

# GÉNÉRALISATION DES DONNÉES SENSIBLES

Arthur D. Chapman et  
Oliver Grafton

**Citation suggérée:**

Chapman, A. D. et Grafton, O. (2008). *Généralisation des données sensibles*, version 1.0. Trad. Chenin, C. Copenhague: Global Biodiversity Information Facility, 27 pp. Disponible en ligne sur [http://links.gbif.org/gbif\\_generalisation\\_donnees\\_sensibles\\_manual\\_fr\\_v1.pdf](http://links.gbif.org/gbif_generalisation_donnees_sensibles_manual_fr_v1.pdf)

ISBN/DoI: non disponible.

Persistent URI:

[http://links.gbif.org/gbif\\_generalisation\\_donnees\\_sensibles\\_manual\\_fr\\_v1.pdf](http://links.gbif.org/gbif_generalisation_donnees_sensibles_manual_fr_v1.pdf)

Langage du document: Français (Titre original: *Generalising Sensitive Data*)

Date de publication de la traduction en Français: Mai 2011.

Date de publication de la version originale en Anglais: 2008

Copyright © Global Biodiversity Information Facility, 2011

Licence:



Cette travail est publié sous une licence Creative Commons Paternité 3.0

<http://creativecommons.org/licenses/by/3.0/deed.fr>

La traduction en Français de ce texte a été cofinancée par le Système Mondial d'Information sur la Biodiversité (GBIF) et la Direction Générale de la Coopération au Développement Belge (DGD) au travers du projet *Central African Biodiversity Information Network* (CABIN) coordonné par le Musée Royal d'Afrique Centrale (RMCA). Le GBIF France a aussi contribué à cette traduction.



GBIF



DGD



CABIN



RMCA



GBIF France

**Dégagement de responsabilité:** Même si tous les efforts ont été mis en œuvre par l'auteur et l'éditeur pour veiller à ce que le contenu de ce texte soit précis et à jour, tous les renseignements contenus ici sont donc présentés «tels quels», sans autre déclaration ou garantie de quelque nature que ce soit. Ni l'auteur ni l'éditeur sont en aucun cas responsable de tout dommage pouvant découler de l'utilisation de l'information contenue dans le présent texte.

## Généralisation des données sensibles

Introduction.....	1
Principes.....	2
Déterminer la sensibilité .....	4
Critères pour déterminer la sensibilité .....	4
Catégories de sensibilité .....	9
Lister les taxons sensibles.....	12
Généraliser l'information textuelle.....	13
Généraliser l'information spatiale.....	14
Documentation.....	15
Documentation et métadonnées .....	15
Documenter la sensibilité.....	16
Authentification et autorisation.....	18
Références bibliographiques.....	19
Annexe : Scénarii d'utilisation des critères 1 et 2 pour déterminer la sensibilité.....	20
Glossaire .....	23
Index .....	24

Ce article est équivalent à :

Chapman, A. D. et O. Grafton. 2008. *Guide to Best Practices for Generalising Primary Species-Occurrence Data*, version 1.0. Copenhague: Global Biodiversity Information Facility. 27 pp.  
ISBN: 87-92020-06-2 (disponible sous forme d'un document PDF indépendant sur le site <http://www.gbif.org/orc>)

## Introduction

La dissémination sans précaution de données primaires d'occurrence d'espèces (par exemple les localisation exactes de taxons rares, menacés et ayant une valeur commerciale) est depuis le premier jour une préoccupation constante du Secrétariat du GBIF. Dès le début 2006, le GBIF a initié un processus visant à traiter ce problème, particulièrement en relation avec les données destinées à être partagées à travers le réseau GBIF et mises à disposition sur le portail de données du GBIF.

Un inventaire des approches actuelles pour diminuer la précision de telles données a été lancée en Février 2006 avec notamment un inventaire en ligne a été conduit à l'aide du logiciel « Survey Monkey<sup>1</sup> ». Un rapport séparé sur les résultats a été publié sur le site GBIF<sup>2</sup> début Juin 2006 (Chapman 2006). Et un atelier d'experts s'est ensuite tenu début Mars 2007, consacré aux divers aspects techniques en jeu (Chapman 2007a).

Un rapport final sur la manière de traiter les données primaires d'occurrence d'espèces sensibles a été établi à l'issue de ce travail et de ces échanges, et il fut présenté au GBIF en Avril 2007 (Chapman 2007b). Il est disponible sur le site du GBIF. Ce rapport faisait un certain nombre de recommandations, dont beaucoup sont incluses dans ce document-ci.

La dernière étape de ce processus a consisté à rédiger un Guide des meilleures pratiques. Ce document devrait être considéré comme un guide de référence pour les institutions, les fournisseurs de données et les points nodaux GBIF, à partir duquel ils pourraient développer leur propres règles de fonctionnement. Organismes et institutions devraient produire leur propre document interne, qui incorpore les pratiques exposées dans ce chapitre et les documents reliés comme le Guide des meilleures pratiques pour le géo-référencement (Chapman et Wieczorek 2006) en adaptant le tout à leur propre environnement de travail.

Il est aussi important de comprendre l'impact que ces approches restrictives pour les données sensibles peut avoir sur la science de la biodiversité, et, tout en restreignant la disponibilité ou la résolution de certaines données, il faut essayer de ne pas restreindre excessivement les usages potentiels de ces données. Pour cette raison, un ensemble de principes est exposé ci-dessous. Parmi ceux-ci un principe clé est de répondre au besoin de rendre l'information sur la biodiversité libre d'accès chaque fois que c'est possible, dans l'intérêt de la science, de l'environnement et de la biodiversité elle-même.

Deux aspects que ce document ne couvre pas, parce qu'ils demandent une discussion et un consensus plus élaborés avant que des recommandations robustes puissent être faites, sont le respect de chaque être vivant et le développement d'accords relatifs au partage des données et aux droits sur les données. Ces deux aspects ont des implications légales et varient considérablement d'une juridiction à l'autre. Des recommandations ont été faites dans le Rapport sur la manière de traiter les données d'occurrence d'espèces sensibles (Chapman 2007) remis au GBIF qui doit poursuivre la réflexion sur ces aspects.

*« L'expression 'meilleure pratique' désigne généralement la meilleure manière de faire quelque chose ; elle est utilisée couramment dans le monde de la gestion, de l'ingénierie logicielle et de la médecine, et de plus en plus aussi pour les gouvernements [...] L'expression 'meilleure pratique courante' est souvent plus précise, laissant entendre qu'il reste une marge d'amélioration. »*

*(Wikipedia: Bonne Pratique).*

---

<sup>1</sup> Survey Monkey <http://www.surveymonkey.com>

<sup>2</sup> [http://www.gbif.org/prog/digit/sensitive\\_data/Summary\\_of\\_Responses\\_-\\_03.pdf](http://www.gbif.org/prog/digit/sensitive_data/Summary_of_Responses_-_03.pdf)

## Principes

*L'information sur la biodiversité devrait être en libre accès et partagées à l'échelle mondiale afin d'être utilisées à des fins non commerciales pour l'aide à la décision, l'éducation, la recherche et autres applications d'intérêt public. Avoir accès à l'intégralité de l'information disponible sur la biodiversité devrait réduire le risque de dommage à l'environnement et aider à préserver un avenir durable. Dans les cas où sa diffusion serait dangereuse l'accès à l'information détaillée devrait être contrôlé.*

Voici un ensemble de grands principes relatifs au partage des données en général, et à celui des données sensibles en particulier :

