# Annexe A

# SOLUTION POUR MASQUER LES DONNÉES (MD)

## Énoncé des besoins

### 1. Besoins

Le Bureau du directeur général des élections du Canada (BDGE) a besoin de « masquer » des données sensibles qui serviront au développement, à la mise à l'essai, aux essais d'acceptation par l'utilisateur et à la pré-production. Le Bureau du directeur général des élections du Canada (BDGE) est aussi désigné par Élections Canada (EC), nom sous lequel il est plus connu.
Le masquage des données est défini comme étant le remplacement de données sensibles par des données réalistes, mais non réelles.

Ces besoins portent sur les éléments suivants :

a) Logiciel de masquage de données sous licence
b) Maintenance logicielle sous licence
c) Services de soutien logiciel sous licence
d) Supports et documentation
e) Maintenance logicielle optionnelle
f) Services de soutien logiciel optionnel
g) Formation optionnelle
h) Services professionnels optionnels

### 2. Buts

La solution de masquage de données vise à permettre à EC de développer la capacité de « masquer » les données sensibles qui serviront au développement, à la mise à l'essai, aux essais d'acceptation par l'utilisateur et à la pré-production. La solution doit fonctionner suivant les paramètres et dans les conditions qui sont énoncés dans le présent document. Il s'agit de paramètres de travail et de paramètres techniques.

### 3. Objectifs spécifiques

La solution de masquage de données doit permettre d'atteindre les objectifs suivants :

1. « Masquer » les données sensibles qui servent au développement, à la mise à l'essai, aux essais d'acceptation par l'utilisateur et à la pré-production.
2. La solution de masquage de données doit être reproductible.
3. L'intégrité référentielle des données doit être préservée dans les données masquées.
4. Aucune correspondance bijective ne doit être établie lors du masquage des données.
5. Les données masquées produites doivent être lisibles humainement.
6. La solution doit comporter à la fois des règles de masquage des données intégrées pour les procédures de masquage communes et des règles de masquage définies par l'utilisateur.
7. La solution doit permettre de masquer les données en place.

8. La solution doit permettre de générer des données masquées qui ne soient pas réversibles.
9. La solution doit permettre d'utiliser des règles multiples en même temps.
10. La solution doit permettre d'emmagasiner des règles de masquage des données en vue d'une réutilisation ultérieure.
11. La solution doit permettre à EC de mieux respecter les politiques en matière de collecte de données et d'utilisation des données.
12. La solution doit permettre de mieux faire connaître les bonnes pratiques de gestion de l'information.

## *4. Portée*

Comme on l'a mentionné ci-haut, il faut une solution de masquage des données. Initialement, cette solution sera déployée dans l'environnement de R-D (recherche et développement) d'EC pour qu'on puisse s'assurer qu'elle répond aux besoins énoncés par EC.

Lorsqu'EC sera satisfait de la solution de masquage des données, EC la déploiera au niveau de ses diverses bases de données de développement et de mise à l'essai.

EC a besoin d'un logiciel sous licence, ainsi que de services de maintenance et de soutien logiciels connexes pour la solution de masquage des données.

EC pourrait avoir besoin d'obtenir des services optionnels professionnels auprès du fournisseur, sur une base ponctuelle.

EC pourrait avoir besoin d'obtenir des services de formation optionnels auprès du fournisseur, à la demande de l'autorité chargée du projet de solution MD.

### Utilisateurs

Les utilisateurs de la solution de masquage des données sont les suivants :

1. Administrateur du masquage des données
2. Administrateur de base de données
3. Coordonnateur de la protection des renseignements personnels

## *5. Environnement informatique d'EC*

EC se sert de plus de 200 applications spécialement développées et de plus de 20 applications commerciales (COTS) pour soutenir et simplifier le travail, accéder aux données et traiter des millions de transactions concernant des électeurs. Ces applications sont critiques pour répondre continuellement aux besoins d'EC et à ceux des citoyens canadiens.

