The British National Bibliography: Who uses our Linked Data?

Corine Deliot
The British Library,
United Kingdom
corine.deliot@bl.uk

Neil Wilson
The British Library,
United Kingdom
neil.wilson@bl.uk

Luca Costabello
Fujitsu Ireland, Ireland
luca.costabello@ie.fujitsu.com

Pierre-Yves Vandenbussche
Fujitsu Ireland, Ireland
pierre-yves.vandenbussche@ie.fujitsu.com

Abstract

The British Library began publishing a Linked Open Data (LOD) version of the British National Bibliography (BNB) in 2011 as part of its open metadata strategy. Although organisational benefits have been gained, it has been challenging to identify how data is used and by whom. System logs capture basic information and anecdotal usage is received via user feedback, but a lack of analytics tools has made it difficult to gain an understanding of service usage to support sustained investment. This paper describes a project between the British Library and Fujitsu Ireland that examined the insights gained from the development and application of Linked Data analytics. The results suggest such analytics offer LOD publishers many benefits, the most important being an ability to accurately assess service impact and target limited resources more effectively. By doing so publishers can begin to manage LOD services as efficiently as their web counterparts and continue the realisation of Linked Data's potential.

Keywords: British Library; library; Linked Open Data; publication; usage analysis; analytics

1. Introduction

The British Library and Linked Open Data: The British Library is the national library of the United Kingdom with responsibility for distributing metadata describing its collections and recording UK publishing output in the British National Bibliography (BNB)[1]. Originally, these services were aimed at the library community and operated on a commercial basis. However, in 2010 the British Library adopted an open metadata strategy in response to Government calls for improved access to public sector data to promote transparency, economic growth and research. It was also believed that enabling the wider re-use of library data would increase its community value, improve access to information and culture and maintain the relevance of library services.

Simultaneously with the Library's open data initiative was a growing interest in Linked Data's potential for creating new information resources and reaching new users. Such opportunities were felt compelling enough to warrant practical experimentation. Despite a steep technical learning curve for library staff, a Linked Open Data (LOD) representation of the BNB was launched in 2011. The move proved influential among the library community in moving the Linked Data 'debate' from theory to practice (Alemu et al., 2012). Unlike other experimental services, the LOD BNB has continued to evolve with regular monthly updates, the inclusion of new links (e.g. to the International Standard Name Identifier (ISNI)[2]) and content (e.g. serials). The value of the Library's work was recognised by the dataset getting a five star openness rating on Data.gov.uk[3]

[1] http://bnb.data.bl.uk

[2] http://www.isni.org

[3] https://data.gov.uk/dataset/the-linked-open-british-national-bibliography

and being awarded Open Data Institute certification[4]. Such recognition supported the justification for continued work on Linked Data at a time when resources were under considerable pressure from the economic downturn. However, it has always been recognised that continued funding would inevitably depend on hard evidence of significant levels of regular and systematic usage to prove the service met community needs.

Assessing the value of Linked Data services: Despite a continuing interest in Linked Data and numerous related projects, RDF data continues to appeal to a more specialist audience than other, simpler open data formats (e.g. .CSV). The true value of Linked Data services has also been difficult to quantify with limited options available for Linked Data publishers keen to find out how their triples are used and which user groups are attracted to them. Due to the open access approach, evolving technologies and new usage patterns, it can be difficult to accurately gauge service impact. Similarly, attribution may be problematic when services are assembled from multiple sources. The situation can also be complicated by activity-based charging for Linked Data hosting platforms where it is vital to distinguish between innovative forms of legitimate use and abusive activity requiring preventative action.

