GSJ データベースへのアクセスの集計・解析

吉川敏之[1]・島田幸子[1]・谷島清一[1]

1．はじめに

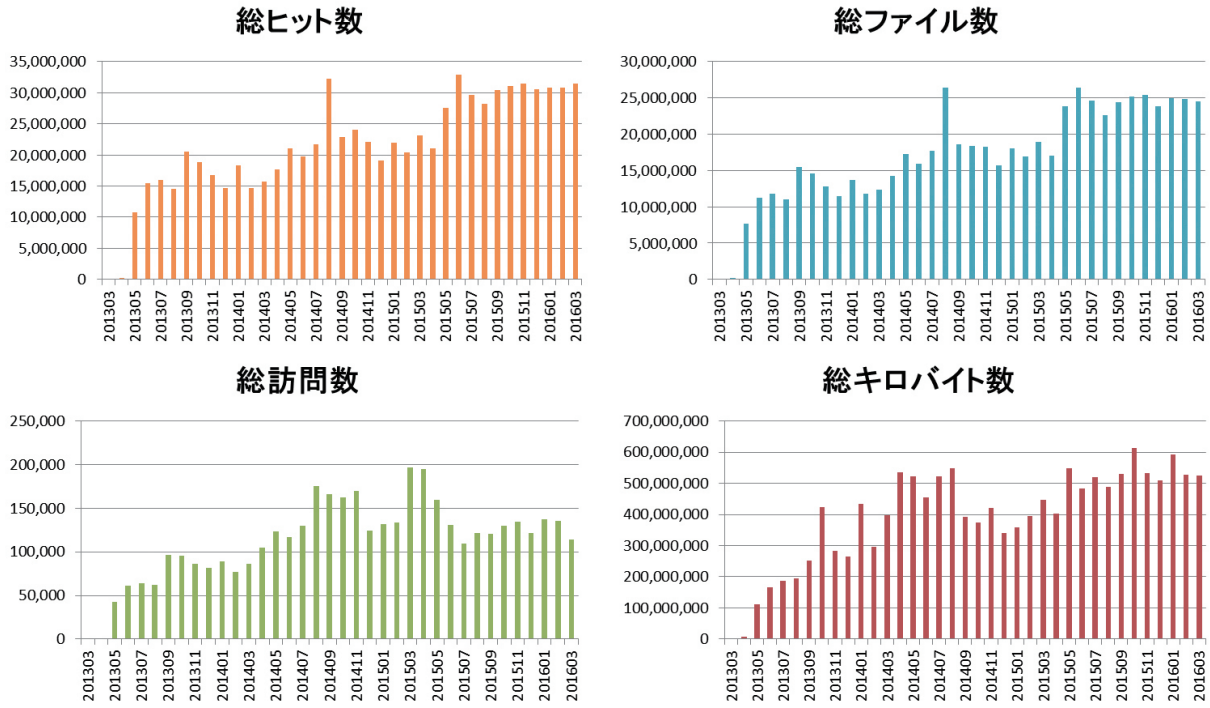
産業技術総合研究所（以下，産総研）では，2015 年度から始まる第四期において社会と研究現場との「橋渡し」機能に注力することを宣言した（産業技術総合研究所，2016）．これは，「我が国最大級の公的研究機関として，日本の産業や社会に役立つ技術の創出とその実用化や，革新的な技術シーズを事業化に繋げる」ことを目的としている．そのためには，研究成果がどのように利用されているかという市場・ユーザー調査が欠かせない．

地質調査総合センター（以下，GSJ）では，研究成果である地質情報を広く社会に使ってもらうために，印刷出版およびウェブサイトからの配信により公開している．このうち印刷物や CD・DVD 等のメディアについては，在庫管理状況から頒布数を算出することができる．ウェブサイト経由の利用については，データの再配布が可能である現在，正確な利用件数はわからないが，サーバへのアクセス件数によりおおよその傾向を把握することは可能である．

地質情報基盤センター（2014 年度までは地質調査情報センター）では，2013 年度から始まった GSJ データベース（以下，gbank）のアクセス状況について，サーバへのアクセスログを取得してきた．また，その情報は可視化ツールである Webalizer を用いて毎月のレポートを作成し，イントラ上で公開・共有してきた．これまでに約 3 年分のデータが蓄積したので，アクセスの経年変化の傾向の概要と，2015 年度の特徴について報告する．