Les chiffres qui suivent donnent un aperçu du niveau actuel de soutien fourni par la technologie de l'information (TI) à EC :

- Environ 400 postes de travail
- Environ 300 appareils Blackberry ou téléphones cellulaires
- Environ 1600 serveurs (physiques ou virtuels)
- Site Web d'hébergement externe
- Réseaux internes reliant plus de 630 points de service
- Liens externes avec de nombreux partenaires (provinces, territoires, bureaux techniques mobiles, autres ministères comme Vital Stats, Agence du revenu du Canada et certains organismes électoraux provinciaux)
- Plus de 200 applications développées sur mesure

- Environ 500 produits logiciels (y compris des produits destinés au développement logiciel).
- Téléphonie : 1 million d'appels (dont certains du système de réponse vocale automatisée)
- Courriel : 100 000 (plus d'un million de pourriels sont bloqués chaque semaine)
- Plus de 80 téraoctets de données.

**a) Plateforme des serveurs**

| Système d'exploitation | Technologie utilisée | Configuration de l'UC | Nombre de serveurs physiques | Nombre de MV actuellement sur tous les serveurs | Nombre moyen de bases de données |
|---|---|---|---|---|---|
| OEL | HP - DL 580 | 4 cœurs | 9 | 65 | 3 (par MV) |
| Itanium | HP – Itanium | 2 cœurs | 2 | s.o. | 3 |
| Unix | HP – Unix | 2 cœurs | 2 | s.o. | 5 |
| | | | | | |

**b) Plateforme des postes de travail**

EC utilise actuellement Microsoft Windows XP SP3 avec Microsoft Office 2010, mais pourrait passer à une version plus récente.

**c) Authentification et courriel**

EC utilise actuellement Microsoft Active Directory 2000 pour authentifier les utilisateurs accédant au réseau, mais passera à Microsoft Active Directory 2008R2. EC se sert actuellement de Microsoft Exchange 2007 pour le courriel.

**d) Plateforme d'informatique décisionnelle**

EC se sert actuellement de la série d'outils d'informatique décisionnelle (ID) IBM Cognos.
EC détient actuellement des licences d'utiliseurs désignés pour la solution IBM Cognos.

**e) Navigateur Internet**

EC se sert actuellement de Windows Internet Explorer version 8 (IE 8) comme navigateur Web pour récupérer, obtenir ou parcourir de l'information sur le Web (Internet et intranet).

## *6. Politiques d'EC, directives du Secrétariat du Conseil du Trésor du Canada (SCT), directives et normes informatiques relatives à la gestion de l'information (GI)*

**a) Exigence relative au code d'éthique**

L'entrepreneur doit se comporter qu'il dirigera et ses employés, contractuels et agents, l'instruction de se conduire, en tout temps, pendant l'exécution des services fournis dans le cadre d'un contrat éventuel découlant de la présente demande de propositions (DP), de la manière conforme aux valeurs et à l'éthique prescrite pour les fonctionnaires dans le Code de valeurs et d'éthique du secteur public du Secrétariat du Conseil du Trésor, en vigueur à compter du 2 avril 2012, et de respecter de manière générale la lettre et l'esprit du Code dans toutes ses relations avec le Canada ou en son nom. La version électronique du Code se trouve à l'adresse suivante : http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=25049

**b) Politique du Conseil du Trésor sur l'utilisation des réseaux électroniques**

L'entrepreneur doit se comporter qu'il dirigera et ses employés, contractuels et agents, l'instruction de se conduire, en tout temps, pendant l'exécution des services fournis dans le cadre d'un contrat éventuel découlant de la présente demande de propositions (DP), de manière conforme à la Politique du Conseil du Trésor sur l'utilisation des réseaux électroniques. La version électronique de la Politique se trouve à l'adresse suivante : http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=12419&section=text

**c) Politique du Conseil du Trésor sur l'obligation de prendre des mesures d'adaptation pour les personnes handicapées dans la fonction publique fédérale**

La version électronique de la Politique se trouve à l'adresse suivante : http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=12541

**d) Norme du Conseil du Trésor sur l'accessibilité des sites Web**

La version électronique de la Politique se trouve à l'adresse suivante : http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?evttoo=X&section=text&id=23601

**e) Normalisation des sites Internet (NSI)**

La version électronique de la Politique se trouve à l'adresse suivante : http://www.tbs-sct.gc.ca/clf2-nsi2/index-fra.asp