- 1. A chaque fois que c'est possible, l'information environnementale devrait être en accès libre pour tous. Ceci bénéficie en général à l'environnement en accroissant la prise de conscience, en permettant une meilleure prise de décision et en réduisant les risques de dommage.*
- 2. Dans un petit nombre de cas, l'accès public à l'information peut être dommageable pour l'environnement. Il devrait être reconnu que dans de tels cas, il se peut que l'accès à l'information doive être contrôlé. La diffusion est quand même préférable par défaut et toute restriction devrait être limitée au strict nécessaire.*
- 3. Toute donnée jugée sensible devrait inclure une date de révision de son statut de sensibilité, ainsi qu'une justification argumentée de ce statut. Le délai avant révision peut être court ou long selon la nature de la sensibilité. A chaque fois qu'un fournisseur de données reçoit une demande d'extension d'accès aux données sensibles, il devrait éviter de présupposer la perpétuation de la sensibilité et utiliser cette occasion pour la réévaluer.*
- 4. Si la précision doit être réduite pour la diffusion, elle ne devrait affecter qu'une copie au moment de la diffusion. Les données ne devraient jamais être altérées, falsifiées ou supprimées dans l'enregistrement qui a été saisi..*
- 5. La documentation est essentielle pour de nombreuses raisons, et lorsque les données ont été restreintes en accès ou en précision pour la diffusion, il est important que cette information soit mémorisée dans les métadonnées conservées avec l'enregistrement.*
- 6. Lorsque les données sont restreintes en accès ou en précision pour la diffusion (par exemple par omission du nom du collecteur ou de l'information textuelle sur la localité, etc.) ceci devrait être documenté en remplaçant l'information occultée par une inscription appropriée : le champ ne devrait pas être laissé blanc ni vide.*
- 7. Il y a d'extrêmement bonnes raisons de ne pas restreindre les données de collecte associées (comme les numéros de collecte, qui sont en séquence, le nom du collecteur, etc.) à cause des limitations que cela impose aux procédures de qualité et de validation des données, et à l'efficacité des techniques dites « push<sup>3</sup> » filtrées.*

---

<sup>3</sup> « La technique 'push' est un mode de communication client-serveur dans lequel le dialogue est lancé par le serveur. Cette technique s'oppose au fonctionnement « classique » des transactions Web où le client ouvre le dialogue, et tire vers lui l'information (*pull*). Le client s'abonne au service et dès qu'une nouvelle information est disponible, elle est envoyée par le serveur.

8. *Un utilisateur de données sensibles devrait respecter toute restriction d'accès ou de précision que le fournisseur aura placée sur ses données. Si on lui accorde la levée de la restriction, il ne doit pas rompre la confidentialité de l'information.*
9. *Les fournisseurs de données devraient respecter le besoin qu'ont les utilisateurs d'accéder aux données et à la documentation afin de déterminer l'« aptitude à l'usage » des données, et d'assurer que leurs analyses sont robustes et ne donnent pas des résultats trompeurs.*

---

Des applications telles que la visioconférence ou la messagerie instantanée sont de parfaits exemples d'utilisation de cette technique : dès que le serveur central reçoit un message (ou n'importe quel autre type de données) de l'un des participants, il l'envoie aux autres ». (extrait de Wikipedia le 23/07/2010) (NdT).

---

## Déterminer la sensibilité

La première étape consiste pour chaque détenteur d'information à identifier toute donnée qui est considérée comme « sensible ». Une information est qualifiée de « sensible » si sa diffusion publique est susceptible d'entraîner un effet néfaste sur le taxon ou l'attribut en question ou sur un être vivant. Un certain nombre de facteurs doivent être pris en compte lorsqu'on détermine la sensibilité, y compris le type et le niveau de la menace, la vulnérabilité du taxon ou de l'attribut, le type d'information, et le fait que l'information soit déjà en accès public. Ces facteurs fournissent des critères objectifs pour déterminer la sensibilité.

Deux exemples de critères de sensibilité qui fournissent un point de départ pour le développement d'autres critères sont donnés par le Réseau national de la biodiversité du Royaume Uni (en anglais : « National Biodiversity Network », NBN) (National Biodiversity Network 2002, 2004), et par le Département de l'environnement et de la conservation des Nouvelles Galles du Sud en Australie (en anglais, « Department of Environment and Conservation ») (« Department of Environment and Conservation, 2007).

Nous présentons ci-dessous une série de critères pour déterminer la sensibilité des taxons et des données ainsi que des commentaires recommandés comme métadonnées pour documenter les arguments de cette détermination. Les deux premiers sont destinés aux fournisseurs de données sur la biodiversité et ceux qui éditent des listes de taxons potentiellement sensibles : ils se réfèrent largement aux taxons eux-mêmes. Les deux derniers sont destinés aux fournisseurs des données sur la biodiversité et traitent de l'évaluation des données détenues et que les fournisseurs envisagent de rendre accessibles : ils ne sont pas adaptés à l'édition de listes de taxons sensibles.

Les critères servent à déterminer :

1. <b>Risque de dommage</b>	Indique si le taxon est sujet à une activité humaine dommageable.
2. <b>Impact du dommage</b>	Evalue la sensibilité du taxon à une activité humaine dommageable.
3. <b>Sensibilité des données</b>	Evalue si la diffusion des données augmentera le dommage.
4. <b>Décision de diffusion &amp; Catégorie de sensibilité</b>	Décision raisonnée concernant la diffusion des données, et détermination des catégories de sensibilité, et donc du niveau de précision et des attributs diffusables.

Un ensemble de scénarii mettant en œuvre les critères 1 et 2 ci-dessous pour déterminer les seuils de sensibilité de taxons est joint dans l'Annexe à ce chapitre.

### ***Critères pour déterminer la sensibilité***

La première étape du processus de détermination de la sensibilité consiste à évaluer si le taxon est ou non menacé par une activité humaine dommageable et si la disponibilité des données sur la biodiversité associées augmentera la probabilité d'occurrence de l'activité dommageable. Si ce n'est pas le cas, alors il ne semble pas y avoir de raison de le lister

comme un taxon potentiellement sensible. Il est recommandé d'utiliser les expressions fournies ci-dessous, en ajoutant une justification décrivant les aspects spécifiques de la menace, par exemple :

« Le taxon est menacé par une activité humaine dommageable : il est susceptible d'être attaqué par *Phytophthora* qui est transporté par les véhicules. »

<b>1. RISQUE DE DOMMAGE</b>	
<b>Evaluer si le taxon est menacé par une activité humaine dommageable.</b>	
1.1. Le taxon est-il menacé par une activité humaine dommageable ?	<p><b>Oui</b> : Documenter en utilisant l'expression <b>1a</b> avec la justification appropriée. Aller en 1.2</p> <p><b>Non</b> : Documenter en utilisant l'expression <b>1b</b> [Le taxon n'est pas sensible] Aller en 3</p>
1.2. Y a-t-il des indices d'occurrence actuelle ou récente de l'activité dommageable ?	<p><b>Oui</b> : Documenter en utilisant l'expression <b>1c</b> avec la justification appropriée. Aller en 1.3</p> <p><b>Non</b> : Documenter en utilisant l'expression <b>1d</b> avec la justification appropriée. Aller en 1.3</p>
1.3. La disponibilité des données sur la biodiversité associées augmentera-t-elle la probabilité d'occurrence de l'activité dommageable ?	<p><b>Oui</b> : Documenter en utilisant l'expression <b>1e</b> avec la justification appropriée. Aller en 2</p> <p><b>Non</b> : Documenter en utilisant l'expression <b>1f</b> avec la justification appropriée. Aller en 2</p>

1a – Le taxon est menacé par une activité humaine dommageable.
1b – Il n'y a pas de risque significatif lié à une activité humaine.
1c – Il y a des indices avérés de dommage actuel ou récent causé au taxon.
1d – Il y n'a pas d'indice avéré de dommage actuel ou récent causé au taxon.
1e – La disponibilité des données sur la biodiversité augmentera la probabilité d'occurrence d'activité humaine dommageable.
1f – La disponibilité des données sur la biodiversité n'augmentera pas la probabilité d'occurrence d'activité humaine dommageable.

