Project

Project Acronym:	AthenaPlus
Grant Agreement number:	325098
Project Title:	Access to cultural heritage networks for Europeana

Deliverable

Deliverable name:	**Review on Linked Open Data Sources**
Deliverable number:	**D4.2**
Delivery date:	October 2013
Dissemination level:	Public
Status	V1.0
Authors (organisation)	Werner Köhler (UNIMAR) Regine Stein (UNIMAR)
Contributors (organisation)	All partners
Reviewers (organisation)	Gordon McKenna (CT) Nikolaos Simou (NTUA)

Revision History

Revision	Date	Author	Organisation	Description
	2013-07	Regine Stein Nasos Drosopoulos Nikolaos Simou	UNIMAR NTUA NTUA	Outline and scope of the deliverable
	2013-08		All partners	Survey
V0.1	2013-08-16	Werner Köhler	UNIMAR	First draft version
V0.2	2013-09-30	Regine Stein	UNIMAR	Second draft version
V0.3	2013-10-04	Gordon McKenna Nikolaos Simou	CT NTUA	Review: suggestions and approval
V1.0	2013-10-07	Werner Köhler Regine Stein	UNIMAR UNIMAR	Final version
		Maria Teresa Natale	ICCU	Formal check

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Project Coordinator: Istituto centrale per il catalogo unico delle biblioteche italiane
Address : Viale Castro Pretorio 105 – 00185 Roma
Phone number : +3906 06 49210 425
E-mail : info@athenaplus.eu
Project WEB site address : http://www.athenaplus.eu

Table of Contents

1   EXECUTIVE SUMMARY

The objective of this deliverable is to provide an overview of the Linked Open Data (LOD) sources that could be used for linking the AthenaPlus content.

The report is composed of five parts: It starts with a brief introduction to Linked (Open) Data with focus on some technical and formal aspects (chapter 2). The following part presents the results of a survey among the AthenaPlus partners on Linked Open Data (chapter 3). Then selected approaches to find LOD sources and gather descriptive information about them as requested for the deliverable are described (chapter 4). Subsequently the core list of LOD sources identified as candidates is presented. This includes general and technical details for the data sources examined, including the type of content described, the amount of resources served and the protocols supported for consuming data as far as they were determinable (chapter 5), and is followed by some conclusions (chapter 6). The results of this deliverable will directly inform further work in task 4.3 and the related deliverable D4.6.

2 INTRODUCTION

2.1 Background

The concepts of Linked Data (LD) and Linked Open Data (LOD) have been widely discussed and presented in the cultural heritage community and therefore are explained very briefly and with a focus on the technical aspects and requirements. For more comprehensive reading one may consult the online material and books recommended in appendix 1 References. [1]

2.1.1 Linked Data in the Semantic Web

LD is a method to connect different data sources with each other using web technologies. The concept of LD is embedded in and is part of the strategy for the development of the Semantic Web. The following illustration of the *Semantic Web Stack* shows how various standard technologies on different layers are combined to develop the Semantic Web. The illustration was introduced by Tim Berners-Lee[2] in a more simple form 2000 and then was altered as the layers became more concrete. The structure of the Semantic Web Stack is still in progress.
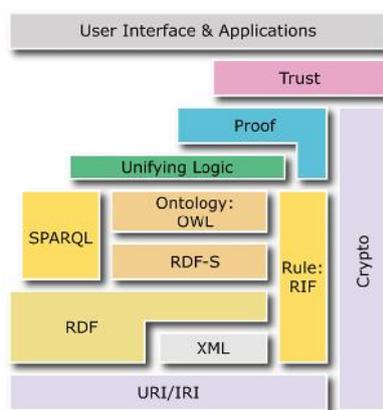
Figure 1: The Semantic Web Stack January 2007[3][4]

Actually the left part of the illustration between the layers "URI/IRI" and "Unifying Logic" with the boxes "SPARQL", "Ontology: OWL", "RDF-S", "RDF" and "XML" is most relevant for the concept of Linked Data. From a technical point of view Linked Data simply can be seen as a combination of URIs, RDF and SPARQL. For a view more focused on context and semantics the other aspects like schemes (RDF-S), ontologies (OWL) and rules (RIF) become more relevant.

