

GUIDE PRATIQUE DE  
LA VERSION 5  
NOTES TECHNIQUES POUR  
LES FOURNISSEURS

Tasha Mellins-Cohen

COUNTER

# SOMMAIRE

---

<b>INTRODUCTION .....</b>	<b>3</b>	<b>MODIFICATIONS DES DONNÉES .....</b>	<b>13</b>
<b>TRAÇAGE DE L'UTILISATION .....</b>	<b>4</b>	<b>Déclaration a posteriori des erreurs dans les données d'utilisation .....</b>	<b>13</b>
<b>Balilage de page.....</b>	<b>4</b>	<b>Prise en compte du changement de titre d'une revue .....</b>	<b>13</b>
Exemples de balises de page.....	4		
Fichiers journaux .....	5		
Cookies.....	6		
<b>Variations anormales d'utilisation.....</b>	<b>6</b>	<b>MISE À DISPOSITION DES RAPPORTS COUNTER .....</b>	<b>14</b>
Pic d'utilisation .....	6	<b>Sushi .....</b>	<b>15</b>
Creux d'utilisation .....	7	<b>Fichiers de données séparées par un délimiteur .....</b>	<b>15</b>
<b>Recherches.....</b>	<b>7</b>	<b>AUDIT .....</b>	<b>16</b>
<b>TRAITEMENT DES DONNÉES .....</b>	<b>9</b>	<b>Processus d'audit.....</b>	<b>16</b>
<b>Données à prendre en compte .....</b>	<b>9</b>	<b>Résultats possibles de l'audit.....</b>	<b>17</b>
<b>Données à ignorer.....</b>	<b>9</b>	Succès (Pass).....	17
<b>Codes de retour .....</b>	<b>9</b>	Succès avec conditions (Qualified pass) .....	17
<b>Sessions.....</b>	<b>9</b>	Échec (Fail).....	17
<b>Doubles clics .....</b>	<b>10</b>		
<b>Éléments et titres pris en compte une seule fois par session.....</b>	<b>11</b>		
<b>Recherches fédérées ou issues d'un robot.....</b>	<b>12</b>		
<b>Outils de téléchargement en masse .....</b>	<b>12</b>		

# INTRODUCTION

---

Ce guide vise à présenter simplement la mise en œuvre de la norme COUNTER aux éditeurs et autres fournisseurs ; il ne constitue pas un document de spécifications pour le développement informatique. Il accompagne le *Guide de la version 5 pour les fournisseurs*, qui explicite les notions de *mesure*, *d'attribut* et de *rapport* propres à cette version de COUNTER.

Pour de plus amples informations sur la mise en œuvre, veuillez consulter le *Code de bonnes pratiques COUNTER, version 5, annexe D : directives de mise en œuvre*.

# TRAÇAGE DE L'UTILISATION

---

Le traçage de l'utilisation peut être réalisé de plusieurs manières, COUNTER n'ayant pas de recommandation particulière en la matière. Les deux méthodes les plus courantes sont le *balisage de page* et *l'analyse de fichier journal*. Leurs avantages et inconvénients respectifs sont résumés ci-après. Nous insistons sur le fait que les données destinées aux rapports COUNTER doivent refléter fidèlement l'activité réelle des utilisateurs ; le mode de traçage de l'utilisation pouvant légèrement varier d'une plate-forme à l'autre, il ne nous est pas possible de fournir de méthode de nettoyage de données spécifique permettant de satisfaire aux exigences COUNTER.

## BALISAGE DE PAGE

Les balises de page sont des fragments de code informatique intégrés dans les pages du site Web. Elles sont généralement écrites en JavaScript, même si d'autres langages peuvent être utilisés, comme Java. Leur fonction est de recueillir des informations et de les consigner dans une base de données. D'autres scripts, écrits dans des langages comme JQuery ou AJAX, eux-mêmes couplés à un langage de script côté serveur comme PHP, permettent ensuite de manipuler puis d'enregistrer ces données, offrant ainsi un contrôle complet sur leur représentation. Les scripts de stockage et de manipulation des données peuvent également accéder à des informations supplémentaires liées à l'application Web cliente ou à la session de l'utilisateur ; par exemple, ils peuvent consulter les données du système de gestion des accès.