2．アクセスの集計結果

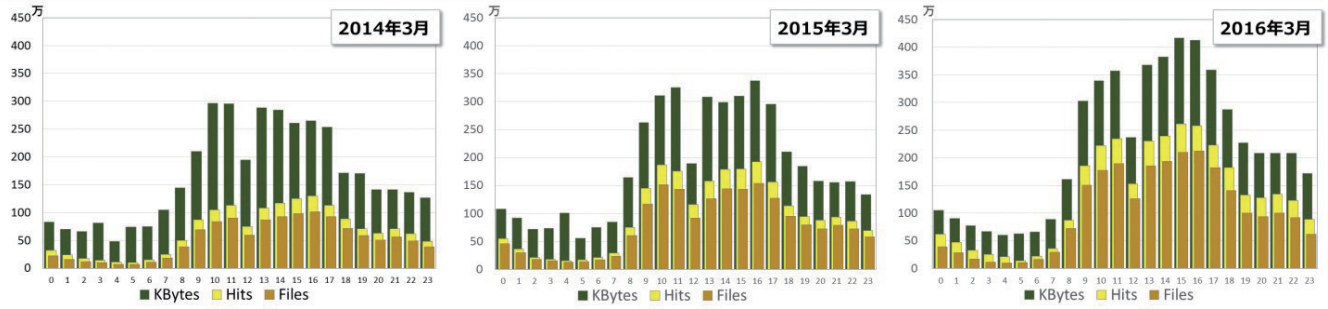
gbank へのアクセスのうち，基本的な集計結果を第 1 図～第 3 図に示す．なお，ヒット，ファイル，キロバイト等の定義と集計の仕方は，付録 1 の説明を参照されたい．



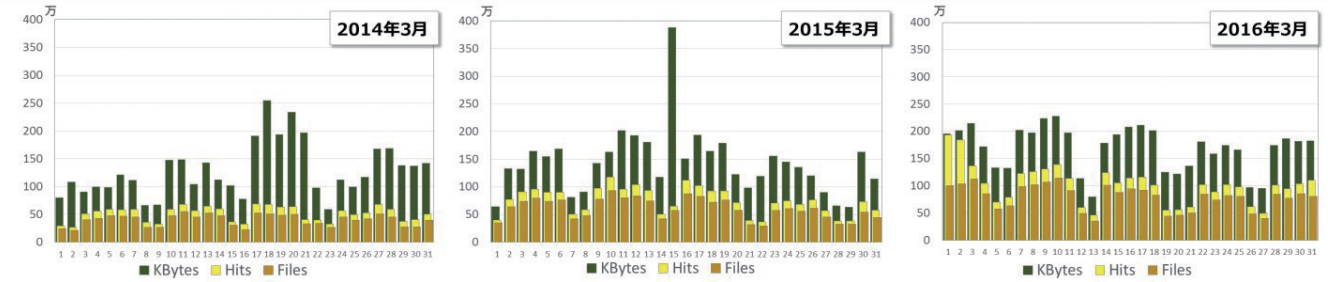
第 1 図　3 年間の月別アクセスの推移
2013 年 3 月から 8 月までは産総研 RIO-DB からの移行期間で，すべてのデータベースが gbank に整備された状況になったのは 2013 年 9 月からである．

1）産総研 地質調査総合センター 地質情報基盤センター

キーワード：アクセスログ，データベース，集計，解析，オープンデータ，webalizer



第 2 図　一日のアクセス数の変化を示したグラフ
左から 2014 年 3 月，2015 年 3 月，2016 年 3 月の結果．なお，スケールを揃えるため，キロバイト のみ 1/10 した数字を使っている（実数はメモリの数字の 10 倍）．



第 3 図　一ヶ月のアクセス数の変化を示したグラフ
左から 2014 年 3 月，2015 年 3 月，2016 年 3 月の結果．なお，スケールを揃えるため，キロバイト のみ 1/10 した数字を使っている（実数はメモリの数字の 10 倍）．

3．データベース毎のアクセスについて

2016 年 3 月末現在，GSJ には 28 のデータベース・システムがある（サーバ上のアカウント単位）．2015 年度を通じたアクセス数上位 10 データベースのアクセスランキング変動結果を第 4 図に示す．なお，上位 10 データベース・システムは，年間を通じて 11 位以下との入れ替わりは発生しなかった．

4．アクセス元について

アクセスログには，どこからアクセスしてきたかを記録できる機能（リファラー）がある．記録されるのは URL のみなので，そこがどのようなサイトなのかを GSJ からアクセスして確認している．ただし，非公開あるいはアクセス制限を行っているようなサイトの場合には GSJ からのアクセスが拒否されるので，残念ながらアクセス元の実態を確かめることができない．