For linking resources found in two different data sources the same identifiers need to be used on both sides. Usually identifiers already exist in any source in form of data keys. While one can match such keys directly between two data sources in order to connect them, a more sustainable way is to mint persistent URIs for the actual resources in a dataset and match them or even reuse from the outset published identifiers which may be provided in widely used authority files, controlled vocabularies, ontologies. Hence links to the resources of third parties that also use or have matched their resources to the same authority files or ontologies are automatically established, thereby bypassing the need for extra linking processes.

2.1.2 Linked Open Data

The "O(pen)" in LOD means that data should be provided with open licences when publishing on the web so that it can be used by others without barriers.

Tim Berners-Lee proposed four rules for Linked Data:[5]

---

[1]   [all hyperlinks and query results in this document retrieved between August 27 and September 29, 2013]

[2]   http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

[3]   http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/layerCake-4.png

[4]   http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#%2824%29

[5]   Berners-Lee, Tim. (2006–2009). *Design Issues: Liked Linked Data*. World Wide Web Consortium (W3C). http://www.w3.org/DesignIssues/LinkedData.html

1.	Use URIs as names for things;
2.	Use HTTP URIs so that people can look up those names;
3.	When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL);
4.	Include links to other URIs, so that they can discover more things.

In 2010, he added a five star rating system as a "road to good linked data"[6].

★	Available on the web (whatever format) *but with an open licence, to be Open Data*
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

In addition the W3C Government Linked Data Working Group for Open Government Linked Data (OGLD) has given recommendations for publishing LD with the "Linked Data Cookbook" 2011[7]. These recommendations can be generalized and are transferable into the domain of publishing cultural heritage content.

The rules, the rating system, and the cookbook recommendations have informed the approach for selecting LOD sources in this deliverable.

2.2 Role of this Deliverable in the Project

The objective of this deliverable is to provide an overview of the Linked Open Data sources that could be used for linking the AthenaPlus content. It is part of WP4 "Terminologies and semantic enrichment" and especially of Task 4.3 "Semantic enrichment of cultural metadata with Linked Open Data". The main objective of Task 4.3 will be the creation of a semantic repository for the AthenaPlus project that will include all the AthenaPlus metadata in a semantic rich representation giving special effort to the linking process. To achieve this objective, it is necessary to identify possible sources that fit to the AthenaPlus content, have a sufficient amount of resources, and support technical protocols suitable for linking processes.

The focus of this deliverable is on the formal analysis of possible sources. Content analyses necessary for the actual linking processes are not an objective of this deliverable, it rather creates the basis for deliverable D4.6 "Linking of metadata to external data sources". D4.6 will build on D4.2 and present the external data sources to which the cultural content of the AthenaPlus will eventually be linked, reporting on the analysis and highlighting examples for using underlying terminologies to improve linking processes.

---

[6]   ibid.
[7]   http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

3 ATHENAPLUS SURVEY ON LINKED OPEN DATA

UNIMAR decided that it would be informative to survey the AthenaPlus partners (and associated providers). This was undertaken to find out:
- Information about the respondents to the survey (organization type, and country);
- Their knowledge and use of linked open data, and linked open data initiatives;
- Their approaches how to get information about linked open data.

3.1 How the Survey was carried out

The survey had 19 questions and was in three parts:
1. *Partner Information*;
2. *Using Linked Open Data*;
3. *Publishing Linked Open Data.*

For full documentation of the survey questions see appendix 3.

The survey was disseminated via E-Mail with attachments as Word-Document (File extension "doc") and as Open Office document (file extension "odt") to the AthenaPlus partner list, requesting partners to fill in one of these forms.