Le balisage de page est la méthode de référence dans le domaine du Web analytique (utilisée par exemple par Google Analytics). L'une des différences majeures entre l'analyse de fichier journal et le balisage de page est la suivante : dans le second cas, la mise à jour du décompte est déclenchée par l'ouverture de la page dans le navigateur, non par une requête adressée au serveur. Il en résulte une représentation potentiellement plus fidèle de l'utilisation réelle puisque les balises de page décomptent les visites des pages en cache au même titre que les requêtes adressées au serveur.

Le balisage de page est particulièrement utile aux organisations qui n'ont pas accès à leurs propres serveurs Web ; ainsi, avec l'expansion de l'hébergement dans le Cloud, cette méthode devient progressivement une référence pour recueillir les données analytiques. En matière de mise en œuvre, le balisage de page et l'analyse des données ainsi enregistrées peuvent être réalisés en interne ou confiés à des sociétés prestataires, pratique largement répandue.

### Exemples de balises de page

Nous donnons ci-dessous quelques exemples de balises de page disponibles dans l'outil Google Gestionnaire de balises, toutes directement applicables aux rapports COUNTER.

Balise	Définition Google	Application aux rapports maîtres en version 5
Page vue	Balise élémentaire qui se déclenche à chaque visite d'une page du site	Pas d'utilisation spécifique
Évènement	Balise qui se déclenche sur certaines interactions ou certains évènements, comme un clic sur un bouton	Différenciation entre <i>investigations</i> , <i>requêtes</i> , <i>recherches</i> et autres types de mesures
Durée	Balise qui détecte les vitesses de chargement sur la page	Détection des doubles clics

Pour plus d'informations sur la manière de baliser vos pages pour la génération des rapports COUNTER, consultez en premier lieu le site <https://www.google.com/analytics/>

## Fichiers journaux

Les fichiers journaux sont des fichiers au format texte où sont consignées toutes les requêtes HTTP reçues par un serveur Web, chacune comportant le nom d'hôte ou l'adresse IP de l'utilisateur, la date et l'heure de la requête, le nom du fichier faisant l'objet de la requête, l'état et la taille du message de réponse HTTP, l'URL correspondante et la signalétique du navigateur. Une requête HTTP correspond à une tentative d'accès à une page Web par un utilisateur ; la réponse HTTP est le message contenant cette page, renvoyé au navigateur de l'utilisateur.

La plupart des serveurs Web génèrent des fichiers journaux en standard, dans un format prédéfini qui peut être spécifique au type de serveur, ce qui permet de disposer des données nécessaires à la production des rapports COUNTER sans avoir à faire évoluer le site Web.

Les fichiers journaux étant stockés dans un format standard sur les serveurs propres du fournisseur, ils peuvent être exploités par une grande variété de programmes analytiques pour fournir des résultats cohérents sur la durée. Par ailleurs, ces fichiers sont indépendants du navigateur des utilisateurs, ce qui garantit un traçage homogène de l'utilisation à destination des rapports COUNTER.

Notez cependant que les visites de pages en cache ne donnent pas lieu à une requête au serveur et ne sont donc pas tracées dans ces fichiers journaux, même si elles représentent potentiellement une proportion significative des pages vues.

Chaque type de serveur Web utilisant un format propre pour ses fichiers journaux, l'analyse de ces fichiers est le plus souvent réalisée par l'équipe de maintenance des serveurs, c'est-à-dire, généralement, l'équipe interne du fournisseur.

Pour plus d'informations sur l'exploitation des fichiers journaux à destination des rapports COUNTER, nous vous recommandons de consulter votre équipe de développement informatique.