**f) Lignes directrices sur l'accessibilité du contenu Web (Web Content Accessibility Guidelines ou WCAG)**

Les lignes directrices sur l'accessibilité du contenu Web (WCAG) couvrent une large gamme de recommandations visant à rendre le Web plus accessible. Le respect de ces lignes directrices rendra le contenu accessible à plus de personnes handicapées, notamment les personnes aveugles ou malvoyantes, sourdes ou malentendantes, celles qui ont des difficultés d'apprentissage, dont les fonctions cognitives ou les mouvements sont limités, les personnes qui ont des difficultés d'élocution, les personnes photosensibles ou celles qui ont plusieurs de ces problèmes à la fois. Le respect de ces lignes directrices rendra aussi le contenu Web plus utile pour les utilisateurs en général. La version électronique de la Politique se trouve à l'adresse suivante : http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505/

Pour mesurer l'accessibilité des sites Web, le World Wide Web Consortium (w3c) a établi une série de 14 lignes directrices ou principes généraux régissant la conception accessible. Chaque ligne directrice est assortie d'un certain nombre de points de contrôle. Ces points de contrôle sont établis en fonction de la gradation WCAG 1.0, WCAG 2.0 et WCAG 3.0, d'après leur impact sur l'accessibilité.

- [Priorité 1 / WCAG 1.0] Il faut qu'un développeur de contenu Web vérifie ce point de contrôle. Autrement, il sera impossible pour un ou plusieurs groupes d'avoir accès à l'information contenue dans le document. Ce point de contrôle est une condition fondamentale pour permettre à certains groupes de pouvoir se servir des documents Web.
- [Priorité 2 / WCAG 2.0] Il faut qu'un développeur de contenu Web vérifie ce point de contrôle. Autrement, il sera difficile pour un ou plusieurs groupes d'avoir accès à l'information contenue dans le document. Ce point de contrôle permettra d'éliminer des obstacles importants à l'accès des documents Web.
- Priorité 3 / WCAG 3.0] Un développeur de contenu Web peut vérifier ce point de contrôle. Autrement, il sera un peu difficile pour un ou plusieurs groupes d'avoir accès à l'information contenue dans le document. Ce point de contrôle permettra d'améliorer l'accès aux documents Web.

L’accessibilité peut se mesurer à 3 niveaux; ces niveaux sont les suivants :

- Niveau de conformité A: tous les points de contrôle Priorité 1 WCAG de chaque ligne directrice ont été vérifiés;
- Niveau de conformité AA : tous les points de contrôle Priorité 1 et 2 WCAG ont été vérifiés;
- Niveau de conformité AAA : tous les points de contrôles Priorité 1, 2 et 3 ont été vérifiés pour chaque guide

## 7. Exigences obligatoires

| Réf. no | Description |
|---|---|
| M1 | La solution de masquage des données doit être fonctionnelle et complète au niveau de l'exécution, de l'habilitation et du soutien, inclure tous les éléments qui contribuent à la composition du tout ou des parties, comme l'indiquent la DP, ses appendices et ses annexes. La liste complète du nom et de la version de chaque composante logicielle sous licence livrée dans le cadre de la solution doit être fournie.<br><br>La solution doit permettre à EC d'atteindre les objectifs spécifiques dont les détails sont donnés dans la section 3 de l'énoncé des besoins.<br><br>La solution de masquage des données doit fonctionner dans l'environnement informatique décrit dans la section 6 du présent document sur l'environnement informatique. |
| M2 | La solution de masquage des données fournie à EC doit être reproductible.<br><br>Autrement dit, si les données de production n'ont pas changé, l'exécution du masquage des données doit donner exactement les mêmes résultats, chaque fois. |
| M3 | La solution de masquage des données doit permettre de préserver l'intégrité référentielle des données.<br><br>Autrement dit, les relations préexistantes avec d'autres éléments de données définies à l'intérieur de la base de données Oracle doivent être maintenues après l'exécution du masquage des données (cas d'utilisation : Électeur-adresse) |
| M4 | La solution de masquage des données ne doit pas créer de relation bijective.<br><br>Autrement dit, le masquage des données ne doit pas redonner le même résultat pour toutes les occurrences d'une valeur donnée (par exemple, toutes les occurrences de « Smith » ne doivent pas toujours redonner « Jones ») |
| M5 | La solution de masquage des données doit produire des données masquées lisibles humainement.<br><br>Autrement dit, le résultat du masquage des données doit se présenter à l'utilisateur final sous formes de données de production réalistes (Exemple : « Smith » apparaît sous la forme de « Jones » – pas « xcvtr ») |