L'étape suivante consiste à déterminer si le taxon est sensible à cette activité humaine, ou s'il est suffisamment robuste pour ne pas être affecté.

<b>2. IMPACT DU DOMMAGE</b>	
<b>Evaluer la sensibilité du taxon à l'activité humaine dommageable.</b>	
2.1. Est-ce que le taxon a des caractéristiques qui le rendent significativement vulnérable à l'activité humaine dommageable ?	<p><b>Oui</b> : Documenter en utilisant l'expression <b>2a</b> avec la justification appropriée. Aller en 2.2</p> <p><b>Non</b> : Documenter en utilisant l'expression <b>2b</b> avec la justification appropriée. Aller en 2.2</p>

2.2. Est-ce que le taxon est vulnérable à l'activité humaine sur toute son aire de répartition, ou y a-t-il des zones (comme des aires protégées, ou d'autres régions du monde) où le taxon n'est pas aussi menacé ?	<b>Oui</b> : Documenter en utilisant l'expression <b>2c</b> avec la justification appropriée. Aller en 3
	<b>Non</b> : Documenter en utilisant l'expression <b>2d</b> avec la justification appropriée. Aller en 3

2a – Le taxon a des caractéristiques qui le rendent significativement vulnérable à l'activité humaine.
2b – Le taxon n'est pas significativement vulnérable à l'activité humaine.
2c – Le taxon est vulnérable à l'activité humaine sur toute son aire de répartition.
2d – Le taxon n'est pas vulnérable à l'activité humaine sur toute son aire de répartition. <b>et/ou</b> il y a des zones où le taxon est présent mais ne court pas un risque significatif.

S'il a été décidé que le taxon est menacé par une activité humaine avec un risque d'impact significatif, il faut décider si la diffusion des données en question sur ce taxon augmenterait le risque et l'impact.

<b>3. SENSIBILITE DES DONNEES</b>	
<b>Evaluer si la diffusion des données augmentera la menace.</b>	
3.1. Est-ce que le contenu et détail des données sur la biodiversité est tel que leur diffusion permettrait à quelqu'un de nuire au taxon ?	<b>Oui</b> : Documenter en utilisant l'expression <b>3a</b> avec la justification appropriée. Aller en 3.2
	<b>Non</b> : [Les données ne sont pas sensibles] Documenter en utilisant l'expression <b>3b</b> avec la justification appropriée Aller en 4
3.2. Est-ce que l'information est déjà dans le domaine public, ou déjà connue des individus ou groupes susceptibles d'entreprendre une activité dommageable ?	<b>Oui</b> : Documenter en utilisant l'expression <b>3d</b> avec la justification appropriée. Aller en 3.3
	<b>Non</b> : Documenter en utilisant l'expression <b>3c</b> avec la justification appropriée. Aller en 3.3
3.3. Est-ce la divulgation menacerait un partenariat ou une relation dont le maintien est essentiel pour atteindre un objectif spécifique de conservation ?	<b>Oui</b> : Documenter en utilisant l'expression <b>3e</b> avec la justification appropriée. Aller en 3.4
	<b>Non</b> : Documenter en utilisant l'expression <b>3f</b> avec la justification appropriée. Aller en 3.4
3.4. Est-ce la divulgation permettrait de déduire les localisations d'éléments sensibles en combinant cette information avec celle qui est déjà dans le domaine public ?	<b>Oui</b> : Documenter en utilisant l'expression <b>3g</b> avec la justification appropriée. Aller en 4
	<b>Non</b> : Documenter en utilisant l'expression <b>3h</b> avec la justification appropriée. Aller en 4

3a – Le contenu et détail des données sur la biodiversité est tel que leur diffusion permettrait à quelqu'un de nuire au taxon.
3b – Le contenu et détail des données sur la biodiversité, si elles étaient diffusées, ne permettrait pas à quelqu'un de nuire au taxon.
3c – L'information <u>n'est pas</u> dans le domaine public, et n'est pas déjà connue d'individus ou de groupes susceptibles d'entreprendre une activité dommageable.
3d – L'information est déjà dans le domaine public, et connue d'individus ou de groupes susceptibles d'entreprendre une activité dommageable.
3e – La divulgation des données <u>menace</u> un partenariat ou une relation dont le maintien est essentiel pour atteindre un objectif spécifique de conservation.
3f – La divulgation des données <u>ne menace aucun</u> partenariat ou relation essentiels pour la conservation.
3g – La divulgation <u>permettrait</u> de déduire les localisations d'éléments sensibles en combinant cette information avec celle qui est déjà dans le domaine public.
3h – La divulgation <u>ne permettra pas</u> de déduire les localisations d'éléments sensibles en combinant cette information avec celle qui est déjà dans le domaine public.

La dernière étape consiste à effectuer une évaluation globale sur la base des trois critères ci-dessus, et à documenter la décision globale en combinant les justifications qui accompagnent les décisions élémentaires qui supportent la décision globale. Une fois qu'on a décidé si les données doivent ou non être diffusées, il est important de déterminer la catégorie de sensibilité, et le niveau de généralisation pour diffuser les données.

<b>4. DECISION SUR LA DIFFUSION &amp; CATEGORIE DE SENSIBILITE</b>	
<b>Prendre une décision globale quant à la diffusion des données et déterminer la catégorie et le niveau de généralisation</b>	
4.1. Globalement, en considérant les critères 1 à 3 ci-dessus et tout élément important dans le contexte, est-ce que la diffusion des données augmente le risque de dommage à l'environnement ou à un être vivant ?	<b>Oui : Documenter en utilisant l'expression <i>4a.</i></b>  <b>Aller en 4.2</b>
	<b>Non : Documenter en utilisant l'expression <i>4b.</i></b>  <b>Aller en 4.5</b>
4.2. Est-ce que le taxon est particulier et d'une grande importance biologique, fortement menacé par l'exploitation, la maladie, ou tout autre menace identifiable, de sorte que même une information de localisation générale pourrait le menacer ? Ou encore est-ce que la diffusion de quelque partie que ce soit de l'enregistrement pourrait causer un dommage irréparable à l'environnement ou à un individu ?	<b>Oui : Documenter en utilisant l'expression <i>4c,</i> regrouper tout l'argumentaire et documenter la décision de ne pas diffuser les données.</b>  <b>Aller à la Catégorie 1</b>
	<b>Non :</b>  <b>Aller en 4.3</b>

4.3. Est-ce que le taxon est tel que la diffusion de localisation précises au-delà de 0,1 degré (~10 km) lui imposerait une menace de perturbation ou d'exploitation ? Ou encore l'enregistrement contient-il de l'information hautement sensible dont la diffusion pourrait entraîner un dommage extrême à un individu ou à l'environnement ?	<b>Oui :</b> <i>Documenter en utilisant l'expression 4d, regrouper tout l'argumentaire et documenter la décision de diffuser les données.</i> <b>Aller à la Catégorie 2</b>
	<b>Non :</b> <b>Aller en 4.4</b>
4.4. Est-ce que le taxon est tel que la diffusion de localisation précises au-delà de 0,1 degré (~10 km) lui imposerait une menace de collecte ou de dégâts délibérés ? Ou encore l'enregistrement contient-il de l'information sensible dont la diffusion pourrait entraîner un dommage à un individu ou à l'environnement ?	<b>Oui :</b> <i>Documenter en utilisant l'expression 4e, regrouper tout l'argumentaire et documenter la décision de diffuser les données.</i> <b>Aller à la Catégorie 3</b>
	<b>Non :</b> <b>Aller en 4.5</b>
4.5. Est-ce que le taxon est susceptible d'être faiblement ou moyennement menacé si des localisations précises (précision supérieure à 0,001 degré ou 100 m) sont rendues publiques où il y aurait un risque de collecte ou de dégâts délibérés ?	<b>Oui :</b> <i>Documenter en utilisant l'expression 4f, regrouper tout l'argumentaire et documenter la décision de diffuser les données.</i> <b>Aller à la Catégorie 4</b>
	<b>Non :</b> <i>Documenter en utilisant l'expression 4g, regrouper tout l'argumentaire et documenter la décision de diffuser les données.</i> <b>Les données devraient être rendues publiques</b>