Pour plus d'informations sur les fichiers journaux en général, consultez la documentation d'Amazon Web Services sur ce sujet, bien conçue et très utile, à l'adresse <https://aws.amazon.com/fr/documentation/>

## **Cookies**

Les cookies sont des petits fichiers stockés sur l'ordinateur de l'utilisateur. Prévus pour mémoriser une quantité limitée de données concernant l'accès par un navigateur spécifique à un site Web particulier, ils sont accessibles au serveur Web et à l'ordinateur de l'utilisateur. Des balises de page peuvent être utilisées pour gérer le processus d'affectation de ces cookies au navigateur de l'utilisateur ; si la méthode de traçage employée est l'analyse des fichiers journaux, c'est le serveur Web qui doit être configuré pour prendre en charge ce processus.

Veillez à respecter les dispositions réglementaires relatives au dépôt de cookies avant toute mise en œuvre.

## **VARIATIONS ANORMALES D'UTILISATION**

D'un établissement à un autre, le seuil à partir duquel des variations extrêmes dans les statistiques d'utilisation sont considérées comme anormales peut varier ; un pic d'utilisation peut s'avérer fondé en de nombreuses occasions, c'est pourquoi nous ne pouvons pas proposer de méthode systématique pour répondre à cette situation. Les approches suivantes permettent cependant la détection d'anomalies potentielles ou d'autres événements inhabituels. Nous conseillons de les utiliser pour déclencher une intervention humaine de vérification des différentes valeurs plutôt que pour couper systématiquement tous les accès.

### **Pic d'utilisation**

On définit un pic d'utilisation, signe d'une potentielle anomalie, de la manière suivante : la mesure de l'utilisation d'un produit donné rapportée par un client sur un certain mois est au moins égale à 300 % de la moyenne mensuelle sur les douze derniers mois, augmentés de cent unités.

## CAS PRATIQUE

Pour la période de juin 2017 à mai 2018, les utilisateurs d'une certaine université ont généré une moyenne de 100 *requêtes* par mois sur un élément de contenu du titre X (*Total\_Item\_Requests*).

Or, en juin 2018, cette même mesure passe à 450 : il s'agit là d'un pic d'utilisation. En effet, cette valeur est supérieure au seuil obtenu ainsi :

- Moyenne mesurée sur les 12 derniers mois :  $Total\_Item\_Requests = 100$
- 300 % de cette moyenne :  $Total\_Item\_Requests = 300$
- plus 100 unités pour obtenir le seuil :  $Total\_Item\_Requests = 400$

## Creux d'utilisation

On définit un creux d'utilisation, signe d'une potentielle anomalie, de la manière suivante : la mesure de l'utilisation d'un produit donné rapportée par un client sur un certain mois est inférieure à 1 % de la moyenne mensuelle sur les douze derniers mois, sous réserve que cette moyenne soit au moins égale à 20 unités.

## CAS PRATIQUE

Pour la période de juin 2017 à mai 2018, les utilisateurs d'une certaine université ont généré une moyenne de 500 *requêtes* par mois sur un élément de contenu du titre X (*Total\_Item\_Requests*).

Or, en juin 2018, cette même mesure passe à 4 : il s'agit là d'un creux d'utilisation.

En effet, cette valeur est inférieure au seuil obtenu ainsi :

- Moyenne mesurée sur les 12 derniers mois :  $Total\_Item\_Requests = 500$
- 1 % de cette moyenne (seuil) :  $Total\_Item\_Requests = 5$

## RECHERCHES

Une *recherche* doit être décomptée à chaque fois que le système génère une opération visant à ramener un nouveau jeu de résultats. Ainsi, même si le système applique plusieurs modes de recherche (p. ex., une recherche en correspondance exacte, une recherche d'occurrence de mot dans le sujet et une recherche globale) pour ramener un seul jeu de résultats triés par pertinence, une seule recherche doit être décomptée.