A total of 28 responses were received from the AthenaPlus partners and associated providers.

3.2 Partner Information

3.2.1 Respondent Type

Following types of organisations responded to the survey (multiple answers allowed):

Respondent Type	Number of Respondents	%
Museum	13	46,4%
Library	3	10,7%
Archive	5	17,9%
Sound archive	0	
Publisher	0	
Aggregator	5	17,9%
Other	14	50,0%

Among the "Other" participants are ministries, government agencies, university research centres and service institutes which often also serve as aggregators for other institutions (4 of 14). Aggregators and others sometimes answered in place of their associated providers, sometimes they reported back about their own activities or both. Because it was difficult to differentiate the answers in these cases they were adopted simply as one answer without changes.

The distribution of responses to respondent types reflects the distribution of the partners in the project. It is apparent that respondents represent mostly museums and other institutions, other specific sectors like libraries are less represented.

3.2.2 Countries

Here are the figures for the countries where respondents are based:

Country	Number of Respondents
Belgium	3
Croatia	1
Cyprus	1
Czech Republic	1
Estonia	1
France	3
Germany	2
Greece	1
Hungary	2
Italy	2
Lithuania	2
Poland	1
Romania	2
Spain	2
Sweden	3
United Kingdom	1

The distribution of responses reflects the distribution of the partners in the project.

Taken as a whole, it can be concluded from the information about respondents that the sample is fairly representative for the AthenaPlus partner consortium.

3.3   Using Linked Open Data

The questions concerning the using of LOD aimed to find out if the concept of LOD was well known and if the partners already had made experiences in consuming LOD.

3.3.1   Level of Familiarity with the Concept of LOD

The question "Are you or your organisation familiar with the concept of Linked Open Data (LOD)?" was answered as follows:

Familiar with the concept of LOD	Number of Respondents	%
Yes:	25	89,3%
No:	3	10,7%

This reflects that the concept of Linked Open Data is well known in the community. In particular, progress can be stated over the last two years: In a similar survey carried out in the Linked Heritage project[8] in 2011 still 25% of the respondents answered 'No' to the same question.

The three answers with "No" came from three different countries and three different types of institution so no specific area of "Non-familiarity" can be spotted.

3.3.2   Experience of Using LOD

The question "Have you or your organisation had experience of using LOD in connection with your collections?" was answered as follows:

---

[8] D2.1 Best practice report on cultural heritage linked data and metadata standards p. 19,
http://www.linkedheritage.org/getFile.php?id=229 [Published November 2011]

Experience of Using LOD	Number of Respondents	%
Yes:	6	21,4%
No:	22	78,6%

This shows that although the concept of LOD is well known in the community there is still little experience to apply it.

The six partners, who responded with "Yes" came from different countries and institutional areas. They were asked to detail their experiences. Sources mentioned in the answers were: ArhivX LOD[9], GeoNames[10], GND[11], ICC[12], VIAF[13]. Only two partners use RDF and SPARQL as genuine LOD techniques. One partner actually is implementing these techniques. The other partners still use manual or semi-automatic processes or proprietary systems. The quality of LOD sources is rated from "very good" (GeoNames, ICC, VIAF) to "sufficient" (GND).

3.3.3 Knowledge / awareness of National LOD Projects

Furthermore, partners were asked about their notice of national LOD projects in the field of cultural heritage. The question "Do you or your organisation know of any LOD projects or initiatives in your country in the field of cultural heritage?" was answered as follows:

Knowledge / awareness of National LOD Projects	Number of Respondents	%
Yes:	19	67,9%
No:	9	32,1%

The responses show that there is a potential for cooperation at national level, so that possibly LOD connections can be created at this level. The projects specifically mentioned are taken into account for the selection of LOD sources in section 5.1.