| Réf. no | Description |
|---|---|
| M6 | La solution de masquage des données doit fournir à la fois des règles intégrées de masquage des données pour les procédures de masquage communes et des règles de masquage définies par l'utilisateur.<br><br>Exemples :<br>Règle de masquage intégrée : (exemple : règle de masquage du NAS)<br>Règle de masquage définie par l'utilisateur : (exemple : nom de famille correspondant à une adresse). |
| M7 | La solution de masquage des données doit permettre de masquer les données en place.<br><br>Autrement dit, la solution doit permettre à EC de masquer les données directement dans la base de données demandée sans qu'il soit nécessaire de procéder à des ETL pour charger les données dans un environnement particulier de masquage des données afin de masquer les données, pour ensuite les extraire de cet environnement particulier et recouvrir l'environnement original. |
| M8 | La solution de masquage des données doit permettre de masquer les données statiques.<br><br>Dans le domaine de la technologie de masquage des données, c'est c e qu'on appelle couramment le masquage de données statiques (SDM). Autrement dit, le résultat de l'opération de masquage des données DOIT être stocké directement dans la base de données en remplacement du contenu original. Il n'est en aucune façon possible alors de visualiser les données sensibles statiques, à l'intérieur de la base de données ou à partir d'une sauvegarde de la base d e données. |
| M9 | La solution de masquage des données doit générer des données masquées non réversibles.<br><br>Autrement dit, il ne doit y avoir aucun moyen de voir ou de déduire la valeur originale de l'élément de données après l'exécution de la procédure de masquage des données. |
| M10 | La solution de masquage des données doit permettre d'utiliser des règles multiples en même temps.<br><br>Exemple : 1) reproductibilité, 2) sexe, 3) unité familiale<br><br>John Smith 100 rue Main (père) - devient - Frank Black 100 rue Main<br>Barb Smith 100 rue Main (mère) - devient - Nancy Black 100 rue Main<br>Bob Smith 100 rue Main (fils) - devient - Dave Black 100 rue Main<br>Sue Jones 100 rue Main (grand-mère) - devient - Linda Franks 100 rue Main<br>George Smith 200 rue Main (voisin) - devient - Pat Brown 200 rue Main |

| Réf. no | Description |
|---|---|
| M11 | La solution de masquage doit permettre de faire appel à la technique de masquage par « substitution ».<br><br>Voici les techniques acceptées de masquage :<br>• Substitution, par exemple : remplacement/ substitution des données par d'autres<br>• Suppression, par exemple : suppression de la colonne de données sensibles<br>• Caviardage, par exemple : des astérisques au lieu des 6 premiers chiffres du NAS<br>• Généralisation, par exemple : remplacement d'une valeur par un intervalle ou une série, par exemple 21 pour une valeur comprise entre 20 et 30.<br>• Remaniement, par exemple : substitution de données, utilisation des données existantes comme base pour les données de remplacement.<br>• Randomisation, par exemple : replacement des données par une donnée de valeur randomisée. |
| M12 | La solution de masquage des données doit permettre à EC de fournir des données de substitution source.<br><br>Autrement dit, la procédure de masquage des données doit permettre à EC de se servir du registre actuel des électeurs comme source de valeurs de substitution des noms, dans le cas du masquage du nom des électeurs. |
| M13 | La solution de masquage des données doit permettre de conserver les règles de masquage des données en vue d'une réutilisation ultérieure.<br><br>Autrement dit, la solution doit permettre à l'utilisateur de configurer les règles de masquage des données, puis de les réutiliser à de multiples reprises à des dates ultérieures. Cette capacité permettra d'éviter d'avoir à entrer à nouveau les règles de masquage des données chaque fois qu'on procède à des opérations de masquage. |
| M14 | La solution de masquage des données doit permettre de masquer les données de la base de données version Oracle Enterprise (base de données standard d'EC).<br><br>Cette capacité ne doit pas être considérée comme limitant la solution uniquement à la base de données Oracle, mais la base de données Oracle doit être l'un des environnements de base de données primaires et non pas un environnement exigeant des ajouts après-coup. La série de caractères surtout utilisée pour la base de données Oracle est WE8MSWIN1252, mais parfois, la série AL32UTF8 est utilisée. |
| M15 | La solution de masquage des données doit permettre de masquer les données sur une plateforme virtuelle Oracle Enterprise LINUX (OEL). |