A l'inverse, les opérations suivantes génèrent des recherches multiples :

- Recherche transmise à plusieurs bases de données à la fois, avec des jeux de résultats présentés chacun dans une fenêtre (écran de type « bento » compartimenté en plusieurs fenêtres) ou dans un onglet (présentation multi-onglet).
- Restriction d'un résultat de recherche par application d'un critère sur les facettes ou d'un filtre occasionnant une nouvelle recherche.
- Navigation dans une liste de sujets ou dans un fichier d'autorité, chaque clic sur un sujet ou une vedette lançant une recherche qui retourne un jeu de résultats.

Remarque : la résolution de lien ne compte jamais comme une recherche.

## CAS PRATIQUE

Un utilisateur clique sur le lien suivant, présent dans un courriel qu'il a reçu :

<http://www.pubalpha.org/showArticle?id=12345>

Le site résout ce lien et « accède » à l'*id* 12345. Aucune *recherche* n'est décomptée puisqu'il s'agit d'une simple opération de résolution de lien.

Après avoir lu l'article en question, l'utilisateur lance une recherche sur « histoire des antibiotiques » sur la plate-forme *pubalpha.org*, ce qui ajoute 1 à la mesure *Searches\_Platform*. Cette recherche retourne un jeu de plusieurs milliers de résultats, aussi l'utilisateur applique-t-il une restriction sur les facettes pour ne conserver que les revues ; cette nouvelle interaction ajoute une unité supplémentaire au compteur *Searches\_Platform*.



# TRAITEMENT DES DONNÉES

---

## DONNÉES À PRENDRE EN COMPTE

- Activité d'origine légitime, à l'initiative d'un utilisateur
- Requêtes fructueuses et valides : les pages qui ne se chargent pas ou les requêtes incorrectes sont à ignorer
- Éléments de contenu pertinents : les autres éléments fournis (comme les feuilles de style) sont à ignorer.

## DONNÉES À IGNORER

- Activité issue de robots
- Pages dont le chargement échoue
- Requêtes incorrectes
- Éléments fournis autres que des éléments de contenu (comme les feuilles de style).

## CODES DE RETOUR

Les codes de retour indiquent si une page Web a été correctement transmise, c'est-à-dire si elle a bien été chargée dans le navigateur de l'utilisateur. Dans les fichiers journaux du serveur Web, les requêtes reflétant un chargement réussi sont celles présentant un code de retour NCSA (*National Center for Supercomputing Applications*) valant entre 200 et 299, ou 304. La norme des codes de retour est définie et tenue à jour par le NCSA. Les événements pris en compte par le serveur Web doivent respecter cette norme. Pour plus d'informations, consultez [www.nca.illinois.edu](http://www.nca.illinois.edu).

## SESSIONS

Dans sa version 5, le Code de bonnes pratiques comprend quatre mesures où les éléments sont décomptés une seule fois par session. Par exemple, si un utilisateur télécharge le texte intégral d'un chapitre d'ouvrage trois fois durant la même session, la mesure *Unique\_Item\_Request* n'augmentera que de 1 unité. Pour évaluer ces mesures, il est donc nécessaire de tracer les sessions utilisateur.

Une session utilisateur est typiquement définie par un identifiant de session (ou ID de session) et une date de transaction. Si l'ID de session n'est pas explicitement disponible, il doit être généré en accolant plusieurs éléments : date de transaction, heure (valeur entre 0 et 24, sans minutes), et l'une des informations suivantes :

- Un identifiant d'utilisateur (ID utilisateur) enregistré
- Un cookie d'utilisateur enregistré
- Une combinaison de l'adresse IP et de l'agent d'utilisateur.

