**NBDCファンディング事業の成果DB等アクセス情報取得に関するガイドライン**

２０１７.８.１

Ver. 3.0

JSTバイオサイエンスデータベースセンター（NBDC）

# はじめに

　NBDCの統合化推進プログラムや基盤技術開発の成果として、DBやツールなどへのアクセス数は利用状況を把握するための重要な指標の１つです。各研究チームの集計手法、アクセス情報の集計方法を統一することで、実際の利用状況の把握や、成果としてのアピールをより良く行えるものと考えます。

そこで、成果DBのアクセス情報の集計方法について、NBDCで本ガイドラインを作成・提供しています。各課題で運用していただくことで、成果を適切にアピールできるようにし、本事業が専門家以外の方にも評価されることを目指します。なお、各課題の負担をできるだけかけないように、ガイドラインの見直しも含めて配慮して参ります。ご協力のほど宜しくお願い致します。なお、本ガイドラインは<https://github.com/biosciencedbc/accesslog/>からダウンロードできます。

# 基本方針

アクセス集計の基本方針は以下の通りです。

1. アクセスログ集計ツールをAWStats に統一します。
   * バージョンについては、「３. AWStatsの設定上の留意点」の「(５)AWStatsのバージョンについて」を参照ください。
2. 各研究チームは3ヶ月ごと（7月、10月、翌年1月、4月の15日を予定）に集計結果をNBDCへ提出します。提出に関しては都度にJST担当よりメールでのご案内をします。
3. 提供されたアクセス情報は編集の上、3ヵ月ごとにNBDCポータルサイトより一般公開します。
   * 参考URL：<http://biosciencedbc.jp/about-us/usage-of-database>
4. アクセスログ集計対象サイト一覧（https://github.com/biosciencedbc/accesslog/）に含まれる全てのサイトをアクセス数提出ならびに一般公開の対象とします。また、技術的な理由でAWStatsでのアクセス集計が困難な場合も、NBDCと協議の上、対象外とすることができます。

（５）研究開発期間終了後もアクセス情報の提供をお願いいたします。アクセス情報は、DBの利用状況を示す重要な指標の一つであり、基盤整備を実施している事業の性質上、その把握が必要と考えています。なお、提供されたアクセス情報は、今後予定している追跡評価等においても利用する予定です。アクセス情報の提供の方法については、個別にチームとご相談させていただきます。

# Ver.２.1からの主な変更点

1. データベース開発者ならびに運用者自身の研究活動において当該サイトを利用する場合で除外すると利用状況が正しく表せない場合には、除外しなくても良いこととしました（参照：３．AWStatsの設定上の留意点（３）集計から除外するIPアドレス(SkipHosts設定)について）。
2. robots.pmを最新版に差し替えました（参照：３．AWStatsの設定上の留意点 (４)ロボット/クローラIPアドレス（robots.pmファイル）について）。
3. 収集対象項目に、ページ数（国別）を追加しました（参照：4. 集計対象と集計方法について　（６）ページ数（国別）（図 4 ページ数（国別））。

# AWStatsの設定上の留意点

集計の際には以下の点に留意して、AWStats\_model.confの設定ならびにAWStatsの集計処理を実行してください。

1. 集計対象サイト

アクセスログ集計対象サイト一覧（https://github.com/biosciencedbc/accesslog/）に含まれる全てのサイトを原則とします。

1. 集計頻度

　月単位（デフォルト）

1. 集計から除外するIPアドレス(SkipHosts設定)について

　原則、開発者ならびに運営者のIPアドレスを除いてください。ただしデータベース開発者ならびに運用者自身の研究活動において当該サイトを利用する場合で除外すると利用状況が正しく表せない場合には、除外しなくても良いこととします。

1. ロボット/クローラIPアドレス（robots.pmファイル）について
   * 原則、AWStatsのデフォルト設定をそのまま利用します（※）。
   * エントリの追加は行いません。
   * 誤検知によって通常のアクセスがロボット/クローラと判定される場合が顕著な場合は、エントリの削除（コメントアウト）を行ってもよいこととします。
   * エントリを削除した場合は、本事業内で集計条件を統一するため、削除したエントリをNBDCまでお知らせください。
   * 一部コメントアウトしたrobots.pmファイルは、github上で共有します。コメントアウト箇所はdiffを取って確認できます。

https://github.com/biosciencedbc/accesslog/blob/master/robots.pm

* + robots.pmファイルはAWStats Ver. 7.6 (2017年7月時点での最新版)に含まれるものを使います。

1. AWStatsのバージョンについて

* （４）で指定したVer. 7.6 (2017年7月時点での最新版)に含まれるrobots.pmファイルを適用できるバージョンであれば構いません。

1. Webサーバでログ取得する際には、NCSA combinedフォーマットで取得してください。Apacheでの設定例は以下のページなどを参照ください。
   * http://httpd.apache.org/docs/current/mod/mod\_log\_config.html