Access logs have been generated since 2011 and indicate BNB LOD usage can vary significantly from a few hundred thousand to several million transactions per month for no clear reason. The limited availability of LOD analytics tools compared to web equivalents restricted the value of these logs and their interpretation has been a continuing challenge due to the time and effort required to extract useful information. The absence of reliable analytics has made it difficult to clearly identify and prioritise system developments with anecdotal feedback occasionally taking the place of quantitative information on usage patterns or client applications. From a resource management perspective, the requirement to justify all expenditure in a difficult economic environment makes such information increasingly important. Similarly, usage-based charging for hosting coupled with significant variations in usage patterns can make accurate capacity planning problematic. Even determining the impact of raw LOD dumps offered in parallel to the endpoint has proven difficult due to the unwillingness of users to register for access or offer feedback. Interestingly this situation is contrasted by the willingness of over 1500 global users to register for the Library's Z39.50[5] open library data service[6].

A collaborative investigation of Linked Data analytics: A key theme of the Library's open metadata strategy is collaboration to promote experimentation beyond the library domain. The Library is particularly interested in areas where partners can offer rare insights or technical expertise. An offer from Fujitsu Research to collaborate in the exploration of Linked Data analytics was therefore welcomed as an opportunity to both examine BNB LOD usage and to potentially develop tools of interest to the wider LOD community. The results of the collaboration together with their potential for assisting other publishers of Linked Open Data are described below.

2. Publication as Linked Open Data

This section offers background on the Library's LOD publication: its architecture, data model and challenges encountered. Full details can be found in a previous publication (Deliot, 2014).

[4] https://certificates.theodi.org/en/datasets/1063/certificate

[5] Z39.50 refers to ISO 23950 and ANSI/NISO Z39.50. It is an international standard client/server protocol developed by the library community and maintained by the Library of Congress for searching and retrieving records from remote bibliographic databases. http://www.loc.gov/z3950/agency/

[6] http://www.bl.uk/bibliographic/datafree.html#m21z3950

2.1. Challenges

When the BNB LOD project started in 2010 the world of library Linked Data was evolving rapidly with little consensus on many issues, e.g. re-use existing ontologies or create your own? (Hannemann and Kett, 2010). Some challenges arose from a requirement to work with converted legacy data since numerous changes in technology and standards over the 60+ years of BNB's existence necessitated careful normalisation processes. In addition, a transition from the flat data structure of the library domain MARC21[7] (MAchine Readable Cataloging) format to the open RDF entity-based model was problematic as, despite its name, MARC21 was not designed for machine actionability as currently understood. Assigning URIs to bibliographic entities originally represented as text strings involved compromises imposed by the available tools. Inevitably, some challenges also arose from the data modelling decisions made, e.g. imposing formal structure on transcribed text - a more complex process than treating it as a literal. Due to the steep learning curve, it was decided to concentrate on data modelling and conversion activities and use an externally-hosted SPARQL endpoint offered by Talis[8]. This practice continued with the later migration to the TSO OpenUp platform[9].

2.2. Data modelling

When the British Library decided to publish the BNB as Linked Open Data, there was little internal expertise in RDF or domain modelling. The Library therefore used Talis to train and mentor staff and assist development of data models for books[10] and serials[11]. The modelling process stepped back from MARC21 concepts to identify what such records expressed about "things in the world", whether concepts or material objects e.g. bibliographic resources, persons, organisations, etc. The intention was to model a defined part of the bibliographic domain accurately rather than just convert MARC21 to RDF and to focus on the main entities present in the data rather than attempt to replicate MARC21's complex structure and content.

The British Library data model has two main features. Firstly, in order to make the dataset useful beyond the library domain, resources such as books and serials are modelled in accordance with the popular understanding of their meaning rather than more abstract models (e.g. FRBR[12]). Secondly, publication is modelled as an event due to known future requirements to represent forthcoming publications and extend the model to cover further lifecycle events (e.g. acquisition, launch, etc.). To increase interoperability and minimise the overhead of maintaining an extensive British Library ontology, entities and relationships were described using existing RDF vocabularies and ontologies (e.g. Dublin Core, FOAF, etc.). New classes and properties were only defined and documented in the British Library RDF schema[13] where required for the data model. Where possible URIs were assigned to British Library entities following accepted patterns and best practices (Davidson, 2009).