リファラーのデータを基にした 2015 年度の主要アクセス元の集計結果を第 5 図に示す．なお，月によって順位の途中に抜けがあるのは主に上述のような理由で確認不能サイトが存在するためであるが，まれに単一の URL から突発的な高アクセスが発生することがある．また，ドメインを基にしたアクセス元のグラフも第 6 図に示す．

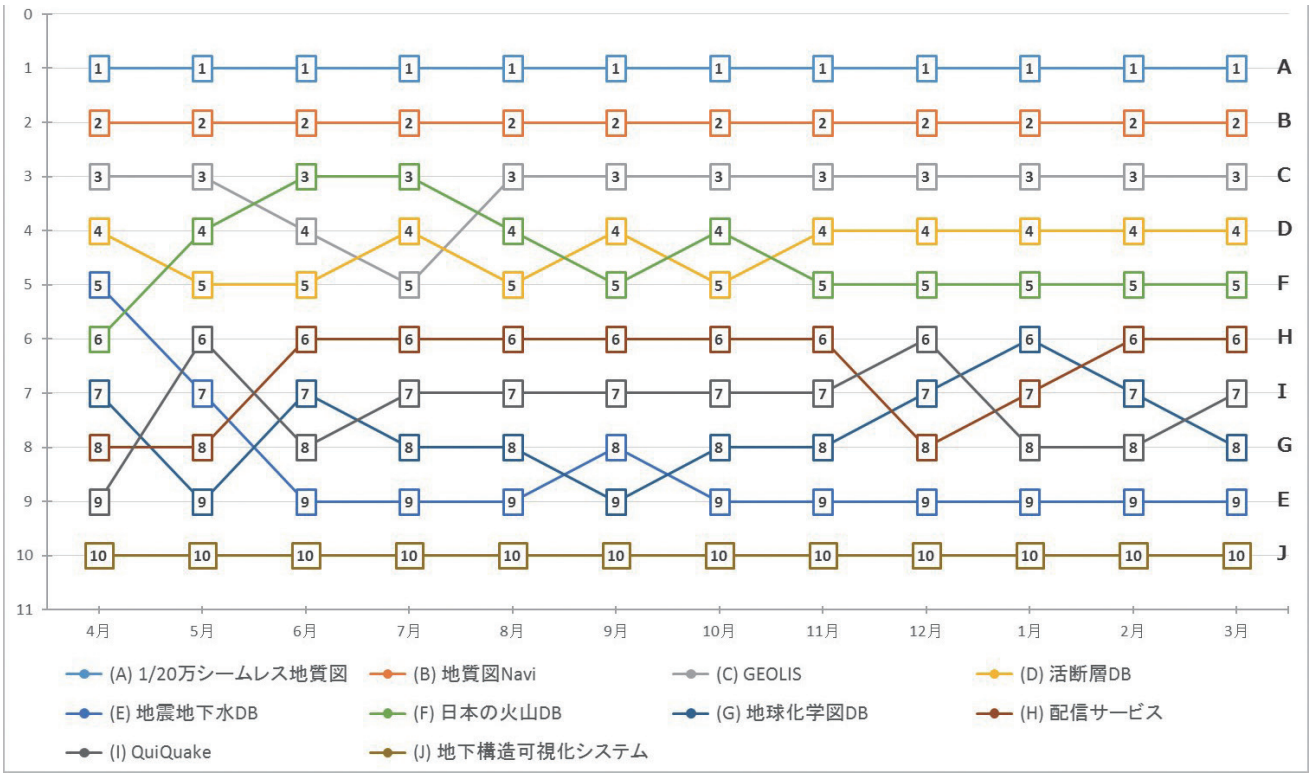
5．考察

全般的なアクセス傾向の特徴は，以下のように考えられる．