## CAS PRATIQUE

On considère la transaction suivante :

- Date et heure de transaction : 2017-06-15 13:35
- Adresse IP : 192.1.1.168
- Agent d'utilisateur : Mozilla/5.0

L'ID de session correspondant peut alors être généré selon le schéma [Adresse IP] | [Agent d'utilisateur] | [Date] | [Heure, sans minutes], ce qui donne :

192.1.1.168 | Mozilla/5.0 | 2017-06-15 | 13

## DOUBLES CLICS

Un double clic sur un hyperlien doit être compté comme une seule requête. Selon la norme COUNTER, deux clics consécutifs sont considérés comme un double clic s'ils sont espacés d'au plus 30 secondes. Par exemple, un clic à 10:01:00 suivi d'un second à 10 :01:29 doit être considéré comme un double clic ; si le second clic avait lieu à 10:01:35, il serait compté comme un clic séparé. Un double clic peut aussi être déclenché par un clic sur l'icône de rafraîchissement de la page ou le bouton de retour arrière.

Dans tous ces cas considérés comme deux tentatives pour accéder au même article en PDF à moins de 30 secondes d'intervalle, seule la seconde requête doit être retenue. Il en va de même pour toute requête supplémentaire éventuellement générée dans les 30 secondes suivantes : seule la dernière requête est à conserver.

Différentes méthodes sont possibles pour tracer les doubles clics en fonction de la manière dont l'utilisateur est authentifié sur le site. Les options sont énumérées ci-dessous, dans l'ordre croissant de leur fiabilité (l'option 4 est donc la plus fiable) :

1. Authentification de l'utilisateur par son adresse IP : c'est à cette adresse IP que les doubles clics seront rattachés. Si plusieurs utilisateurs exploitent la même adresse IP, le phénomène suivant peut se produire à l'occasion : des clics effectués par des utilisateurs différents seront interprétés comme un double clic issu d'un même utilisateur (uniquement dans le cas où ces utilisateurs cliquent exactement sur le même contenu dans une fenêtre de 30 secondes).
2. Mise en place d'un cookie de session : c'est ce cookie qui enregistrera les doubles clics.
3. Mise en place d'un cookie d'utilisateur : c'est ce cookie qui enregistrera les doubles clics.
4. Utilisateur connecté sous son propre profil : les doubles clics seront rattachés à ce nom d'utilisateur.

## ÉLÉMENTS ET TITRES PRIS EN COMPTE UNE SEULE FOIS PAR SESSION

Deux *mesures* COUNTER, *Unique\_Item\_Requests* et *Unique\_Item\_Investigations*, prennent en compte l'utilisation des éléments une seule fois au sein d'une même session. Un *élément* est une unité de contenu à laquelle les utilisateurs accèdent : articles, chapitres d'ouvrages et contenu multimédia, par exemple ; il doit porter un identifiant unique comme un DOI (*digital object identifier*). Le *titre* est l'entité parent à laquelle les éléments appartiennent, comme la revue ou la base de données. Si un utilisateur génère une requête pour le même élément à plus d'une occasion dans une même session, elle doit être décomptée une seule fois.

De même, deux *mesures* COUNTER, *Unique\_Title\_Requests* et *Unique\_Title\_Investigations*, prennent en compte l'utilisation des titres une seule fois au sein d'une même session. Ici, un *titre* serait typiquement un ouvrage ou une revue servant de contenant pour des *éléments*.

Si un utilisateur génère une requête pour le même titre à plus d'une occasion dans une même session, elle doit être décomptée une seule fois.

### CAS PRATIQUE

Un utilisateur lance une recherche sur l'histoire des antibiotiques sur la plate-forme éditeur Alpha. Parmi les résultats de sa recherche, il choisit d'ouvrir trois résumés d'article et un enregistrement vidéo. Ces quatre éléments sont différents, mais deux des articles se trouvent être issus de la même revue. Le décompte correspondant est alors le suivant :

- Total\_Item\_Investigations : 4
- Unique\_Item\_Investigations : 4
- Unique\_Title\_Investigations : 3

Après lecture des résumés, l'utilisateur télécharge la version PDF de deux des articles, tous deux de la même revue. Le nouveau décompte est alors :

- Total\_Item\_Investigations : 6
- Unique\_Item\_Investigations : 4
- Unique\_Title\_Investigations : 3
- Total\_Item\_Requests : 2
- Unique\_Item\_Requests : 2
- Unique\_Title\_Requests : 1

## RECHERCHES FÉDÉRÉES OU ISSUES D'UN ROBOT

L'usage croissant de recherches fédérées et le développement des robots Web sont susceptibles d'influer à la hausse sur les statistiques d'utilisation, c'est pourquoi la norme COUNTER exige que ce type d'accès soit identifié dans les rapports.