# 集計対象と集計方法について

　NBDCが収集対象としているアクセス集計項目は以下の通りです。

1. 訪問者数（図 1　訪問者、訪問数、ページ数(サイト全体）)
2. 訪問数（図 1　訪問者、訪問数、ページ数(サイト全体）)
3. ページ数（サイト全体）（図 1　訪問者、訪問数、ページ数(サイト全体））
4. ダウンロード数（サイト全体）（図 2　Download数（サイト全体））
5. ロボット/スパイダーの訪問数（図 3 ロボット/スパイダーの訪問数）
6. ページ数（国別）（図 4 ページ数（国別））



図 1　訪問者、訪問数、ページ数(サイト全体）

件数、バイト数は集計対象外です。

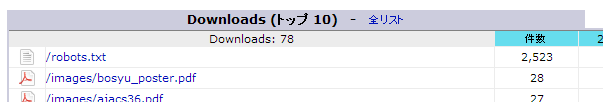


図 2　Download数（サイト全体）

件数の合計がダウンロード数(サイト全体)です。

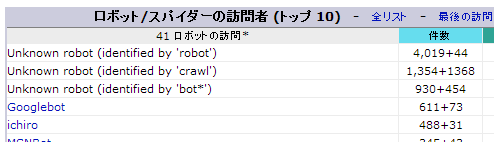


図 3 ロボット/スパイダーの訪問数

件数の合計がロボット/スパイダーの訪問数です。



**図 4 ページ数（国別）**

ページ数の内訳がページ数（国別）です。

# 集計結果の提出方法について

## 提出フォーマット

1. 集計プログラムAWStats.plが出力する集計結果ファイル（参照：図 4）です。
2. １サイト１ヵ月単位で１ファイル存在します。
3. ファイル名のフォーマットは、awstats[mmyyyy].[サイトURL].txt[[1]](#footnote-1)を想定しています。例えば、NBDC広報サイト(http://events.biosciencedbc.jp.txt)の2013年1月の集計ファイル名は、awstats012013.events.biosciencedbc.jp.txtとなります。
4. 設定ファイル: awstats.[サイトURL].confの変数DirDataで指定したディレクトリにファイルが作成されます。

図 4 AWStats集計結果ファイルの先頭部分

AWSTATS DATA FILE 7.1 (build 1.971)

# If you remove this file, all statistics for date 201103 will be lost/reset.

# Last config file used to build this data file was /etc/awstats/awstats.mailonpix.com.conf.

# Position (offset in bytes) in this file for beginning of each section for

# direct I/O access. If you made changes somewhere in this file, you should

# also remove completely the MAP section (AWStats will rewrite it at next

# update).

BEGIN\_MAP 27

POS\_GENERAL 1993

POS\_TIME 2671

＜＜以下、続く＞＞

## 提出方法

NBDCが用意するファイルアップロードサイト(FTPSプロトコル利用)にアップロードします。ただし、NBDCから自動的にファイルを取得できれば、提出者が用意するサーバからダウンロードする方法も可とします。提出期限が近づきましたら、NBDCより連絡いたします。

# 更新履歴

　2017/7 Ver.3.0

1. 集計対象となるサイトを、添付「アクセスログ集計対象サイト一覧.xlsx」を廃止し、基本方針（４）の通りに定めることに変更した。
2. 研究開発期間終了後のアクセスログ集計については、基本方針（５）の通りに定めることに変更した。
3. データベース開発者ならびに運用者自身の研究活動において当該サイトを利用する場合で除外すると利用状況が正しく表せない場合には、アクセス数に含めることを可能にした（参照：３．AWStatsの設定上の留意点（３）集計から除外するIPアドレス(SkipHosts設定)について）。
4. 収集対象項目に、ページ数（国別）を追加（参照：4. 集計対象と集計方法について　（６）ページ数（国別）（図 4 ページ数（国別））

　2015/8/13 Ver.2.1

1. 「６−２. 提出方法」を更新

・NBDCサーバにアップロードする方法以外を追加

・アップロードサイトを変更

　2014/9/19 Ver.2.0

1. 「３. AWStatsの設定上の留意点(６)」の0を追加。
2. 基本方針（３）の機能単位でのアクセス数取得について、運用ならびに利用方針が未定のため、Ver.2.0はURLとそれが指し示す機能の対応表(URL機能説明表.xls)の提出は不要とした。
3. 提出フォーマットをAWstatsが出力するファイル（旧提出フォーマット１（参照：図 6 AWStats集計結果ファイルの先頭部分））に統一。Excelファイルにアクセス数を転記したもの（旧提出フォーマット２）は受け付け不可に変更。詳細は「６. 集計結果の提出方法について」を参照。
4. 提出日は第一金曜日から、該当月の15日に変更。
5. １. 基本方針の(５)ならびに(６)を追加。

* 2013/3/27 Ver.1.1

1. 「３. AWStatsの設定上の留意点」「(４)ロボット/クローラIPアドレス（robots.pmファイル）について」
   * robots.pmファイルのカスタマイズについて追記
2. 「３. AWStatsの設定上の留意点 」「(５)AWStatsのバージョンについて」
   * Ver.7.0以下を利用する場合の注意点について追記

* 2013/3/6 Ver.1.0 初版

以上

1. 2013年度ver.1.1ではAWStats[mmyyyy].[サイトURL].txtとAWSが大文字でしたが、AWStatsのデフォルトに合わせて小文字に変更いたします。ただし、大文字のファイルも受け付けます。 [↑](#footnote-ref-1)