2.3. The 'Extract, Transform, Load' workflow

To generate Linked Data for the service, relevant BNB MARC21 records are selected from the dataset and passed through a series of character set conversion, data normalisation and matching processes prior to the addition of British Library-minted and external URIs. To place the data in a

[7] https://www.loc.gov/marc/bibliographic/

[8] https://talis.com/ Talis closed its generic semantic web division due to insufficient commercial interest in July 2012

[9] http://www.tso.co.uk/our-expertise/technology/openup-platform

[10] http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf

[11] http://www.bl.uk/bibliographic/pdfs/bldatamodelserial.pdf

[12] http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

[13] http://www.bl.uk/schemas/bibliographic/blterms

broader context, links to trusted resources selected from both library and general domains are included. The enhanced file is converted to RDF/XML and N-Triples and quality checked. The resulting data dumps are uploaded to the website[14]. N-Triples files together with VoID[15] descriptions are loaded to the Linked Data platform, where users can access the data via a SPARQL endpoint[16] and content negotiation (dereferencing)[17].

3. Data usage analysis methods

3.1. Technology shortcomings

Bringing access analytics to Linked Data requires an understanding of the different modes of Linked Data publication specificities, i.e. dataset dump, SPARQL endpoint and HTTP dereferencing. While monitoring access to a dataset dump is no different from any other file and can be undertaken with traditional Web analytics tools (Fasel and Zumstein, 2009), these applications do not suffice for the two other publication methods. Google Analytics[18] and other popular web analytics platforms[19] (e.g. Open Web Analytics[20], PIWIK[21]) are not designed for linked datasets (e.g. they do not provide access metrics for SPARQL). Getting insights from SPARQL endpoint access requires the parsing of queries issued and extraction of useful information such as the load of a query or the type of resources requested. Similarly, HTTP dereferencing necessitates the support for HTTP 303 content negotiation[17] not handled by traditional Web analytics tools. In the literature, few initiatives propose Linked Data-specific traffic metrics: Möller et al. (2010) propose a list of Linked Data-specific metrics that cover HTTP and SPARQL access to RDF (e.g. ratio between 303 and 200 HTTP requests, number of RDF-aware agents, SPARQL query features, machine vs human classification based on user-agent strings). The well-established USEWOD workshop series[22] is the reference for Linked Data usage mining and provide a dataset of anonymised linked datasets access logs. We reused and extended metrics defined in (Fasel and Zumstein, 2009; Möller et al., 2010).

3.2. System

To assist analysis of the BNB access logs, Fujitsu Ireland developed a hosted analytics platform for Linked Datasets. An online demo[23] shows one month of traffic insights of The British National Bibliography (BNB) data set. The system mines the logs of registered Linked Data publishers and extracts traffic insights. The analytics system is designed for RDF data stores with or without a SPARQL engine, supports Linked Data HTTP dereferencing with HTTP 303 patterns, load-balancing scenarios and filters out search engines and robot activity. The system offers Linked Data-specific features which are currently not supported by classic web analytics tools (e.g. visitor sessions). Clients are not tracked, thus preserving visitors' privacy. To better identify workload peaks of a SPARQL endpoint, SPARQL queries are qualified as heavy or light according to SPARQL syntactic features.