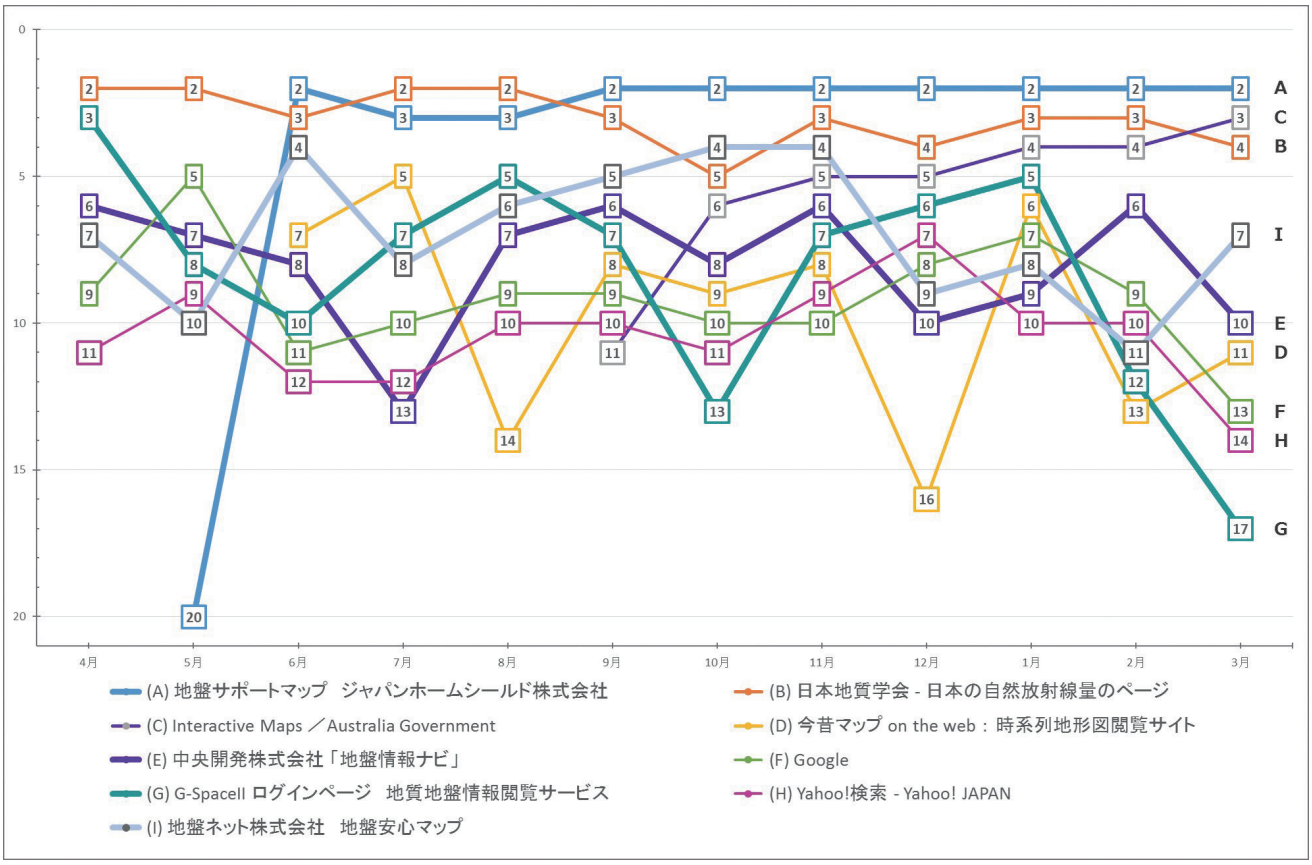
3 年間のアクセス変化の特徴として，gbank 全体のアクセスが増加傾向にあるのは間違いない（第 1 図）．これは，サーバの運用状況からも裏付けられており，近年ほどアクセスの集中によるサービスの高負荷状況が発生しやすくなっている．また，突発的な地質災害や，報道によってアクセスが増えることも確認されており，2014 年 8 月の総ヒット数・総ファイル数の顕著なピークは，広島県で発生した土砂災害に伴い，地質図の情報がニュースに取り上げられたことに起因している．