Les méthodes les plus courantes pour détecter des recherches fédérées ou automatisées sont les suivantes :

- Un moteur de recherche fédérée peut utiliser une adresse IP spécifique, qu'il suffit de reconnaître pour distinguer les utilisations correspondantes.
- Dans le cas d'une interface HTML standard (p. ex. pour la capture de données d'écran), l'identifiant du navigateur dans les fichiers journaux du serveur web permet d'identifier les activités issues d'une recherche fédérée.
- Pour les activités conformes au protocole Z39.50 (décrit à l'adresse [http://www.niso.org/standards/resources/Z39.50\\_Resource](http://www.niso.org/standards/resources/Z39.50_Resource)), l'authentification se fait habituellement au moyen d'une combinaison nom d'utilisateur-mot de passe. Il suffit alors de créer un couple (nom d'utilisateur, mot de passe) unique réservé au moteur de recherche fédérée.
- Si une passerelle API (*Application Programming Interface*) est disponible, configurez une instance de cette passerelle réservée aux outils de recherche fédérée ; il est aussi recommandé de veiller à ce que la recherche fédérée inclue un identifiant dans ses requêtes adressées à la passerelle.

Une liste COUNTER des outils de recherche fédérée et des robots Web figure dans les annexes du Guide complet de mise en œuvre ; elle est examinée et mise à jour régulièrement.

Dans le cas d'une utilisation fédérée ou automatisée issue d'un utilisateur, et donc légitime, correspondant à une fouille de texte et de données (en anglais, *Text & Data Mining* ou TDM), l'attribut *Access\_Method* dans les rapports doit contenir « TDM » (au lieu de la valeur habituelle « regular ») pour indiquer cette particularité.

## OUTILS DE TÉLÉCHARGEMENT EN MASSE

Les articles en texte intégral téléchargés via un outil automatique ou semi-automatique de téléchargement en masse comme *Quosa* ou *Pubget* ne doivent être décomptés que lorsque l'utilisateur a cliqué sur l'article téléchargé en texte intégral pour l'ouvrir.

# MODIFICATIONS DES DONNÉES

---

## **DÉCLARATION A POSTERIORI DES ERREURS DANS LES DONNÉES D'UTILISATION**

Si des erreurs sont découvertes dans les rapports COUNTER, le fournisseur doit les corriger dans les trois mois suivant leur détection et informer ses clients des corrections.

## **PRISE EN COMPTE DU CHANGEMENT DE TITRE D'UNE REVUE**

En cas de changement de titre d'une revue sans changement de son code d'identification (ISSN ou DOI), la revue doit être considérée comme un titre unique au regard des rapports COUNTER, qui ne doivent mentionner que le nouveau titre, l'ancien étant donc à éliminer.

En cas d'attribution d'un nouvel identifiant DOI ou ISSN, la revue doit être considérée comme équivalente à deux titres, avec deux jeux de données différents. Un rapport pour une certaine période ne peut pas contenir de données pour les deux identifiants à la fois, ancien et nouveau. Tout rapport généré spécifiquement pour l'ancien identifiant fera état d'une utilisation nulle à partir du mois où le nouvel identifiant entrera en vigueur.

# MISE À DISPOSITION DES RAPPORTS COUNTER

---

Les rapports COUNTER doivent être mis à disposition dans deux formats. Le format principal est JSON, destiné à une transmission via le protocole SUSHI ; le format secondaire est celui de fichiers de valeurs séparées par un délimiteur (comme ceux produits par un tableur).