3.4 Publishing Linked Open Data

3.4.1 Experience of Publishing LOD

The question "Have you or your organisation had experience of publishing LOD in connection with your collections?" was answered as follows:

Experience of Publishing LOD	Number of Respondents	%
Yes:	4	14,3%
No:	24	85,7%

The responses show that actually the level of experience of publishing data with LOD techniques is quite low and that the partners are less experienced in publishing LOD than in consuming it. Nevertheless, the sources already published by the partners are of course most suitable to be used by

---

[9] http://data.arhivx.net
[10] http://www.geonames.org/
[11] http://www.dnb.de/gnd
[12] http://www.icc.cat
[13] http://viaf.org/

the other partners as well. The projects published by the partners are taken into account for the selection of LOD sources in section 5.1.

3.4.2 *Plan to publish LOD*

The responses to the question "Does your organisation plan to publish LOD in the near future?" were as follows:

Plan to publish LOD	Number of Respondents	%
Yes:	21	75,0%
No:	7	25,0%

A majority of 75%, that is three of four partners, are planning to publish LOD in the near future. The responses to the additional question about the details of the planned publication show different approaches, using different formats and technologies. The answers show that partners seem to know well how to address a specific implementation, e.g. which ontologies they may reuse, but also that the shape of the final realization is not yet fixed.

Among the ontologies to be used for publication are mentioned Dublin Core and SKOS (6 times each), CIDOC CRM (5 times), LIDO, own ontology(3 times each), EDM , ESE, FOAF, FRBRoo (2 times each).

As for serving the data many institutions answered that it is not yet fixed and / or it depends on other project's decisions. However, data dump (4 times) and SPARQL (3 times) were explicitly mentioned.

One institution informed that specific focus will be given to reuse of and connecting with existing LOD sources.

3.4.3 *Plan to connect with new LOD sources*

The question "Does your organisation plan to connect with new LOD sources in the near future?" was answered as follows:

Plan to connect with new LOD sources	Number of Respondents	%
Yes:	14	50,0%
No:	13	46,4%
No answer:	1	3,57%

50 % of responses expressed the intention to connect with new LOD sources in the near future. 50 % of positive responses expressed the intention to connect with the Getty thesauri when they will be published as LOD[14]. Other LOD sources listed more than one time were: DBpedia[15], Geonames, LOC[16], SOCH[17], VIAF, and Wikidata[18]. Only once were mentioned: British Museum[19], CLAROS data[20], Iconclass[21], Plejades[22], and GND.

Furthermore it was mentioned that sources should be trustworthy and provide a SPARQL endpoint.

3.4.4 *Becoming Aware of New LOD Sources*

There were 20 responses (71,43%) to the question "How do you notice that there are new LOD sources as candidates to connect with?". Six participants gave the answer that the question does not concern

---

[14] http://www.getty.edu/research/tools/vocabularies/lod/index.html
[15] http://dbpedia.org
[16] http://id.loc.gov/
[17] http://www.ksamsok.se
[18] www.wikidata.org
[19] http://collection.britishmuseum.org/
[20] http://data.clarosnet.org/
[21] http://www.iconclass.nl
[22] http://pleiades.stoa.org/home

them or they do not look for new LOD sources systematically. The majority replied that they would become aware of new LOD sources "mainly by keeping in touch with the community" which means by conferences, colleagues, social media, web scouting, mailing lists and newsletters. One response explicitly mentioned as possible sources of information: Open Directory Project[23], PSI[24], LLD XG[25], and Datahub[26]. In addition also OpenGLAM[27] and CEST[28] were mentioned.

3.4.5 *Language Support*

With regard to the language support as a prerequisite to connect with LOD sources almost all participants require generally the support of their own national languages or English, in addition languages of neighbouring countries are listed.
Two partners interpreted the question in a different way and answered that RDF has to be supported. Interestingly these were the two partners with apparently most experience in LOD techniques.