[14] http://www.bl.uk/bibliographic/download.html

[15] https://www.w3.org/TR/void/

[16] http://bnb.data.bl.uk

[17] https://www.w3.org/TR/cooluris/

[18] http://analytics.google.com

[19] https://en.wikipedia.org/wiki/List_of_web_analytics_software

[20] http://www.openwebanalytics.com

[21] http://piwik.org

[22] http://usewod.org/

[23] http://52.49.205.156/analytics/

The system, illustrated in Figure 1, starts by parsing the BNB access logs from March 2014 to April 2015. Access information from robots and search engine crawlers are filtered out to remove noise from usage insights. The system extracts traffic metrics from the logs. It includes traditional metrics such as location of visitors, referrer website as well as Linked Data-specific ones. A list of the metrics extracted is presented in the following section. Traffic metrics are stored in a data warehouse equipped with an SQL-compliant MOLAP[24] unit that answers queries with sub-second latency. The front-end queries the RESTful APIs exposed by the MOLAP Unit, generates a web UI. Figure 2 shows three different screenshots of the Web UI as used by the British Library Linked Data team to get insight on (a) the most popular RDF classes (including visitors' mistakes); (b) the distribution of heavy and light SPARQL queries; and (c) the visitor location over a given period of time. Data can alternatively be accessed directly from the APIs for analysis by other tools or visualisation interfaces. Although the system has been developed in collaboration with the British Library, it is generic and can be used by any Linked Data publisher, providing they have access to their Linked Data server access logs. Distribution plan and licence for the tool can be provided upon request[25].

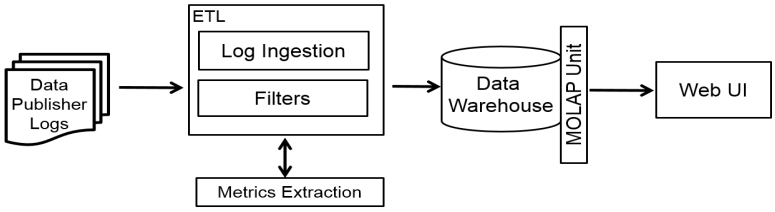
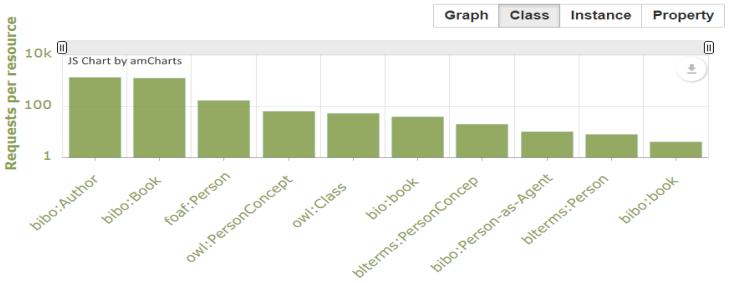
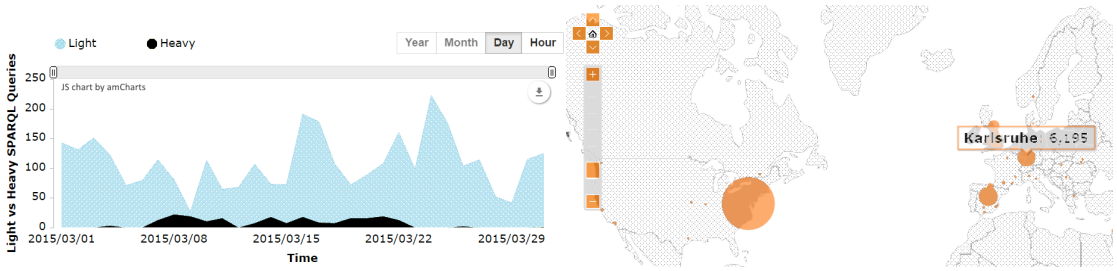


FIG. 1: Architecture of the analytics platform for Linked Data publishers



(a)



(b) (c)

FIG. 2 Screenshots from the web UI (selected timeframe: March 2015): (a) most popular RDF classes (including visitors' mistakes), (b) distribution of heavy and light SPARQL queries, and (c) visitor location.