ただし，総訪問者数のグラフだけは傾向が異なり，直近のアクセス数やピークの現れ方は約 1 年前よりも落ち着いている．この理由については最後に考察する．

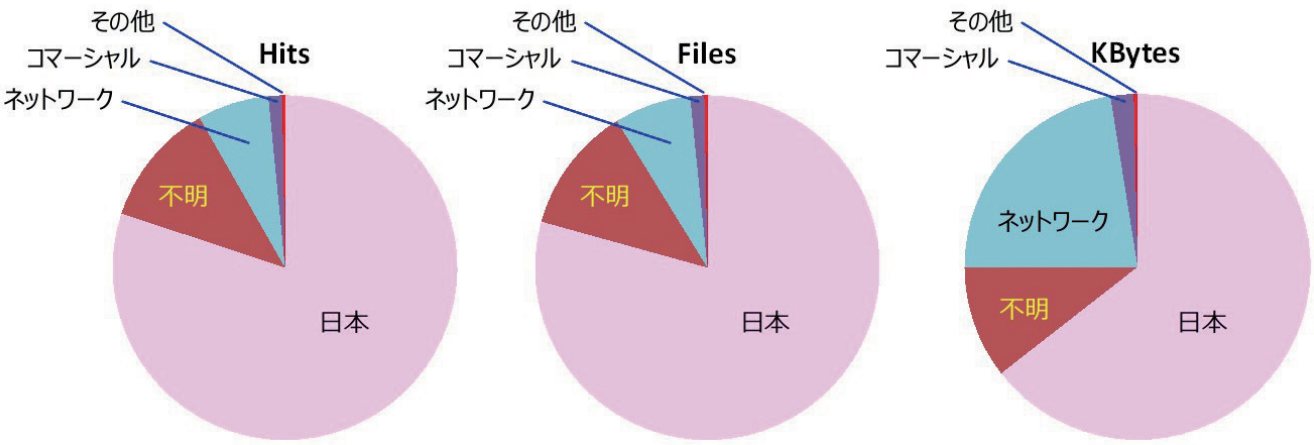
一日のアクセス変化（第 2 図）を見る限り，日本時間の日中のアクセスが卓越し，昼時に一時的に低下する特徴が明瞭である．このことから，gbank の利用者層は日本国内のユーザーが大半であることを示唆する．ドメインを基にした国別アクセス元の結果（第 6 図）からもそれが裏付け



第 4 図　2015 年度の GSJ データベースアクセス数トップ 10 の変動を示したグラフ
総ヒット数（Hits）の数字に基づく．



第 5 図　2015 年度のアクセス元上位サイトの変動を示したグラフ
総ヒット数（Hits）の数字に基づく．なお，1 位は年度を通して Direct Request（お気に入りやブックマーク，URL 直接入力からのアクセス）で変わらない．



第 6 図　2016 年 3 月のアクセス元をドメイン別に表したグラフ.

られる．また，一日のアクセスのグラフ 3 年分（第 2 図）を比較すると，アクセス数の山が年々高くなってきており，こちらの結果からも gbank 全体のアクセス数が増えていることがわかる．

一ヶ月のアクセス数の変化を示したグラフ（第 3 図）からは，平日にアクセス数が多く，土・日曜日および休日に低下する傾向が顕著である．これは主に仕事での利用が卓越していることを示唆する．一日のアクセス変化のグラフでも，19 時以降のアクセスは日中に比べて急減している．ただし，仕事以外の趣味や興味関心を主体とすると考えられる土・日・休日や夜のアクセス数は，過去と比べて現在は明らかに増加している（第 2 図および第 3 図）．

データベースのアクセスランキング変動結果（第 4 図）は，比較的変化が少ない．人気のコンテンツはほぼ固定していることがうかがえる．1 および 2 位の 20 万分の 1 日本シームレス地質図と地質図 Navi は，いずれも地質図等の画像をタイル配信しており，拡大縮小や地域の選択毎に多数のタイルアクセスが発生する構造になっている．また，20 万分の 1 日本シームレス地質図のタイル画像は，地質図 Navi や活断層データベース等の他のデータベース・システムでも利用しており，これらの連携先を利用したユーザーのアクセスも含まれている．

アクセス元の集計結果（第 5 図）では，2015 年度には 2 つの特筆すべき変化が見られた．

1）民間の地盤情報サイトからのアクセス増加

2）オーストラリアからのアクセス増加

1）の民間の地盤情報サイトは，2015 年度末現在，上位に 4 サイトが名前を連ねている．このうち最上位の「地盤サポートマップ（ジャパンホームシールド株式会社）」は 2015 年夏からアクセスが現れるようになった新しいサービスのようで，ランキングに登場して以来，最上位を維持していることから，gbank アクセス数を純増させる一因に挙げられる．他の地盤情報サイトについても，ほぼ安定して上位を維持している．

2）のオーストラリア（Geoscience Australia）からのアクセスは，年度の後半になって増え続けている．Geoscience Australia はオープンデータの先進国であるオーストラリアの地球科学系ポータルサイトで，その地質図ビューアーが，自国の地質図の他にニュージーランドおよび東・東南アジアの地質図 WMS を参照している．後者は gbank から発信されているため，Geoscience Australia の利用者が増えるほど，gbank へのオーストラリアからのアクセスも増える構造になっている．2015 年度後半にアクセスが増え続けている理由は不明だが，Geoscience Australia の機能追加・拡充によるものか，あるいは南沙諸島の領有権問題が国際的に大きく取り上げられるようになったことに起因していると思われる．