3.4.6 *Expectations and Criteria*

There were 16 responses (57,14%) to the request to "describe the expectations and criteria (if there are some) for connecting with LOD sources". Three partners without experience with LOD explained to have no expectations yet. Among the positive answers the most common expectations were: enlarging accessibility of data in a broader content, increasing the visibility of collections, extend the semantic relations between various collections, development of cross-domain interdisciplinary networks of knowledge, possibility of re-contextualizing the resources for improved research infrastructure. Recognized as an added value for the own collections was the possibility to enrich your own data via (inter)national connections. One reply mentioned the prospect of easy access to valuable information for scientific research and the purpose to create educational apps.

However, there are not only positive expectations. A possible disadvantage would be to lose control over the published data. In addition, a possible decrease in the quality is feared e.g. due to links to non-qualified sources. The question was raised whether an overload of links may lead to a loss of visibility and/or accessibility. Other questions focused on copyright and sometimes to the lack of staff, expertise and funding.

As one may expect there were less answers to criteria than to expectations because for developing criteria there is a need to have concrete expectations in the first place. Criteria were listed as follows: need for trustful sources, provider (and data) need to be reliable, relevance, thoroughness and quality of the information, regular updates, popularity, persistent URIs. Also important is the simplicity of the linking process which should be more formalized. Two responses of aggregators with broader experience added as technical criteria the use of URLs, API connectors and widgets that expose the LOD sources.
To sum up what is required: a high quality of the LOD sources in terms of metadata, sustainability, trust and ease of use, at the technical level the availability of the data in SKOS or more generally in RDF.

3.4.7 *Quality check*

The question "Do you check the quality of LOD sources?" was answered as follows:

Check quality	Number of Respondents	%
Yes:	9	32,14%
No:	12	42,86%
(Not answered):	7	25,00%

---

[23] http://www.dmoz.org/rdf.html
[24] http://epsiplatform.eu/
[25] http://www.w3.org/2005/Incubator/lld/
[26] http://datahub.io/
[27] http://openglam.org/
[28] http://www.projectcest.be/index.php/Linked_Open_Data_value_vocabularies

The answer to this question must be seen in connection with the questions about the experience with publishing or using LOD sources relating to the collections. Of six participants who had no experience with LOD sources, only one replied "No". Four answers with "Yes" came from partners who had no experience in using LOD sources.

Two responses describe the criteria in detail: (1) "there are several criteria: the complexity and accuracy of an ontology, the stability of URIs, the adoption within the community, the method of serving (SPARQL endpoints vs. dumps), the format (rdf/xml vs. csv) …", (2) "the indicators that are usually evaluated are: producer of the data and the authors, relevance, standardization and use of standard formats, sources used, additional information and illustration of information, etc.". Most of the respondents actually check sources more or less manually via test queries and by looking at sample data, or semi-automatically via scripts.. The reliability of the data provider often is mentioned as important. Also for checking quality it is (good) practice to ask colleagues and experts, and to get information through the community network.

4 FINDING LINKED OPEN DATA SOURCES FOR ATHENAPLUS

To find qualified LOD sources and gather the relevant information about them was not as easy as expected. To get the results presented in chapter 5 some experiences had to be gained which resulted in the following approaches.

4.1 … by asking the community

The responses of the partner survey to the question "How do you notice that there are new LOD sources as candidates to connect with?" revealed that the common practice to find new LOD sources mostly follows a word-of-mouth recommendation (see 3.4.4). The sources suggested by the partners in the survey are described in section 5.1.

4.2 … by exploring the datahub

Searching for LOD sources in the Web, using Google of course results in many hits but the authors were looking for a more systematic approach, preferably using LOD techniques, e.g. querying SPARQL endpoints. As a focal point to find publications of LOD sources the datahub[29] was analyzed as it is also recommended in the "*Best practice report on cultural heritage linked data and metadata standards*"[30].

In this section an overview of the datahub GUI is given (4.1.2.1), followed by a description of the approach to find LOD sources with the datahub (4.1.2.2).