[24] Multidimensional Online Analytical Processing

[25] http://innovation.ie.fujitsu.com/contact-us/

3.3. Metrics

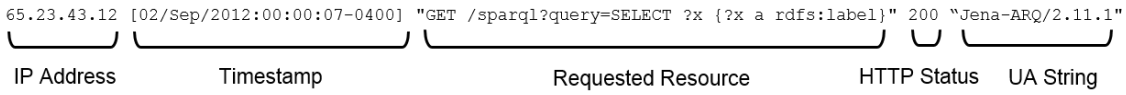


FIG. 3: Linked Dataset access record (Apache Commons Log file Format)

Figure 3 shows an example of access logs processed by the system. A number of key performance indicators for web analytics have been proposed in the academic library domain (Fagan, 2014). The system extracts some of the metrics described by (Fagan, 2014) from access logs, and extends the work of (Möller et al., 2010). Such metrics are grouped in three categories:

Content Metrics. How many times RDF resources have been accessed (cf. Figure 2a). Unlike traditional tools, 303 URIs are correctly interpreted, and the number of times resource URIs appear in SPARQL queries[26] is also counted. Aggregates are provided by family of RDF resource (i.e. instances, classes, properties, graphs).

Protocol Metrics. Information about the data access protocols used by visitors. Includes SPARQL-specific metrics such as the count of malformed queries, SPARQL query type (*SELECT*, *ASK*, *DESCRIBE*, and *CONSTRUCT*) or the detection of light and heavy SPARQL queries (cf. Figure 2b).

Audience Metrics. Besides traditional information about visitors (e.g. location, network provider), these measures include Linked Data-specific metrics such as details of visitor sessions or language tags in queries (cf. Figure 2c).

4. Results

4.1. Overview

While top level statistics for daily and cumulative monthly usage had been regularly recorded (e.g. the BNB dataset dump has been downloaded on average 40 times a month from April 2014 to April 2015), there had been limited detailed examination of log files until the project began. This was due to the volume of data logged, the resources required to extract useful information together with a need to compare data over time to identify meaningful usage patterns. However, the analysis possible via the Fujitsu system offered a range of new insights, which were further assisted by the use of graphical visualisation techniques available on the platform.

4.2. Traffic and usage

As anticipated, the bulk of the search requests (i.e. 43.7M over 13 months) originated from search engine and Linked Data crawlers together with some robot activity. Google variants and other search engines including Bing or Baidu were found to account for 40.7M of these. While significantly smaller as a proportion, the filtered 252K HTTP and SPARQL queries received over the period were found to increase over time from 18K in April 2014 to 24K in April 2015 (see Figure 4) with the number of SPARQL queries making up a significantly increasing proportion (i.e. 67 in April 2014 to 11.1K in April 2015). Over the period of study SPARQL queries were found to be of predominantly light complexity[27] (e.g. 10.8K light vs 364 heavy in April 2015). The average daily duration of sessions[28] was found to be ~1 hour for visitors using software

[26] Access logs do not contain SPARQL result sets. This is therefore a lower bound estimation.

[27] A SPARQL query is defined as heavy (or light) if it requires considerable (or little) computational and memory resources.

[28] A Session is defined as a sequence of requests issued with no significant interruptions by a uniquely identified visitor.

libraries, and 26 minutes for visitors from desktop browsers. Software libraries' sessions also include on average 24 requests with 11 distinct RDF entities queried, while browsers sessions only account for (on average) 2 requests with 2 distinct resources queried. This suggests that although much usage relates to brief investigative, tutorial or test activity (e.g. from desktop browsers), a smaller number of more expert users are undertaking systematic and structured queries of the site over longer periods (e.g. with scripts and SPARQL). This usage breakdown is also evidenced by a 2:1 ratio of new versus returning visitors, indicating the dataset continually receives new users. However, 48% of such sessions consisted of single resource lookups (bounce rate), indicating the possible need to prioritise retention methods, e.g. suggesting "related links" in the HTML view of each RDF resource.