単一の URL から突発的な高アクセスが発生する原因としては，何らかの開発テストである可能性が高い．運用初期には所内からのアクセスを集計から排除していなかったため，GSJ データベース関係者の開発状況に応じたランキング変動が頻繁に見られた．現在は GSJ 関係者の開発に伴うアクセスは集計から排除しているが，外部の二次利用サイトがテストのために集中的にアクセスすることは予想できないので，結果として突発的な高アクセスが記録に残ることになる．

総訪問者数が増加していないのにそれ以外のアクセス（総ヒット数・総ファイル数・総キロバイト数）が増加していることは，アクセス元の集計結果から推定された 2 つの変化と関係があると予想している．すなわち，民間の

地盤情報サイトのような二次利用サイトから地質図等の画像タイルを利用するユーザーが増えたため，と考えられる．また，2015 年度は GSJ で緊急調査を行うような地震は発生せず，火山も 5 月の口永良部島火山の噴火以降，比較的静穏であることから，報道等に伴うアクセス急増も発生しにくい状況であった．

上述のように，各データベースの連携が進んでデータの利用（共用）が多様化しているのに加え，外部の二次利用サイトがますます増えていくと，ユーザーの利用実態の把握は次第に難しくならざるを得ない．しかし，一方で入り口はどこであれ，よく利用される，またはユーザーに必要とされる情報・データの特徴を示すという点で，アクセス集計には引き続き注目しておく必要がある．

6．まとめ

gbank の創設以来 3 年間，特に 2015 年度のアクセス集計結果について取りまとめ，その傾向や変化を考察した．主な特徴は以下の通りである．

- 全体として gbank へのアクセス数は順調に増えている．報道をきっかけに，一時的に増えることもある．
- gbank の利用者層は，多くが日本国内のユーザーである．
- 仕事に関係した利用が主体で，趣味や興味関心によるアクセスは少ない．
- 人気のコンテンツはほぼ固定しており，変動は少ない．地質図等の画像タイルは GSJ データベースによる相互利用や外部サイトでの二次利用も進み，安定したアクセスがある．
- 外部サイトのうち，民間の地盤情報サイトからのアクセスが増えている．
- 海外からのアクセスも増えている．

本報告における考察は，あくまでもサーバ管理・運営部署としてのものである．地質情報基盤センターではこれらの結果をサーバ管理の方針・計画に活かしている．一方，個々のデータベース・システムからの立場でアクセス数の変化を検討すると，別の考察も成り立つものと予想される．例えば，新規公開や何らかの改修後のアクセス変化は，そのインパクトをはかる指標ともできる．ユーザーからのダイレクトな反応であるアクセス記録を，研究計画の策定や PDCA の手段として有効に活用していただきたい．

出　典

産業技術総合研究所（2016）「産総研：産総研について」，http://www.aist.go.jp/aist_j/information/index.html（2016.5.1 閲覧）．

付録 1：用語の説明

英語	日本語	説明
Hits	ヒット	サーバに対するすべてのリクエスト．存在しないファイルや，ユーザーのキャッシュに入っていて送信しなかったファイルへのリクエストも含まれる．
Files	ファイル	ユーザーからのリクエストに応えて，実際に送信されたファイルの数．ヒットが入力の数とすれば，ファイルは出力の数になる．
Visits	訪問	サイトを訪問したユーザーの数． 同一 IP アドレスからリクエストがあった場合，30 分以内であればカウントされない．30 分を超えると新規訪問と見なされる．
Kbytes	キロバイト	サイトが送信したデータの総量．ログの中に記録された各ファイルのサイズを合計したもの．サーバはこのデータに基づいて課金される．

付録 2：統計処理の説明

Webalizerのデフォルト表示から，以下のような変更を加えている．

- GoogleBotからのアクセスを計上しないように変更．地質情報利用ユーザーのアクセスとは違うため．
- Localhostからのアクセスを計上しないように変更．地質情報利用ユーザーのアクセスとは違うため．
- JavaScriptへのアクセスを計上しないように変更．内部ページ間の遷移や，ツール利用が主のため．

YOSHIKAWA Toshiyuki, SHIMADA Sachiko and YAJIMA Seiichi（2016）Statistical analysis of access log to the GSJ databases.

（受付：2016 年 5 月 10 日）