La mise à disposition des rapports COUNTER doit respecter les principes suivants :

- La livraison est mensuelle.
- Elle doit intervenir dans les quatre semaines suivant la fin de la période couverte par les rapports.
- L'historique des rapports accessible doit inclure au minimum l'année en cours plus les 24 mois qui précèdent, sauf en cas de certification récente (p. ex., au 10 juillet 2018, l'historique disponible doit couvrir 2016, 2017 et la période de janvier à juin 2018).
- Il doit être possible, par une requête, d'obtenir les données d'utilisation concernant une certaine plage délimitée par deux mois à l'intérieur des 24 derniers mois.
- Si aucune requête n'est précisée, le rapport doit montrer par défaut l'intégralité des 24 derniers mois.
- Les rapports doivent être immédiatement accessibles sur un site Web via un accès contrôlé par mot de passe.
- Une option doit exister pour recevoir une alerte par courriel dès qu'un nouveau rapport est disponible.
- Chaque rapport doit appartenir à un fichier séparé ou une page distincte pour éviter la manipulation de fichiers trop lourds.
- Les statistiques des rapports COUNTER doivent être indépendantes du navigateur Web utilisé et prendre en charge les versions courantes de Google Chrome, Internet Explorer et Mozilla Firefox, conformément aux standards du *World Wide Web Consortium* (WC3).

Les éditeurs fournissent les rapports COUNTER à chaque client sur la base de son identifiant client (*Customer ID*). Par exemple, si une école possède un identifiant client différent de celui de son université mère, chacun des deux établissements recevra son propre jeu de rapports COUNTER. Généralement, l'authentification se fait par reconnaissance de l'adresse IP. Ainsi, dans l'exemple précédent, si l'école n'a pas une plage réservée d'adresses IP, il n'est pas possible de différencier l'utilisation émanant de l'école de celle issue de l'université, ce qui implique que seule cette dernière recevra un jeu de rapports.

## CAS PRATIQUE

L'université X et son école rattachée souscrivent chacune à différents abonnements sur la plate-forme éditeur Alpha. Elles partagent la même plage d'adresses IP, ce qui les rend impossibles à différencier aux yeux de la plate-forme : les rapports COUNTER destinés à l'université X incluent aussi l'utilisation émanant de l'école.

L'université Y et son école rattachée souscrivent aussi à différents abonnements sur la plate-forme éditeur Alpha. Elles utilisent des plages d'adresses IP distinctes, ce qui permet de les différencier : les rapports COUNTER destinés à l'université X n'incluent pas l'utilisation émanant de l'école.

## SUSHI

Le protocole SUSHI (ou *Standardized Usage Statistics Harvesting Initiative*) vise à simplifier la collecte des statistiques d'utilisation par les bibliothèques ; la version 5 de COUNTER exige la conformité à ce protocole.

Quatre chemins d'API (*API paths*) doivent être pris en charge :

- *GET/status* : renvoie l'état courant du service API COUNTER\_SUSHI
- *GET/reports* : renvoie une liste des rapports pris en charge par le service API COUNTER\_SUSHI
- *GET/reports/{ReportID}* : renvoie un rapport spécifique, par exemple *GET/reports/TR\_B1* pour le rapport de requêtes sur les ouvrages (hors *OA\_Gold*)
- *GET/members* : renvoie la liste des membres du consortium ou des site (clients multi-sites).

Les spécifications COUNTER relatives au protocole SUSHI sont décrites plus en détail sur le site de COUNTER à l'adresse [www.projectcounter.org/COUNTER\\_SUSHI5\\_0.html](http://www.projectcounter.org/COUNTER_SUSHI5_0.html).