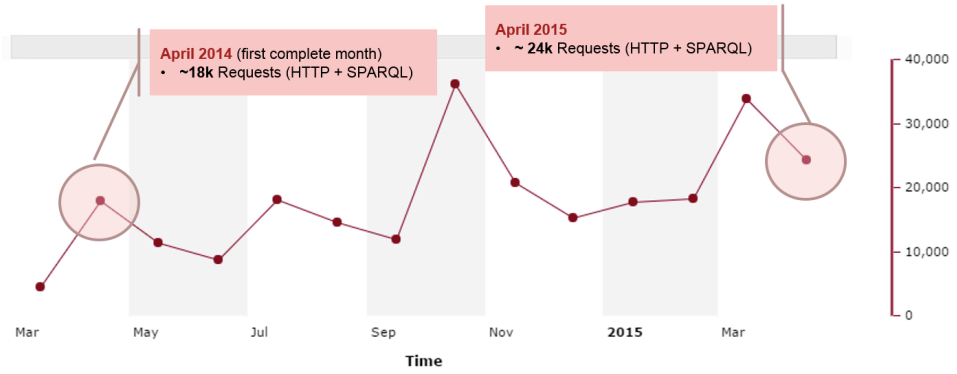


FIG. 4: Evolution of the number of requests from March 2014 to April 2015.

There seemed to be an identifiable correlation between usage peaks and the addition of new metadata elements (e.g. ISNIs[2] in early 2015) or full data refreshes (e.g. September 2014). Some usage patterns could be related to Linked Data tuition sites or events, e.g. a peak in August 2014 following reporting of an international library Linked Data conference in Paris[29] and publication of a Linked Data survey mentioning the BNB[30]. Experimental or educational usage was further underlined by the presence of queries documented in tutorial material published concerning the site, e.g. 1.4K requests for the author C.S. Lewis based on a tutorial query present on the BNB site[31] and 6K requests for 'The Hobbit' relating to a blog tutorial[32]. It was also instructive to quantify previously anecdotal usage by educational institutions. Table 1 presents the ten most queried classes URIs. The fact that several classes queried do not exist in the dataset, e.g. they are misspelled (*bio:birth*), or do not refer to actual terms in existing ontologies (e.g. *owl:PersonConcept*) also reinforces the conclusion that the site is regularly searched by novice users.

TABLE 1: Top 10 classes queried from March 2014 to April 2015

Class URI	Prefixed URI	Present in BNB dataset	Frequency
http://purl.org/dc/terms/BibliographicResource	dcterms:BibliographicResource	Yes	2,115
http://purl.org/ontology/bibo/Author	bibo:Author	**No**	1,429
http://purl.org/ontology/bibo/Book	bibo:Book	Yes	1,307
http://purl.org/vocab/bio/0.1/birth	bio:birth	**No**	591
http://bnb.data.bl.uk/resource/Author	blterms:Author	**No**	476

[29] http://commonplace.net/2014/08/library-linked-data-happening/

[30] http://hangingtogether.org/?p=4137

[31] http://bnb.data.bl.uk/getting-started

[32] https://blog.ldodds.com/2014/10/08/an-introduction-to-the-british-national-bibliography/

http://xmlns.com/foaf/0.1/Person	foaf:Person	Yes	169
http://www.bl.uk/recourse/Author	blterms:Author	**No**	112
http://www.w3.org/2002/07/owl#PersonConcept	owl:PersonConcept	**No**	65
http://www.w3.org/2002/07/owl#Class	owl:Class	Yes	57
http://purl.org/ontology/bio/book	bio:book	**No**	38

The original publication of the BNB Linked Open Dataset generated wide interest amongst global library and open data communities. It was therefore interesting to examine usage from a geographical perspective. The country with the largest source of queries (33.3%) was the United States with the UK next at 21.7% and Germany third at 9.9%. In the library domain, a significant number of requests originated from other state libraries, suggesting a shared exploration of the use of Linked Data at the national library level. The dataset was also found to be used by 350 UK and foreign academic and governmental organisations. The analytics tool also allowed some abnormal activities to be discovered, e.g. an unknown 1-hour spike of 10,000 light SELECT SPARQL queries (October 28th, 2014). The tool identified the requests as originating from a specific city[33], and showed the queries were issued by a Java application. Using this information, staff quickly found thousands of identical queries in access logs and concluded they were probably due to a bug in the client rather than malevolent action.