4.2.1 The datahub GUI

The datahub is an installation of CKAN[31] (Comprehensive Knowledge Archive Network) and provides access to datasets and descriptive information about these datasets, it serves as a registry. CKAN is a web based data management system developed by the Open Knowledge Foundation[32] that is also used by Open Government projects like data.gov.uk or data.norge.no. CKAN offers several features for describing datasets like Groups, Keywords, and Labels, which hence can be used for the exploration of big data.

---

[29] http://datahub.io
[30] http://www.linkedheritage.org/getFile.php?id=229 [Published November 2011]
[31] http://ckan.org/
[32] http://okfn.org/

Figure 2: CKAN Metadata (09/10/2013)[33]

The screenshot above shows one dataset item of the datahub with explanations of the features of CKAN. It has to be pointed out that all the metadata of a dataset is defined by the publisher of the dataset, that is the one who registers a dataset in the datahub.

Each dataset has a unique URL that can be used for SPARQL queries as explained below. There are several categories of labels, which allow the description of the dataset with keywords. The labels of the datahub are categorized as "Groups", "Tags", "Formats" and "Licence" and can be used for a faceted search in the browser. Each category can be used more than one time. The meaning of the categories is not distinct but overlapping. This is not obvious at a first glance and turns out to be a problem for a systematic approach. "Groups" and "Tags" are similar categories that allow for the description of the dataset, but easily can be confounded and consequently produce less precise search results. "Formats" are shown in a prominent way in the search result list but not in the single view of a dataset. In the single view "Formats" are only shown on the left side in the column for faceted search. "Tags" often have overlapping meanings with "Formats" and "Groups". The usability of the label, to find similar datasets sharing the same notations of labels, is hampered by the fact that publishers apparently tend to define new labels which differ little from labels used for other datasets. There is a bunch of labels which hardly can be systematically approached. In other words there is no controlled vocabulary available for the labels. Furthermore, not all categories seem to be retrievable by a SPARQL query, at least it was despite all efforts not obvious for the authors how to query "Groups" through the SPARQL endpoint of the datahub.

For each dataset, in addition to their description data and resources can be uploaded. The publishers can give information and access to example files, SPARQL endpoints, sitemaps, vocabulary etc. Direct access to the data and resources is possible by following the "explore" button. With "Additional Info" publishers can give information about version, amount of items and links, namespace etc.

---

[33] http://ckan.org/features/

The following screenshot of the CKAN website shows the search result page and explains it.
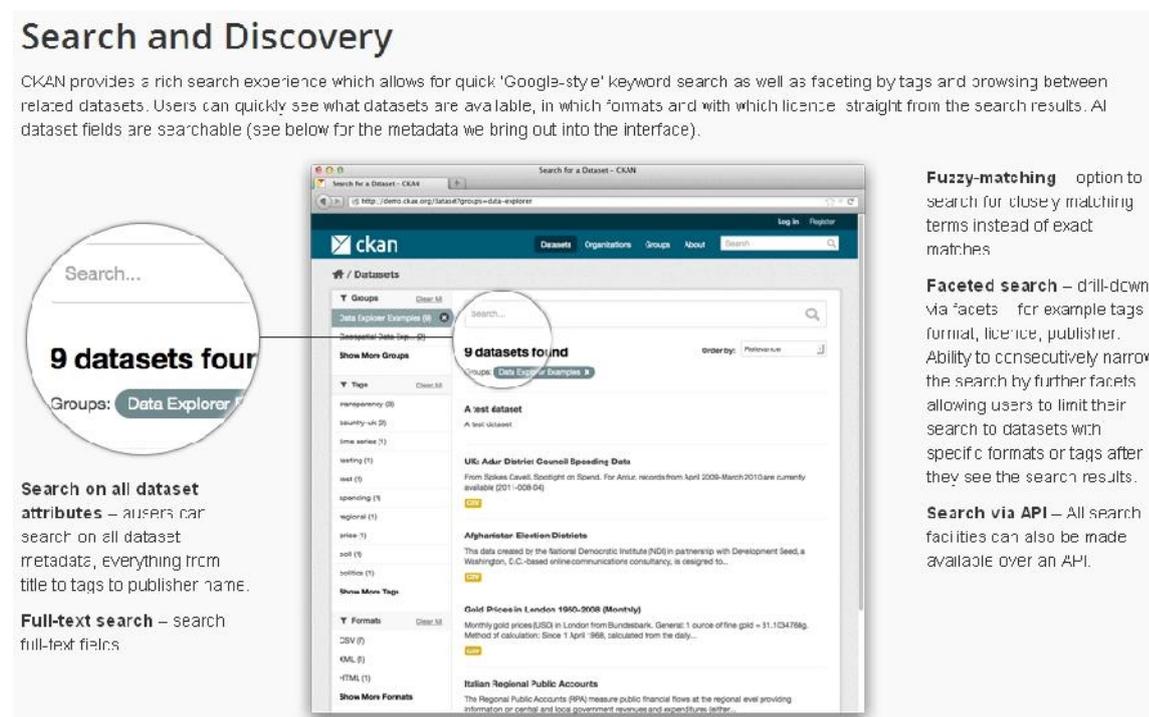
Figure 3: CKAN Search and Discovery (09/10/2013)[34]

