

S.-s. 2 – Révision

14. Au plus tard cinq ans après l'entrée en vigueur de ce standard, le ministère des Services gouvernementaux doit, de concert avec les ministères et les organismes, en évaluer la mise en œuvre et conseiller le ministre des Services gouvernementaux quant à l'opportunité d'y apporter des modifications en vue d'une proposition au Conseil du trésor.

S.-s. 3 – Date d'entrée en vigueur

15. Ce standard entre en vigueur le 11 décembre 2006.

RENSEIGNEMENTS COMPLÉMENTAIRES

R.C. 1 – Autres sigles et définitions

R.C. 1.1 – Sigles

API

Application Program Interface

BMP

Bitmap Graphics

CEI

Commission électrotechnique internationale

CSS

Cascading Style Sheets. Standard de feuilles de style du W3C.

HTML

HyperText Markup Language. Langage de balisage de documents.

HTTP

Hypertext Transfer Protocol. Protocole de transport d'hypertexte.

IETF

Internet Engineering Task Force. Communauté volontaire de personnes intéressées au développement de l'architecture Internet et de ses standards (<http://www.ietf.org/overview.html>).

ISO

Organisation internationale de normalisation (ISO est un symbole provenant de la racine grecque *isos* – qui signifie *même*) ; en anglais : *International Organization for Standardization*.

NFC

Normalization Form C. Forme normalisée C sous Unicode choisissant comme équivalence canonique un caractère précomposé plutôt qu'une séquence combinatoire, quand il existe de telles équivalences.

PDF

Portable Document Format. Format de document portable d'Adobe Systems inc.

RFC

Request for Comments. Chaque RFC fait partie d'une série de notes techniques et organisationnelles dans Internet, série qui a débuté en 1969. Un RFC peut devenir un standard Internet ; entre-temps, plusieurs RFC sont appliqués en fonction des besoins des utilisateurs de l'Internet.

URL

Universal Resource Locator. Localisateur de ressources universel. Convention de dénomination universelle qui sert à localiser et à accéder des ressources dans Internet.

XML

Extensible Markup Language. Langage de balisage de documents défini par le W3C, et plus évolué que le langage HTML.

R.C. 1.2 – Définitions

Script

Série d'instructions servant à accomplir une tâche particulière.

Applet

Petite application écrite en langage Java et qui, insérée dans un document Web, exécute ses objets multimédias en présence d'un navigateur compatible, et ce, directement sur l'ordinateur de l'internaute, peu importe le système d'exploitation utilisé.

Servlet

Programme Java qui utilise des modules supplémentaires figurant dans l'API Java, qui s'exécute dynamiquement sur le serveur Web, et permet l'extension des fonctions du serveur.

R.C. 2 – Références bibliographiques

R.C. 2.1 – Références normatives

GOVERNEMENT DU QUÉBEC. *Cadre de diffusion de l'information gouvernementale dans Internet, Langue de diffusion*.

<http://www.webmaestro.gouv.qc.ca/ress/cadre/Cadre/cadre.htm#langue> .

GOVERNEMENT DU QUÉBEC. *Charte de la langue française*. 2004,

<http://www.canlii.org/qc/legis/loi/c-11/20041104/tout.html> .

GOVERNEMENT DU QUÉBEC. *Guide sur l'emploi et la qualité du français dans les sites Web de l'Administration*,

http://www.oqlf.gouv.qc.ca/ressources/bibliotheque/documentation/guide_web.html .

GOVERNEMENT DU QUÉBEC. *Politique gouvernementale relative à l'emploi et à la qualité de la langue française dans l'Administration*, <http://www.spl.gouv.qc.ca/langue/politique1.html> .

GOVERNEMENT DU QUÉBEC. *Programme d'identification visuelle (PIV) du gouvernement du Québec*, <http://www.piv.gouv.qc.ca> .

GOVERNEMENT DU QUÉBEC. *SGQRI 003, Jeux de caractères codés*. 2006.

GOVERNEMENT DU QUÉBEC. *SGQRI 013, Format des documents diffusés dans le Web*. 2006

GOVERNEMENT DU QUÉBEC. *SGQRI 046-4, Codification, Partie 4 : Langues*. 2006.

INTERNET ENGINEERING TASK FORCE. *RFC 2070, RFC 2070, Internationalization of the Hypertext Markup Language*, <http://www.ietf.org/rfc/rfc2070.txt>. (Traduction française disponible à l'adresse http://alis.isoc.org/web_ml/html/rfc-i18n/rfc-i18n-1.fr.html).

INTERNET ENGINEERING TASK FORCE. *RFC 2277, IETF Policy on Character Sets and Languages, IETF, janvier 1998*, <http://www.ietf.org/rfc/rfc2277.txt>.

INTERNET ENGINEERING TASK FORCE. *RFC 2616, Hypertext Transfer Protocol, HTTP/1.1*, <http://www.ietf.org/rfc/rfc2616.txt>.