## FICHIERS DE DONNÉES SÉPARÉES PAR UN DÉLIMITEUR

Les rapports spécifiés dans la version 5 de COUNTER peuvent tous être fournis sous forme de fichiers de données séparées par un délimiteur :

- Données séparées par une virgule, ou .csv<sup>1</sup>
- Données séparées par une tabulation, ou .tsv

Ces fichiers peuvent être ouverts et lus avec tous les tableurs, dont *Excel*, *OpenOffice Calc*, *Google Sheets*, et *Numbers* pour *Mac*. Même si le formatage du texte (police et couleur) est sans importance dans ce type de format, le respect de l'agencement des données tel que décrit dans la spécification de chaque rapport COUNTER est nécessaire pour obtenir la certification.

---

<sup>1</sup> Note de la version française : bien qu'il existe une variante française du format csv [[https://fr.wikipedia.org/wiki/Comma-separated\\_values#Variantes\\_fran%C3%A7aises](https://fr.wikipedia.org/wiki/Comma-separated_values#Variantes_fran%C3%A7aises)], il convient d'utiliser la version anglaise.

# AUDIT

---

Un point important du Code de bonnes pratiques COUNTER est la nécessité de passer puis de réitérer régulièrement un audit indépendant pour obtenir et conserver la certification. COUNTER s'applique à faire en sorte que cet audit réponde aux besoins des bibliothèques en matière de statistiques d'utilisation fiables tout en restant financièrement abordable. C'est pourquoi les audits sont réalisés en ligne, en suivant le processus décrit dans l'annexe E *Auditing standards and procedures* du Code de bonnes pratiques.

Un audit indépendant est nécessaire dans les six mois suivant la première déclaration de conformité à COUNTER, puis, chaque année. COUNTER reconnaît les audits conduits par un comptable agréé (au Royaume-Uni), par un comptable public certifié (aux États-Unis), par leur équivalent dans les autres pays ou par l'un de ses organismes partenaires.

## PROCESSUS D'AUDIT

COUNTER informe par écrit au moins trois mois à l'avance les fournisseurs ayant déclaré leur conformité de la nécessité de se soumettre à un audit.

Le fournisseur dispose d'un mois pour répondre, en indiquant :

- Le planning de l'audit ;
- L'organisme qui conduira l'audit ;
- Toute demande concernant le processus d'audit.

Quel que soit l'auditeur retenu, l'audit doit être conforme aux exigences et aux tests spécifiés dans l'annexe E du Code de bonnes pratiques. Il est conduit en trois phases :

1. Respect du format et de la structure des rapports d'utilisation
2. Intégrité des statistiques d'utilisation des rapports
3. Mise à disposition des rapports d'utilisation.

Si la conclusion de l'audit est favorable, l'auditeur envoie une copie signée du rapport d'audit au bureau de COUNTER ([lorraine.estelle@counterusage.org](mailto:lorraine.estelle@counterusage.org)).

En cas d'échec à l'audit, l'auditeur envoie au bureau de COUNTER un rapport provisoire décrivant les raisons de l'échec. L'auditeur aide alors le fournisseur à apporter les corrections nécessaires dans un délai convenu avec COUNTER.



# RÉSULTATS POSSIBLES DE L'AUDIT

## **SUCCÈS (*Pass*)**

La conclusion de l'audit est favorable. L'auditeur est susceptible d'inclure dans son rapport d'audit des observations visant à améliorer le processus d'application de la norme COUNTER ; elles ne font cependant pas obstacle à la certification.

## **SUCCÈS avec conditions (*Qualified pass*)**

La conclusion de l'audit est favorable, sous réserve de problèmes mineurs, à régler pour garantir la certification COUNTER. Ces problèmes dits « mineurs » ne portent pas sur les valeurs apparaissant dans les rapports : il peut s'agir par exemple de questions de présentation. Ils doivent être résolus dans les trois mois suivants l'audit pour permettre la certification.

## **ÉCHEC (*Fail*)**

Un problème fait obstacle à la certification COUNTER. À compter de la date de notification, le fournisseur dispose d'un mois pour le résoudre et se soumettre à nouveau à l'audit.