| Réf. no | Description |
|---|---|
| | Cette capacité ne doit pas être considérée comme limitant la solution uniquement à une plateforme LINUX, mais LINUX doit être l'un des environnements primaires et non pas un environnement exigeant des ajouts spéciaux après-coup. La plateforme LINUX est normalement utilisée par EC pour les bases de données Oracle et les technologies de virtualisation sont largement employées à EC. La solution de base doit donc être compatible avec elles. |
| M16 | La taille ou le volume de données masquées par la solution de masquage des données ne doit pas être limitée.<br><br>Les données à masquer avec la solution MD occupent actuellement 800 gigaoctets, la solution doit permettre de traiter ce volume et toute augmentation éventuelle subséquente au fil des ans. |

# Exigences techniques cotées

| Réf. no | Description |
|---|---|
| R1 | La solution de masquage des données devrait être très extensible pour permettre de traiter le volume de données à masquer.<br><br>Autrement dit, la solution devrait permettre à EC de masquer tous les électeurs (env. 45 millions) au sein de la base de données du registre dans n'importe lequel de nos environnements de développement actuels (actuellement au nombre de 4, soit les environnement de développement, de mise à l'essai, d'AQ et de pré-production) en l'espace d'au plus 8 heures, sur la base des configurations existantes des bases de données et des serveurs {configuration actuelle d e la base de données du registre : 4 UC, 10G RAM, OEL 5.8, VM ware 4.1 avec Oracle 11GR2} |
| R2 | La solution de masquage des données devrait comporter une fonctionnalité de rétablissement en cas de défaillance éventuelle.<br><br>Autrement dit, en cas de défaillance éventuelle (par exemple : insuffisance de l'espace disque ou des segments de repositionnement, etc.), la solution devrait permettre de poursuivre le masquage à partir de l'endroit où il a été interrompu une fois que la situation est redevenue normale. EC ne devrait pas avoir à recommencer l'opération de masquage à partir du début simplement à cause d'une défaillance. |
| R3 | La solution de masquage des données devrait permettre l'utilisation de techniques de masquage en plus de la « substitution ».<br><br>Voici les techniques de masquage généralement acceptées :<br>• Substitution, par exemple : remplacement/ substitution des données par d'autres<br>• Suppression, par exemple : suppression de la colonne de données sensibles<br>• Caviardage, par exemple : des astérisques au lieu des 6 premiers chiffres du NAS<br>• Généralisation, par exemple : remplacement d'une valeur par un intervalle ou une série, par exemple 21 pour une valeur comprise entre 20 et 30.<br>• Remaniement, par exemple : substitution de données, utilisation des données existantes comme base pour les données de remplacement.<br>• Randomisation, par exemple : replacement des données par une donnée de valeur randomisée. |
| R4 | La solution de masquage des données devrait permettre de chiffrer toutes les données conservées pour usage interne au moyen des normes de chiffrement du GdC. |