INTERNET ENGINEERING TASK FORCE. *RFC 3066, Tags for the Identification of Languages*, <http://www.ietf.org/rfc/rfc3066.txt>.

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO 639-1 Codes pour la représentation des noms de langue : Code alpha-2*. 2002.

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO 639-2 Codes pour la représentation des noms de langue : Code alpha-3*. 1998.

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO 3166-1, Codes pour la représentation des noms de pays et de leurs subdivisions – Partie 1: Codes de pays*. 1997.

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO/CEI 8859 : 1999, Traitement de l'information – Jeux de caractères graphiques codés sur un seul octet, Partie 15 : Alphabet latin n° 9*. 1999.

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO/CEI 10646-1 : 2003, Technologies de l'information, Jeu universel de caractères codés sur plusieurs octets (JUC)*. 2003.

WORLD WIDE WEB CONSORTIUM. *Cascading Style Sheets, level 2, CSS2 Specification, W3C, mai 1998*, <http://www.w3.org/TR/REC-CSS2>. (Traduction française disponible à l'adresse <http://www.yoyodesign.org/doc/w3c/css2/cover.html>).

WORLD WIDE WEB CONSORTIUM. *HTML 4.01 Spécification*, décembre 1999, <http://www.w3.org/TR/html401/> (Traduction française disponible à l'adresse <http://www.la-grange.net/w3c/html4.01/cover.html>).

WORLD WIDE WEB CONSORTIUM. *Service de validation CSS*, <http://jigsaw.w3.org/css-validator>.

WORLD WIDE WEB CONSORTIUM. *Service de validation HTML*, <http://validator.w3.org>.

WORLD WIDE WEB CONSORTIUM. *Programme utilitaire de validation Tidy*, <http://www.w3.org/People/Raggett/tidy/>.

R.C. 2.2 – Autres références

IANA. Site Web de l'*Internet Assigned Numbers Authority*, <http://www.iana.org>.

GOVERNEMENT DU QUÉBEC. *Programme d'identification visuelle*, <http://www.piv.gouv.qc.ca>

R.C. 3 – Dérogation aux autres standards du gouvernement du Québec

Le présent standard déroge partiellement au standard sur l'identification des langues (SGQRI 046-04) pour l'usage des codets de langue, en raison de la pratique documentée dans le standard Internet RFC 3066.

R.C. 4 – Conformité au concept d'adaptabilité culturelle et linguistique

Le présent standard gère les conditions d'application des principes d'adaptabilité culturelle et linguistique pour les sites Web multilingues.

R.C. 5 – Composition du groupe de travail responsable de l'élaboration du standard

Depuis le 18 février 2005, le ministre des Services gouvernementaux assume, en matière de gestion des ressources informationnelles, la responsabilité d'élaborer et de proposer notamment des standards au Conseil du trésor. Au moment des travaux du groupe de travail interministériel, de 2002 à 2004, les personnes suivantes représentaient les ministères et les organismes suivants :

Rédacteur et chargé de projet :

LA BONTÉ, Alain Secrétariat du Conseil du trésor

Membres du groupe :

ASSAFIRI, Abdallah	Secrétariat du Conseil du trésor
AUDET, Hélène	Société de l'assurance automobile du Québec
BOUCHER, Rose-Ange	Ministère des Transports
BOULANGER, Louise	Commission des normes du travail
BOURGET, Renaud	Régie des rentes du Québec
BRISSETTE, Normand	Ministère de l'Industrie et du Commerce
CAOINETTE, Jacques	Secrétariat du Conseil du trésor
CORRIVEAU, Marc	Régie de l'assurance maladie
CÔTÉ, Madeleine	Ministère de la Culture et des Communications
DOLBEC, Pierre	Régie de l'assurance maladie
FORTIN, Steven	Ministère de l'Emploi, de la Solidarité sociale et de la Famille
FRENETTE, François	Secrétariat du Conseil du trésor
GIRARD, Benoît	Ministère des Relations avec les citoyens et de l'Immigration
GUÉRIN, Réjad	Financière agricole du Québec
HUDON, Yves	Secrétariat du Conseil du trésor
JOHNSTON, Daniel	Ministère des Relations internationales
MANDJEE, Azim	Office québécois de la langue française
MATTE, Pierre	Ministère de l'Éducation
MATTEAU, Julie	Ministère des Transports
MICHAUD, Florent	Société immobilière du Québec
MONTMINY, Jacques	Ministère de la Culture et des Communications
OUELLETTE, Raymond	Ministère de l'Éducation
PERREAULT, Dany	Ministère de l'Environnement
POTVIN, Ginet	Secrétariat du Conseil du trésor

ROY, Jean-Jacques	Secrétariat du Conseil du trésor
SÉGUIN, Jean-François	Ministère de l'Éducation