4a – Tout bien pesé, la diffusion de l'information augmentera le risque de dommage environnemental ou de dommage à un être vivant.
4b – Tout bien pesé, la diffusion de l'information n'augmentera pas le risque de dommage environnemental ou de dommage à un être vivant.
4c – L'espèce est une espèce particulière de grande importance biologique, elle est menacée par l'exploitation, la maladie ou d'autres menaces identifiables, et même une information de localisation générale peut être dommageable, ou encore la diffusion de l'information pourrait causer des dommages irréparables à l'environnement, à un individu, ou à d'autres éléments. <b>[Catégorie 1]</b>
4d – L'espèce est classée comme hautement sensible, et la divulgation de localisations précises exposerait l'espèce à des menaces de perturbation et d'exploitation, et/ou l'enregistrement contient de l'information sensible dont la diffusion pourrait entraîner des dommages graves à l'environnement ou à un individu. <b>[Catégorie 2]</b>
4e – L'espèce est classée comme moyennement à hautement sensible, et la divulgation de localisations précises exposerait l'espèce à des menaces de collecte ou de dégâts délibérés, et/ou l'enregistrement contient de l'information sensible dont la diffusion pourrait entraîner des dommages à l'environnement ou à un individu. <b>[Catégorie 3]</b>

4f – L'espèce est classée comme faiblement à moyennement sensible, et la divulgation de localisations précises exposerait l'espèce à des menaces de perturbation ou d'exploitation. Des données détaillées peuvent être diffusées à certaines personnes sous licence. **[Catégorie 4]**

4g – L'espèce est classée comme faiblement sensible, et la divulgation de localisations précises ne devrait pas exposer l'espèce à une réelle menace, et/ou l'enregistrement contient une information peu sensible, dont la diffusion à peu de chance d'entraîner de dommage à l'environnement ou à un quelconque individu. Les données devraient être rendues publiques « telles quelles ».  
**[Pas sensible sur le plan de l'environnement]**

Au cours de l'enquête en ligne, un certain nombre de personnes ont répondu que des données en attente de publication, des données utilisées dans des recherches en cours, et des données incomplètes ou non vérifiées devraient à leur sens être classées comme sensibles et donc sujettes à restriction dans leur diffusion. Ce type de sensibilité est de courte durée et il est alors important qu'une date de publication ou de revue du classement sensible soit clairement indiquée.

De telles données tomberaient vraisemblablement sous le critère 3.3 ci-dessus et devraient être documentées en conséquence avec comme justification « *en attente de publication* », etc.

**NB.** *Toute donnée considérée comme sensible devrait inclure une date de revue de son statut sensible, associée à l'argumentaire qui justifie ce statut. Le délai avant la revue peut être court ou long selon la nature de la sensibilité.*

Les catégories de sensibilité (ci-dessous) sont largement fondées sur celles du département de l'environnement et de la conservation des Nouvelles Galles du Sud.

## Catégories de sensibilité

Critère	Raisonnement
<b>Catégorie 1</b> – Espèce pour laquelle aucun enregistrement ne sera fourni, ou ne sera diffusé qu'avec une simple indication de présence dans une large région, comme un pays, un bassin versant, etc.	La raison de ne pas diffuser est que : <ol style="list-style-type: none"><li>1. Une espèce <i>particulière</i> de <b>grande importance biologique</b> est <b>gravement menacée</b> par l'exploitation / la maladie ou une autre menace identifiable et même une information générale sur la localisation peut nuire au taxon.</li><li>2. L'information dans l'enregistrement est de telle nature que sa diffusion pourrait causer un dégât irréparable à l'environnement, à un individu ou à d'autres éléments.</li></ol> Les données ne peuvent être fournies que sous des conditions de licence strictes, ou réduites à une indication de présence à l'échelle d'une grande région : pays, bassin versant ou région biogéographique.

Critère	Raisonnement
<p><b>Catégorie 2</b> – Espèce pour laquelle les coordonnées seront accessibles au public “dénaturées” (à 0.1 degré) et/ou les autres informations dans l’enregistrement seront généralisées. Des données à une échelle plus fine (Catégorie 3 ou 4 ou données détaillées) peuvent être fournies au cas par cas sous Licence.</p>	<p>Les raisons de la restriction sont que :</p> <ol style="list-style-type: none"> <li>1. L’espèce est classée comme <i>hautement sensible</i>, et la divulgation des localisations précises <i>exposerait</i> l’espèce à des menaces de perturbation et d’exploitation.</li> <li>2. L’enregistrement contient de l’information <i>hautement sensible</i> dont la diffusion pourrait causer un dommage <i>grave</i> à un individu ou à l’environnement.</li> </ol> <p>Les données sont fournies au public</p> <ol style="list-style-type: none"> <li>1. avec un géoréférencement dénaturé à 0.1 degré (~10 km) et/ou</li> <li>2. avec ses champs sensibles généralisés ou supprimés et remplacés par une information appropriée.</li> </ol> <p>Les données peuvent être fournies à une échelle plus fine sur demande sous condition d’un accord écrit sur ces données, habituellement un Accord de licence sur les données. Quand les données sont fournies aux clients, ceux-ci sont informés de quelles espèces ou quels champs sont sensibles et peuvent avoir leurs coordonnées dénaturées au niveau des données disponibles sous les Catégories 3 ou 4.</p> <p><b>NB.</b> Dans le cas où la sensibilité est liée à des champs autres que le géoréférencement, il peut être plus approprié de classer les enregistrements dans les Catégories 3 ou 4.</p>

Critère	Raisonnement
<p><b>Catégorie 3</b> – Espèce ou enregistrements dont les coordonnées seront rendues publiques sous forme ‘dénaturée’ (à 0.01 degré) et/ou où d’autres informations dans l’enregistrement sont généralisées. Des données à une échelle plus fine (Catégories 3 ou 4 ou données détaillées) peuvent être fournies au cas par cas sous licence.</p>	<p>Les raisons de la restriction sont que :</p> <ol style="list-style-type: none"> <li>1. L’espèce est classée comme étant <i>moyennement à hautement sensible</i>, et la divulgation des localisations précises <i>pourrait</i> exposer l’espèce à des menaces comme la perturbation et l’exploitation.</li> <li>2. L’enregistrement contient de l’information <i>sensible</i>, dont la diffusion pourrait causer des dommages à un individu ou à l’environnement.</li> </ol> <p>Les données sont fournies public</p> <ol style="list-style-type: none"> <li>1. avec un géoréférencement dénaturé à 0.01 degré (~ 1 km) et/ou</li> <li>2. avec les champs sensibles généralisés ou supprimés et remplacés par une information appropriée.</li> </ol> <p>Les données peuvent être fournies à une échelle plus fine sur demande sous conditions d’un accord écrit sur les données, habituellement un Accord de licence sur les données. Quand les données sont fournies aux clients, ceux-ci sont informés de quelles espèces ou quels champs sont sensibles et peuvent avoir leurs coordonnées dénaturées au niveau des données disponibles sous la Catégorie 4.</p> <p><b>NB.</b> Dans le cas où la sensibilité est liée à des champs autres que le géoréférencement, il peut être plus approprié de classer les enregistrements dans la Catégorie 4.</p>
<p><b>Catégorie 4</b> – Espèce ou enregistrements dont les coordonnées seront rendues publiques sous forme ‘dénaturée’ (à 0.001 degré) et/ou où d’autres informations dans l’enregistrement sont généralisées. Des données détaillées sous leur forme originale peuvent être fournies au cas par cas sous licence.</p>	<p>Les raisons de la restriction sont que :</p> <ol style="list-style-type: none"> <li>1. L’espèce est classée comme étant <i>faiblement ou moyennement sensible</i>, et la divulgation des localisations précises pourrait faire courir un risque de collecte ou de dégâts délibérés.</li> <li>2. L’enregistrement contient de l’information <i>sensible</i>, dont la diffusion pourrait causer des dommages à un individu ou à l’environnement.</li> </ol> <p>Les données détaillées peuvent être fournies sous conditions d’un accord écrit sur les données written data agreement, habituellement un Accord de licence sur les données. Quand les données sont fournies aux clients, ceux-ci sont informés de quelles espèces ou quels champs sont sensibles.</p>