| | |
|---|---|
| | La plupart des produits de masquage créent des données de traitement provisoires en vue d'une utilisation à court terme ou à long terme, ou les deux. Par exemple, mentionnons le schéma de mise en correspondance utilisé dans le cas de la technique de substitution. Si ce schéma n'est pas conservé de manière sécurisée, il peut servir à annuler le masquage des données. Toutefois, une mise en correspondance pourrait être nécessaire pour répéter la même opération de masquage des données à une date ultérieure. Pour éviter des failles éventuelles, la solution devrait permettre de conserver toutes ces données pour usage interne de manière sécurisée.<br><br>Pour connaître les normes de chiffrement du GdC, consulter l'adresse suivante : http://www.cse-cst.gc.ca/its-sti/services/crypto-services-crypto/ca-ac-fra.html |
| R5 | La solution de masquage des données devrait comporter une interface graphique (GUI) conviviale.<br><br>Autrement dit, la solution devrait comporter une forme ou une autre d'interface GUI permettant l'interaction de la communauté des utilisateurs, par opposition à des commandes lancées à partir de la ligne de commande du système d'exploitation. |
| R6 | La solution de masquage des données devrait permettre de faire un choix parmi des règles de masquage pré-chargées.<br><br>Autrement dit, la solution devrait permettre à l'utilisateur de ne pas être obligé d'appliquer toutes les règles de masquage à chaque opération de masquage, mais de choisir la série de règles à exécuter à chaque opération de masquage. |
| R7 | La solution de masquage des données devrait être assortie de mesures de sécurité intégrées pour limiter l'accès des utilisateurs.<br><br>Autrement dit, la solution devrait être fournie directement avec une forme quelconque de protection de l'accès contre l'utilisation non autorisée de l'outil. L'exemple le plus simple de mesure de sécurité de l'accès des utilisateurs serait la protection par nom d'utilisateur et mot de passe, avec maintenance des justificatifs d'identité par l'outil. Un mécanisme plus complexe consisterait pour l'outil à recourir à un protocole LDAP comme Active Directory de Microsoft. Une autre possibilité perfectionnée consisterait à permettre à l'outil de soutenir de multiples profils/rôles pour les utilisateurs (par exemple Administrateur, Opérateur, Lecture seule, etc.). |
| R8 | La solution de masquage des données devrait permettre la sauvegarde.<br><br>Autrement dit, la solution devrait offrir les capacités nécessaires pour fonctionner sans interruption avec HP Data Protector (version 6 +). HP Data Protector est l'outil de sauvegarde employé par EC. |

| | |
|---|---|
| R9 | La solution de masquage des données devrait permettre d'empêcher des opérations simultanées visant la même base de données cible.<br><br>Autrement dit, la solution devrait permettre d'éviter, lors d'une opération de masquage dans une base de données, qu'une opération de masquage soit exécutée en même temps dans cette même base de données. On évitera ainsi les problèmes de lancement accidentel ou de recommencement des opérations de masquage dans une même base de données visée. À noter que cela ne signifie pas que la solution sert à un seul utilisateur à la fois, mais plutôt qu'il ne peut pas y avoir d'opérations de masquage en même temps pour une même base de données cible. |
| R10 | La solution de masquage des données devrait permettre de générer à la fois des rapports préconstruits et des rapports configurables.<br><br>Autrement dit, la solution devrait permettre de produire une série de rapports préconstruits (rapports types) fournissant à l'utilisateur des statistiques sur l'exécution du masquage des données et d'autres renseignements consignés pendant l'exécution du masquage des données, comme les erreurs survenues, des dénombrements, etc. La solution doit aussi permettre à l'utilisateur de créer de nouveaux rapports supplémentaires ou de configurer les rapports types originaux. |
| R11 | La solution de masquage des données devrait comporter une fonction intégrée de vérification configurable.<br><br>La solution devrait avoir une capacité intégrée de vérification de toutes les actions (exemples : changement au niveau des règles, etc.) prises de l'intérieur ou de l'extérieur (exemples : entrées des heures d'exécution). |

## Acronymes

| Acronyme | Description |
|---|---|
| BDGE | Bureau du directeur général des élections |
| DGEC | Directeur général des élections du Canada |
| COTS | Commercial (Commercial off the shelf) |
| TI | Technologie de l'information |
| EC | Élections Canada |
| R-D | Recherche et développement |
| OEL | Oracle Enterprise Linux |
| MV | Machine virtuelle |
| RAM | Mémoire vive (Random Access Memory) |
| UC | Unité centrale |
| AQ | Assurance de la qualité |
| ID | Informatique décisionnelle |
| IE | Internet Explorer |
| SCT | Secrétariat du Conseil du Trésor du Canada |
| GI | Gestion de l'information |
| DP | Demande de propositions |
| NSI | Normalisation des sites Internet |
| WCAG | Lignes directrices sur l'accessibilité du contenu Web (Web Content Accessibility Guidelines) |
| W3C | World Wide Web Consortium |
| GUI | Interface graphique (Graphical User Interface) |
| EB | Énoncé des besoins |
| NAS | Numéro d'assurance sociale |