Queries originating from humans rather than machines accounted for 62% of access with desktop browsers being the most popular method (54%). However, a significant increase (95x) in usage by software libraries was identifiable from 83 requests in April 2014 to 7,895 in March 2015. Overall, there is a clear evolution of the type of visits to access the BNB SPARQL endpoint. This migrated from a dominant profile of manual, human browsing of HTML pages generated from the data (issuing HTTP dereferencing) to a majority of access by machines using software libraries (cf. Figure 5). As of April 2015, 65 distinct SPARQL-based client applications were observed, showing a steady growth from the beginning of the study. This suggests that from an initial experimental base, an ecosystem of more mature clients may be developing around the dataset.

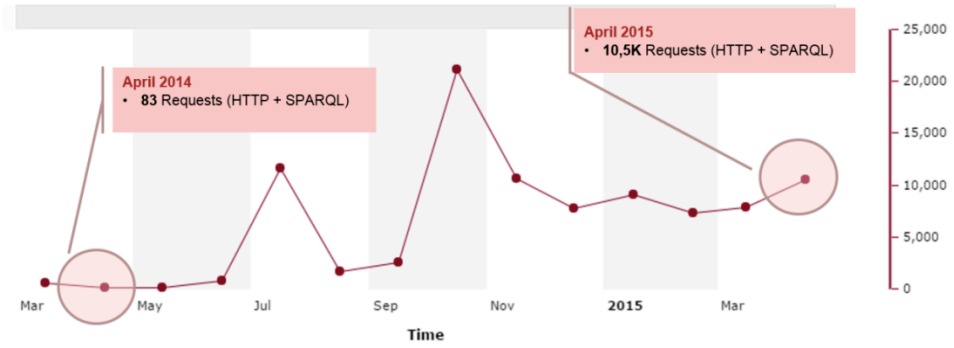


FIG. 5: Evolution of the number of requests issued by Software Libraries.

Client-side HTTP errors account for almost 9% of requests, with 4% of overall requests being '404 Not found' errors (misses). Server-side errors account for 1.5% of total resources (e.g. internal triple store-related errors).

5. Discussion

The role of analytics in assessing Linked Data service value: Creation of comprehensive Linked Data analytics offers Linked Data publishers opportunities to gain new insights into who uses their service and for what purpose and to make development decisions based on concrete

[33] For privacy reasons, we do not provide any specifics about any user.

data. Understanding of BNB LOD usage has significantly improved through the project and its findings will assist future planning. Initially the results have informed development discussions with the current platform provider and will be used to support a case for service continuity based on the identified educational value, the platform's utility for exposing deep library metadata to search engines and indications of a developing Linked Data ecosystem around the service. The results will also be used to guide development and promotion of future Linked Data services and the investigation of a tiered access model to improve performance and prevent misuse. Findings will also assist resource balancing for user documentation activities (e.g. SPARQL/non-SPARQL). Most significantly, the results will inform development of a tender specification for the Library's next Linked Data platform due to go live in 2017.

The wider advantages to publishers offered by Linked Data analytics can be categorised as: organisational, technical, service management, and user support benefits. Some areas, including bounce rate, tracking system performance over time, insights into visitor behaviour and location are common to web analytics; while others have similarities but differing implications. An example of the latter is mobile device usage since this can indicate the impact of a social media campaign or public event via link publication rather than a new requirement for a mobile optimised site.

Organisational benefits: From an organisational perspective, Linked Data analytics offer several important benefits. These include the ability to target scarce resources (staff effort, financial, technical, etc.) more effectively while improving institutional reputation as a trusted creator of LOD services optimised for 'real world' user requirements. The insights gained can also inform an understanding of the relative position of Linked Data platforms in wider institutional systems and resource discovery strategies. Analytics used together with regulated API keys could also assist identification and management of trusted high volume users and potentially lead to new collaboration opportunities.