## Lister les taxons sensibles

Les données sont déjà distribuées de par le monde, y compris au travers de doubles de spécimens, et bien que les données puissent être en diffusion restreinte dans certaines institutions, d'autres qui détiennent des doubles peuvent diffuser la même information. Ceci peut venir de l'ignorance de ce qui peut être considéré comme sensible dans les zones d'habitat naturel du taxon concerné, du fait qu'aucune liste universelle des taxons « sensibles » n'est encore disponible. A ces difficultés s'ajoute le fait qu'un taxon peut être sensible dans une région, et pas dans une autre (et peut même être une espèce envahissante ou nuisible dans cette autre région).

C'est pourquoi il a été recommandé qu'une liste de taxons potentiellement sensibles pour l'environnement soit établie et liée au catalogue<sup>4</sup> de noms du GBIF. Ceci aurait l'avantage d'alerter les fournisseurs de données dans d'autres juridictions qu'une espèce est potentiellement sensible, et via ECat fournirait des liens vers les synonymes. Il est important de noter que cette liste devrait être utilisée pour signaler le besoin éventuel d'une décision sur la sensibilité réelle du partage d'information fondée sur les critères décrits ci-dessus, et pas pour justifier des restrictions de précaution. Les espèces en danger ne sont pas toutes menacées par la connaissance de leur localisation et elles ne devraient pas être considérées comme sensibles *en soi* : la liste des taxons potentiellement sensibles pour l'environnement devrait donc être beaucoup plus courte que toutes les listes existantes d'espèces rares ou menacées.

La liste devrait être créée en utilisant les critères 1 et 2 (se référer au chapitre précédent et aux scénarii en Annexe 1), et devrait contenir des informations supplémentaires comme :

- Nom du Taxon
- Critères et argumentaire pour l'inclusion
- Nom de la personne ou de l'organisme responsable de cette inclusion
- Couverture géographique de la sensibilité (surtout si le taxon n'est sensible que sur une partie de son aire de distribution ou au sein d'une juridiction)
- Catégorie de sensibilité recommandée
- Date de révision

Les juridictions peuvent aussi souhaiter tenir à jour une liste similaire pour leurs propres besoins, et il est alors recommandé qu'elles incluent les informations ci-dessus dans tous les cas. L'avantage de rendre l'information plus largement accessible est qu'elle avertira les autres conservateurs de données que votre juridiction considère ce taxon comme potentiellement sensible, et les utilisateurs seront ainsi avertis qu'ils doivent tenir compte de cette sensibilité lorsqu'ils publient les résultats de leurs analyses, etc.

**NB.** *Toute liste de taxons potentiellement sensibles pour l'environnement devrait être utilisée en faisant attention à ce caractère « potentiel », et toute restriction sur la disponibilité des données devrait être appliquée au cas par cas en prenant en compte les critères listés ci-dessus.*

---

<sup>4</sup> Catalogue électronique [des noms scientifiques] du GBIF (en anglais, « GBIF Electronic Catalogue ») <http://www.gbif.org/prog/ecat>

## Généraliser l'information textuelle

Il y a des cas où l'information contenue dans des champs texte pourrait être considérée comme sensible dans certaines circonstances. Ceci comprend des informations comme :

- Les noms de personnes vivantes
- L'information sur la localité
- La date de collecte
- Le numéro de collecte
- L'habitat
- L'information sur le propriétaire du terrain
- Les noms taxonomiques

Il se peut que certains de ces champs doivent être en accès restreint afin d'empêcher les analyses de corrélation qui pourraient conduire à déduire les localités de récolte d'enregistrements qui sont en accès restreint ou généralisés : par exemple, le nom du collecteur, la date, et les numéros de collecte en séquence. Dans d'autres cas, il peut être nécessaire de cacher le nom d'un taxon dans une liste de récoltes effectuées dans un point chaud de la biodiversité ou dans une localité sensible.

De telles restrictions ne devraient cependant pas couvrir la totalité de l'enregistrement. Les données qui doivent être cachées peuvent être retirées et remplacées par une information appropriée (voir ci-dessous), ou bien généralisées : par exemple, donner seulement le nom de plus haut rang taxonomique lorsque l'espèce doit être en accès restreint.

**NB.** *A chaque fois que des données dans un champ texte sont restreintes ou généralisées pour la diffusion (comme le nom d'un collecteur, la description textuelle de la localité, etc.) cela devrait être documenté en remplaçant les données par une information appropriée : le champ ne devrait pas rester blanc ou vide.*

Les exemples d'information de remplacement comprennent :

- « *nom supprimé pour raisons de protection de la vie privée* » ;
- « *ce spécimen représente une espèce en danger ou menacée. La localité précise a été supprimée de l'enregistrement en ligne afin de protéger l'espèce de collecte excessive. Ces données peuvent être fournies aux chercheurs sur demande* » ;
- « *ce spécimen représente une espèce en danger ou menacée. La localité a été généralisée à la présence dans un maillage de 0,1 degré de résolution. Les données détaillées peuvent être fournies aux chercheurs sur demande* ».

**NB.** *Lorsqu'on doit restreindre des noms taxonomiques (par exemple, ceux de taxons sensibles au sein d'un inventaire), il est possible de les remplacer par des noms de taxons de rang supérieur (genre / famille, etc.), ou d'indiquer seulement la présence de 'x' taxons sensibles, sans préciser leur nom.*

Parfois, les fournisseurs de données peuvent être tentés de restreindre l'information dans des enregistrements liés à un enregistrement sensible (en plus de l'enregistrement sensible lui-même), comme le nom du collecteur et les numéros en séquence des enregistrements

provenant de collectes effectuées au même endroit et au même moment afin de réduire la possibilité que l'enregistrement sensible puisse être retrouvé avec une analyse de corrélation. Cependant, si le nom du collecteur et le numéro de collecte sont supprimés seulement de l'enregistrement sensible et pas des autres, il est peu probable que cette information puisse être retrouvée par une personne extérieure. Pour cette raison, et d'autres (voir l'encadré ci-dessous), il est recommandé que les données des enregistrements voisins ne soient pas restreintes.

**NB.** *Il y a de très bonnes raisons pour ne pas restreindre les données sur les collectes voisines (numéros de collecte en séquence, nom du collecteur, habitat, etc.) à cause des restrictions que cela impose aux procédures de contrôle qualité et de validation des données et aux limites que cela impose à l'efficacité des techniques « Push » filtrées. L'information des enregistrements voisins d'un enregistrement sensible (mais pas celle de l'enregistrement sensible lui-même) ne devraient pas être restreintes à moins que ce ne soit absolument nécessaire..*

## Généraliser l'information spatiale

L'une des demandes les plus fréquentes en matière de généralisation de l'information sur la biodiversité concerne la localité ou le géoréférencement. Traditionnellement cela a été fait de nombreuses manières, avec peu de cohérence entre les méthodes, et très peu de documentation sur ce qui a été fait dans chaque cas. Ceci a considérablement réduit la valeur des données pour l'analyse, et les utilisateurs ignorent souvent que les données ont été modifiées.

La bonne pratique demande de documenter toute généralisation des données afin que les utilisateurs sachent quelle confiance leur accorder.

A la suite de longues discussions entre les fournisseurs et les utilisateurs des données, il a été décidé de recommander que les fournisseurs qui généralisent leurs données utilisent une méthodologie normalisée (voir ci-dessous), et joignent la documentation associée. Comme la plupart des données sur la biodiversité sont actuellement mises à disposition avec des degrés décimaux, la méthode recommandée ne demande pas de modification des formats (comme Darwin Core) autre que de permettre une documentation appropriée en métadonnées.

La méthode recommandée ci-dessous autorise plusieurs niveaux de généralisation selon les Catégories 1 à 4 décrites au chapitre antérieur *Déterminer la sensibilité*.

La méthode recommandée pour la généralisation est la suivante :

Catégorie	Sensible	Géoréférencement
<b>Catégorie 1</b>	Extrême	Géoréférencement non diffusé, éventuellement indication seule du bassin versant, du pays ou de la biorégion.
<b>Catégorie 2</b>	Haute	Géoréférencement arrondi à 0.1 degré
<b>Catégorie 3</b>	Moyenne	Géoréférencement arrondi à 0.01 degré
<b>Catégorie 4</b>	Basse	Géoréférencement arrondi à 0.001 degré
<b>Pas sensible</b>	Pas sensible	Géoréférencement non restreint

## Documentation

Il est important de documenter la méthode et le niveau de généralisation afin que les utilisateurs sachent ce qui a été fait aux données, et quelle est leur fiabilité. Actuellement, aucun des deux schémas XML d'échange de données Darwin Core et ABCD ne prévoient de champ pour les métadonnées qui documentent la généralisation et la restriction. Il a toutefois été recommandé que ces schémas –et les protocoles associés DiGIR et BioCASE- soient modifiés pour prendre en compte ces métadonnées (voir le chapitre *Documentation et Métadonnées* ci-dessous), mais en attendant il est recommandé d'enregistrer ces informations dans des champs commentaires.

En ce qui concerne la généralisation du géoréférencement il est important d'enregistrer le fait que les données ont été généralisées à l'aide d'un maillage géographique décimal, et de mémoriser les deux informations :

- Précision de la donnée fournie (par exemple : 0,1 degré ; 0,001 degré ; etc.)
- Précision de la donnée stockée (par exemple : 0,0001 degré ; 0,1 minute ; 1 seconde ; etc.)

Les métadonnées qu'il est recommandé d'inclure dans l'extension « [Définitions de l'information géospatiale](#) » du schéma Darwin Core (TDWG 2005) sont explicitées dans le prochain chapitre *Documentation et Métadonnées*. Lorsque cette extension (ou quelque chose d'équivalent) aura été adoptée, il est recommandé d'enregistrer et de diffuser les champs correspondant en même temps que les données.

**NB.** Si l'information géographique est généralisée à une large région comme un bassin versant, une région biogéographique ou un pays, alors il ne faut pas fournir le géoréférencement.

## Documentation et métadonnées

Il est important que les données soient documentées avec précision afin que les utilisateurs et autres parties prenantes sachent exactement ce que représentent les données et quelle confiance leur accorder. L'utilisateur doit par exemple être en mesure de déterminer si les données sont adaptées à l'analyse qu'il souhaite effectuer. De nombreux fournisseurs ont indiqué lors de l'enquête que l'une des raisons pour lesquelles ils étaient réticents à diffuser certaines de leurs données était leur crainte qu'elles soient utilisées à tort. Si les données ne sont pas correctement documentées, la probabilité d'usages non pertinents involontaires s'en trouve considérablement accrue puisque l'utilisateur peut les intégrer dans son analyse en croyant à tort qu'il dispose d'un géoréférencement précis, alors qu'en réalité la référence spatiale a été généralisée à des mailles de 10 km, et que la localisation réelle pourrait se trouver n'importe où au sein d'une zone de 100 km<sup>2</sup>. S'il fait tourner un modèle de climat, par exemple, une telle erreur pourrait le conduire à tirer des résultats des conclusions gravement erronées. Pour cette seule raison, il est important à la fois pour les fournisseurs, pour les utilisateurs directs et pour les utilisateurs finaux (comme les gestionnaires de l'environnement, les décideurs, etc.) que les données soient correctement décrites.

En particulier, il devrait y avoir une documentation claire des « Contraintes d'accès » qui pourraient inclure, par exemple, une indication des parties sensibles dans les données (lorsqu'il y en a), des raisons de cette sensibilité et des conditions dans lesquelles la diffusion est possible.

## Documenter la sensibilité

« Les métadonnées remplissent une fonction essentielle dans la communication vers les tiers concernant les contraintes d'accès et les conditions d'utilisation que les créateurs des données souhaitent voir respecter. Elles peuvent être considérées comme une « aide » dans la protection des données et de l'information, du fait qu'elles permettent aux utilisateurs de visualiser les conditions établies par les créateurs des données pour l'accès et l'usage de leur information. En outre, lorsque les données ne sont pas accessibles, les métadonnées permettent de connaître les conditions d'accès via d'autres média (numériques ou non) ainsi que le résumé de leurs contenu » (Llinás, 2005).

L'expression « métadonnées » a généralement été utilisée pour désigner la documentation relative à un jeu de données entier. La documentation au niveau de l'enregistrement a habituellement été traitée seulement sous forme de commentaires. Je préfère cependant désigner la documentation à ce niveau par l'expression « *métadonnées du niveau enregistrement* » (voir le glossaire) et formaliser le processus. Dans le chapitre précédent il a été recommandé, lorsque les données ont été généralisées pour la diffusion, de documenter le niveau de généralisation : par exemple, indiquer que les données ont été généralisées à l'aide d'un maillage géographique décimal, et enregistrer à la fois la précision des données fournies et celle des données stockées. De plus, dans le chapitre « *Déterminer la sensibilité* », une série de processus de documentation ont été recommandés. Certains de ceux-ci peuvent être plus adaptés pour documenter les raisons pour lesquelles on considère un taxon comme potentiellement sensible pour l'environnement (Critères 1 et 2), tandis que les autres (Critères 3 et 4) sont plus adaptés aux données elles-mêmes et permettent de générer les métadonnées du niveau enregistrement. Afin de documenter pleinement les raisons de restreindre les données, il peut toutefois être nécessaire de répercuter au niveau de l'enregistrement la documentation correspondant aux critères 1 et 2 : par exemple, les raisons pour lesquelles les données sont restreintes peuvent être entre autres que le taxon est menacé par une activité humaine.

Pour le moment, ni le schéma Darwin Core, ni le schéma ABCD ne contiennent de champs permettant d'enregistrer le type de métadonnées du niveau enregistrement qui est recommandé ici. Un certain nombre de recommandations ont été proposées au TDWG pour intégrer des champs supplémentaires au schéma Darwin Core sous la forme de l'extension « [Définitions de l'information géospatiale](#) » (TDWG 2005) et/ou dans le schéma Darwin Core lui-même. Les recommandations incluent celles indiquées dans le tableau qui est présenté sur la page suivante.

Le champ « Commentaires sur les données sensibles » ici est peut-être équivalent au champ « Contraintes d'accès » que l'on trouve dans la plupart des métadonnées au niveau du jeu de données. L'information au niveau du jeu de données peut contenir quelque chose comme :

*« Ce jeu de données n'est accessible au public qu'avec une résolution sommaire pour la raison suivante. Certaines informations contenues dans le jeu de données ont trait à des espèces vulnérables aux perturbations ou préjudices de l'activité humaine. Deux espèces (*Adelanthus lindenbergianus*, *Athalamia hyaline*) sont très vulnérables à la collecte. Tous les détails de cette information sensible peuvent être fournis sous*

*licence aux organismes et individus spécifiques qui ont besoin de les connaître afin d'éviter la nuisance à l'environnement. SVP contacter le fournisseur pour plus d'information. »*

En attendant que les schémas d'échange normalisés soient amendés, il est recommandé de documenter les données à l'aide de champs commentaire, et autant que possible d'enregistrer le même type d'information qui serait incluse dans les champs recommandés ci-dessus, c'est-à-dire :

- Le fait que les données sont sensibles ;
- Les principales raisons pour lesquelles les données sont considérées comme sensibles (voir les Critères 1 à 4 dans le chapitre « Déterminer la sensibilité ») ainsi que les justifications afférentes ;
- La date à laquelle la sensibilité devrait être révisée ;
- La précision des données qui sont accessibles ;
- La précision des données originales.