Technical and service management benefits: Some of the more obvious benefits of Linked Data analytics relate to service management or technical aspects of LOD publication. These include the ability to build a comprehensive understanding of usage categories and geographic spread (e.g. search engine, developer, individual) to support service investment. Such information can support accurate cost control and tendering for service hosting options by ensuring only the appropriate systems capacity is specified. Similarly, such fundamental information enables service providers to determine dataset or feature popularity in order to support accurate decision-making on service extension, enhancement or withdrawal. Specific system benefits from the application of analytics include identification of the range and intensity of normal, legitimate usage together with abuse patterns and abusers to support availability. Normally, publishers have no choice but to manually browse data stored in server access logs. However, an efficient analytics system can extract traffic metrics of Linked Datasets and present results via a web interface to relieve publishers of time-consuming log mining. The ability to interpret access patterns and peak usage can also support service performance optimisation via caching of frequently used data and other tuning techniques. Examination of search engine traffic can also support optimisation of hosted data via targeted monitoring of harvest patterns coupled with structured interrogation of search engine sites to assess the results of changes.

User support benefits: From a support perspective, Linked Data analytics offer staff the ability to identify documentation enhancements (e.g. sample SPARQL queries), relevant tutorial examples and improve evidence-based communications on support issues. An aspect of behaviour common to both web and Linked Data sites is the extremely low percentage of users willing to report problems but to simply switch to alternatives instead. The relatively new nature of LOD services and variations in standards compliance (e.g. SPARQL 1.0/1.1) means user expectations and behaviour can vary significantly beyond those of web equivalents. Regular analytics offer the ability for LOD publishers to spot and fix emerging issues to improve user retention and regular usage while also suggesting development or documentation needs.

All of the above ultimately enable LOD publishers to maintain and improve service continuity and performance for the benefit of users. Concrete analytics data also supports better services with targeted characteristics based on observed usage patterns and developments arising from demonstrable user needs.

6. Conclusions

The British Library believes Linked Open Data to be a logical evolutionary step for the established principle of freedom of access to information, offering trusted and authoritative knowledge organisations an important role in the new information landscape. For such organisations, the vision of a global pool of semantically rich, reusable metadata enabling them to concentrate scarce resources on adding unique value is highly attractive. Similarly, the potential value of LOD sites in offering cost-effective exposure of large datasets to search engines, application developers and new modes of resource discovery has great appeal. However, tough economic conditions and the rapid evolution of LOD solutions necessitate hard evidence-based justification for any new expenditure. The Linked data analytics developed by Fujitsu in this project offers publishers the ability to accurately assess the impact of their data and target scarce resources more effectively. In doing so they can begin to develop and manage new LOD services as efficiently as more traditional web services and continue the realisation of Linked Data's potential for the benefit of the wider community.

Acknowledgements

This work has been supported by the 'TOMOE' project funded by Fujitsu Laboratories Limited.

References

Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. New Library World, 113(11/12), (pp. 549-570).

Davidson, Paul. (2009). Designing URI Sets for the UK Public Sector. UK Chief Technology Officer Council. Retrieved, May 25, 2016 from https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector

Deliot, Corine. (2014). Publishing the British National Bibliography as Linked Open Data. Catalogue & Index (174), (pp. 13-18).

Fagan, Jody Condit. (2014). The suitability of web analytics key performance indicators in the academic library environment. The Journal of Academic Librarianship, 40(1), (pp.25-34).

Fasel, Daniel and Darius Zumstein. (2009). A fuzzy data warehouse approach for web analytics. In Miltiadis D. Lytras et al. (Eds.): WSKS 2009, LNAI 5736, (pp.276-285). Berlin, Heidelberg : Springer.

Hannemann, Jan, and Jürgen Kett. (2010). Linked data for libraries. In Proceedings of the World Library and Information Congress: 76th IFLA General Conference and Assembly, 2010.

Möller, Knud, Michael Hausenblas, Richard Cyganiak, Gunnar Grimnes and Siegfried Handschuh. (2010). Learning from linked open data usage: Patterns & metrics. In WebSci10: Extending the Frontiers of Society On-Line, (pp.1-8)