Champ	Commentaires
<b>IndicateurDonnéeSensible</b>	O/N indique si l'observation est sensible ou pas.
<b>RaisonDonnéeSensible</b>	La raison principale de la sensibilité de la donnée. Le format suggéré est soit une liste de choix avec des valeurs dérivées des Critères 1 à 4 ci-dessus (ou un champ texte qui combine les indications 1a à 4g associées à ces critères).
<b>CommentairesDonnéeSensible</b>	Information supplémentaire sur la ou les raison(s) ou argumentaire pour déterminer la pertinence des Critères pour cet enregistrement comme recommandé ci-dessus. [Texte libre]
<b>RévisionDonnéeSensible</b>	Un champ date indiquant quand la nature sensible de la donnée devrait être révisée. Particulièrement important lorsque la sensibilité est seulement liée à l'attente d'une publication.
<b>PrécisionDonnéeFournie</b>	L'échelle ou la précision de la donnée fournie dans l'enregistrement Darwin Core ; peut être codé par exemple comme ceci : <ul style="list-style-type: none"> <li>• 0 = 1 degré</li> <li>• 1 = 0.1 degré</li> <li>• 2 = 0.01 degré</li> <li>• 3 = 0.001 degré</li> <li>• 4 = 0.0001 degré</li> </ul>
<b>PrécisionDonnéeStockée</b>	L'échelle ou la précision de la donnée stockée ; peut-être codée de la même manière : <ul style="list-style-type: none"> <li>• 0 = 1 degré</li> <li>• 1 = 0.1 degré</li> <li>• 2 = 0.01 degré</li> <li>• 3 = 0.001 degré</li> <li>• 4 = 0.0001 degré</li> <li>• Etc. ou</li> </ul> Peut-être encore du texte libre, comme '1 minute', '0.1 minute', '1 seconde', etc. selon comment la donnée est stockée.

## Authentification et autorisation

Selon les recommandations de l'atelier des experts, et aussi de nombreuses réponses à l'enquête en ligne, la responsabilité de déterminer qui peut ou ne peut pas avoir accès aux détails des données sensibles, par exemple avec un accès sécurisé par identifiant et mot de passe, ou sous forme d'un lot de données délivré sous licence, doit rester celle des fournisseurs.

Les membres de l'atelier sont aussi tombés d'accord sur le fait que ce n'est pas le rôle du GBIF de gérer l'identification, la vérification et l'autorisation des utilisateurs, non plus que de contrôler l'authentification sur le portail de données, mais son rôle peut être de fournir des conseils et une méthode d'authentification appropriée aux points nodaux.

Il a été rapporté à l'atelier qu'il n'est pas difficile de résoudre les problèmes techniques liés à l'authentification d'un groupe ou d'un individu, et l'utilisation de rôles. Il y a plusieurs protocoles et modes opératoire bien établis pour l'authentification qui sont couramment utilisés et qui pourraient être aisément adaptés à l'usage des fournisseurs de données.

Le principal problème consiste à déterminer qui devraient être les utilisateurs autorisés et qui sont les utilisateurs de bonne foi. C'est un problème difficile qui demande que l'on prenne le temps de la réflexion. Il n'y a pas sur ce point de recommandation précise à faire dans le présent document sur les meilleures pratiques ; toutefois le rapport antérieur (Chapman 2007b) a fait à ce sujet un certain nombre de recommandations sur la manière d'avancer dans cette réflexion.

Il a été recommandé que le GBIF explore le problème de l'authentification en vue de fournir des mécanismes appropriés qui aident les fournisseurs de données. Il est donc recommandé que les fournisseurs qui souhaitent développer un système d'authentification sécurisé pour leurs bases de données en discutent avec le GBIF, ou avec leur point nodal GBIF.

La recommandation faite au GBIF dans le rapport antérieur (Chapman 2007b) était que :

*Le GBIF explore le problème de l'authentification en vue de fournir des mécanismes appropriés qui aident les fournisseurs de données à identifier les utilisateurs qui peuvent accéder à plus de détails et de quelle manière ils peuvent le faire. Bien que le GBIF ne doive pas (à ce stade tout au moins) s'occuper d'enquêter sur les utilisateurs, ou de placer des contrôles sur le portail, il est clairement dans son rôle de fournir des conseils et d'aider les points nodaux à implémenter une méthode appropriée et robuste pour l'authentification.*

## Références bibliographiques

- Chapman, A.D. 2006. *Questionnaire on Dealing with Sensitive Primary Species Occurrence Data – Summary of responses*. 61 pp. Copenhagen: GBIF.  
[http://www.gbif.org/prog/digit/sensitive\\_data/Summary\\_of\\_Responses\\_-\\_03.pdf](http://www.gbif.org/prog/digit/sensitive_data/Summary_of_Responses_-_03.pdf) [Accessed 8 Apr. 2007].
- Chapman, A.D. 2007a. Workshop on Dealing with Sensitive Species Occurrence Data. Held at: NatureServe Offices, Arlington, Virginia, USA. 6-7 March 2007. Report. Copenhagen: GBIF. 30 pp. <http://www.gbif.org/>
- Chapman, A.D. 2007b. *Dealing with Sensitive Primary Species Occurrence Data. Report*. Report to the Global Biodiversity Information Facility 60pp. <http://www.gbif.org/>. Copenhagen: GBIF.
- Chapman, A.D. and Wiczorek, J. (eds). 2006. *Guide to Best Practices for Georeferencing*. BioGeomancer Consortium. Copenhagen: Global Biodiversity Information Facility. 90pp. ISBN: 87-92020-00-3.  
<http://www.gbif.org/prog/digit/Georeferencing> **See also Chapter 5 in this Manual.**
- Department of Environment and Conservation – NSW. 2007. *Threatened Species Information Disclosure Policy* (Version 3 Amended March 2007).  
[http://www.nationalparks.nsw.gov.au/npws.nsf/content/sensitive\\_species\\_policy](http://www.nationalparks.nsw.gov.au/npws.nsf/content/sensitive_species_policy) [Accessed 15 Mar 2007].
- Llinás, J.V. 2005. *Data and Information on Biodiversity and its Protection in the Digital Realm* Ver. 1. Bogotá, Colombia: Biological Resources Research Institute Alexandre von Humboldt. 43pp.
- National Biodiversity Network Trust. 2002. *NBN Data Exchange Principles*. Version 3.2, April 2002.  
<<http://www.nbn.org.uk/downloads/files/DataExchange%20principles%202002.pdf>> [Accessed 27 Mar 2007].
- National Biodiversity Network Trust. 2004. *The 'Environmental Exception' and access to information on sensitive features*. Version 1.3.2, Countryside Agencies' Open Information Network Environmental Information Regulations Guidance Note No. 1. Linked from [www.nbn.org.uk/eir](http://www.nbn.org.uk/eir) [Accessed 27 Mar 2007].
- TDWG. 2005. *Geospatial Extension to Darwin Core*. Taxonomic Databases Working Group.  
<http://wiki.tdwg.org/twiki/bin/view/DarwinCore/GeospatialExtension> [Accessed 1 Apr 2007].

## Annexe : Scénarii d'utilisation des critères 1 et 2 pour déterminer la sensibilité

Les scénarii suivants montrent comment les indications associées aux critères listés dans le chapitre « Déterminer la sensibilité » peuvent être utilisés pour élaborer des déclarations résumées qui documentent les raisons de considérer un taxon comme sensible. La déclaration résumée (colonne de droite) devrait aussi inclure l'argumentaire, comme les types de dommages spécifiques, etc. Par exemple dans le second scénario (**B**), la déclaration complète peut être quelque chose comme :

*« Les taxons pourraient être menacés par les agents pathogènes transportés sur les roues des engins forestiers, mais l'occurrence n'est pas affectée par la disponibilité des données. »*

Ceci peut s'appliquer à une espèce de plante dans une zone forestière sensible à une attaque de *Phytophthora*, les champignons étant transférés sur les roues des engins forestiers.

### Critère 1

#### Scénario A

Déclaration(s) du critère	Déclaration résumée
1a – Il n'y a pas de risque significatif lié à une activité humaine dommageable.	Le taxon n'est pas sensible.

#### Scénario B

Déclaration(s) du critère	Déclaration résumée
1a – Le taxon est menacé par une activité humaine dommageable.	Le taxon pourrait être menacé mais la probabilité du dommage n'est <b>pas</b> affectée par la disponibilité des données.
1d – Il n'y a pas actuellement d'indice établi d'un dégât avéré au taxon.	
1f – La disponibilité des données sur la biodiversité n'accroîtra pas la probabilité d'occurrence de l'activité humaine dommageable.	

### Scénario C

Déclaration(s) du critère	Déclaration résumée
1a – Le taxon est menacé par une activité humaine dommageable.	Le taxon pourrait être menacé et la probabilité du dommage <b>est</b> affectée par la disponibilité des données.
1d – Il n’y a pas actuellement d’indice établi d’un dégât avéré au taxon.	
1e– La disponibilité des données sur la biodiversité accroîtra la probabilité d’occurrence de l’activité humaine dommageable.	

### Scénario D

Déclaration(s) du critère	Déclaration résumée
1a – Le taxon est menacé par une activité humaine dommageable.	Le taxon est menacé et il y a des indices qui le prouvent, mais la probabilité du dommage n’est <b>pas</b> affectée par la disponibilité des données.
1c – Il y a des indices établis de dégâts actuels ou récents au taxon.	
1f – La disponibilité des données sur la biodiversité n’accroîtra pas la probabilité d’occurrence de l’activité humaine dommageable.	

### Scénario E

Déclaration(s) du critère	Déclaration résumée
1a – Le taxon est menacé par une activité humaine dommageable.	Le taxon est menacé et il y a des indices qui le prouvent, et la probabilité du dommage <b>est</b> affectée par la disponibilité des données.
1c – Il y a des indices établis de dégâts actuels ou récents au taxon.	
1e– La disponibilité des données sur la biodiversité accroîtra la probabilité d’occurrence de l’activité humaine dommageable.	

## Critère 2:

### Scénario F

Déclaration(s) du critère	Déclaration résumée
2b – Le taxon n’est pas significativement vulnérable à une activité humaine dommageable.	Le taxon n’est pas significativement vulnérable à l’activité dommageable, et il n’est pas vulnérable à cette activité sur toute son aire de répartition et il y a des zones où le taxon n’est pas significativement menacé par cette activité.
2d – Le taxon n’est pas vulnérable à une activité humaine dommageable sur toute son aire de répartition <b>et/ou</b> il y a des zones où le taxon n’est pas significativement menacé.	

### Scénario G

Déclaration(s) du critère	Déclaration résumée
2a – Le taxon a des caractéristiques qui le rendent significativement vulnérable à l'activité humaine dommageable.	Le taxon est significativement vulnérable à l'activité dommageable, mais il n'est pas vulnérable à cette activité sur toute son aire de répartition et il y a des zones où le taxon n'est pas significativement menacé par cette activité.
2d – Le taxon n'est pas vulnérable à une activité humaine dommageable sur toute son aire de répartition <b>et/ou</b> il y a des zones où le taxon n'est pas significativement menacé.	

### Scenario H

Déclaration(s) du critère	Déclaration résumée
2a – Le taxon a des caractéristiques qui le rendent significativement vulnérable à l'activité humaine dommageable.	Le taxon est significativement vulnérable à l'activité humaine dommageable, et il est vulnérable à cette activité sur toute son aire de répartition.
2c – Le taxon est vulnérable à une activité humaine dommageable sur toute son aire de répartition.	

## Glossaire

**Authentification :** — désigne la détermination de l'identité d'un utilisateur, ainsi que de ce qu'il est autorisé à accéder. La forme la plus commune d'authentification consiste à utiliser un identifiant et un mot de passe associé, bien que cela procure le plus bas niveau de sécurité.

**Autorisation :** — désigne le processus qui détermine quels individus peuvent se voir attribuer différents droits d'accès pour s'authentifier et accéder aux données.

**Généralisation :** — désigne ici toute modification apportée aux données source afin de cacher un contenu sensible, typiquement en réduisant la précision des données. Par exemple en ne donnant comme indication spatiale qu'un bassin versant, une maille ou un pays, ou en citant seulement le toponyme le plus proche, ou en supprimant une partie des données. En termes géographiques, la généralisation désigne la conversion d'une représentation géographique à une autre de moindre résolution et un contenu informatif réduit ; traditionnellement associée à un changement d'échelle. Aussi parfois désignée par le terme « floutage ».

**Métadonnées du niveau enregistrement :** — désigne la documentation au niveau d'un enregistrement plutôt qu'à celui d'un jeu de données complet. Dans ce document, ces métadonnées traitent largement du statut sensible de l'enregistrement (ou de l'espèce concernée) ainsi que des contraintes d'accès liées à cet enregistrement et de l'information détaillée sur toute généralisation des données.

**Donnée sensible :** — il s'agit de toute donnée que du fait de sa nature, un fournisseur ne veut pas rendre accessible en l'état, comme par exemple une localisation précise d'un taxon menacé.

# Index

## A

accord relatif au partage des données,2  
accord relatif aux droits sur les données,2  
atelier d'experts,2  
authentification,19  
autorisation,19

## C

catalogue électronique des noms du GBIF (ECat),12  
catégories de sensibilité,5,8,10  
CommentairesDonnéeSensible,18  
critères de sensibilité,4

## D

date de collecte,13  
définition d'une donnée sensible,24  
définition de l'authentification,24  
définition de l'autorisation,24  
définition de la généralisation,24  
Département de l'environnement et de la conservation  
des Nouvelles Galles du Sud, en Australie,4  
déterminer la sensibilité,4  
documentation,16  
documenter la sensibilité,16

## E

estimation de l'incertitude maximale,18  
estimation de l'incertitude maximale,18  
extension "Définitions de l'information  
géospatiale",15,17

## G

généralisation de l'information textuelle,13  
Guide des meilleures pratiques pour le géo-  
référencement,2

## H

habitat,13

## I

impact du dommage,5

impact du dommage,6  
IndicateurDonnéeSensible,18  
information sur la localité,13  
information sur le propriétaire du terrain,13  
inventaire en ligne,2

## L

listes de taxons sensibles,12

## M

maillage géographique décimal,15  
métadonnées,16

## N

noms de personnes vivantes,13  
noms taxonomiques,13  
numéro de collecte,13

## P

partage des données,2  
portail de données du GBIF,2  
portail de données GBIF,19  
PrécisionDonnéeFournie,18  
PrécisionDonnéeStockée,18

## R

RaisonDonnéeSensible,18  
Réseau national de la biodiversité du Royaume Uni,4  
RévisionDonnéeSensible,18  
risque de dommage,5  
risque de dommage,5

## S

schéma d'échange de données ABCD,15  
schéma d'échange de données Darwin Core,15  
Secrétariat du GBIF,2  
sensibilité des données,5  
sensibilité des données,7  
survey monkey,2