Furthermore CKAN has an indicator for the use of datasets, data and resources. Often used datasets and resources are promoted by a "popular" badge and a tooltip showing the number of views.

Figure 4: CKAN Highlighting of Popular Datasets and Resources (09/10/2013) [35] [36]

Finally, after a group search a tab for the "activity stream" of the group is visible, so one can easily see whether the group is still in use or not.

4.2.2 Approach to find LOD sources

In a first approach, the authors wanted to explore the datahub via SPARQL queries, starting from the Europeana LOD dataset[37] and looking for relations to other datasets. Surprisingly it was not as easy to understand the structure of the datahub and how to access its content systematically. The only documentation found through the actual website was a link to the CKAN API documentation[38] on the "About"[39] page, which is only helpful for developers. The search page of the CKAN documentation[40]

---

[34] http://ckan.org/features/
[35] http://docs.ckan.org/en/769-docs-reorg/_images/popular-dataset.png
[36] http://docs.ckan.org/en/769-docs-reorg/_images/popular-resource.png
[37] http://datahub.io/de/dataset/europeana-lod
[38] http://docs.ckan.org/en/ckan-2.0/#the-ckan-api
[39] http://datahub.io/about
[40] http://docs.ckan.org/en/ckan-2.0/search.html

returned no results for a search for "sparql". Finally the CKAN dataset itself in the datahub[41] revealed the SPARQL endpoint. Of course a Google search with "datahub sparql" also leads easily to the ckan-dev site[42] and subsequently the link to the SPARQL endpoint of the datahub. However, a minimum documentation should be available on the datahub site itself.

For subsequent exploration of the datahub via SPARQL queries the authors used several tools, e.g. the SPARQLer Query validator[43] and the Virtuoso SPARQL Query Editor of the CKAN SPARQL endpoint[44]. Once familiar with the query mechanisms of the datahub, the SPARQL query documented in appendix 5 proved to return the results presented in chapter 5.

Taken as a whole, a systematic exploration of the datahub via SPARQL queries turned out to be rather costly in terms of labour and time, if not impossible. The approach was only useful to a limited extent for the purpose to retrieve LOD source candidates. In addition, it appears as a shortcoming that there is no controlled vocabulary for keywords ("tags") and labels ("formats"). As already mentioned each data provider can define own keywords and labels and so far no evolvement towards harmonization can be identified, but rather the different denominations still seem to grow. The screenshot of the "OpenGLAM" group of the datahub with an amount of 29 datasets including the "Europeana LOD" group can serve as an example for this observation (cf. appendix 4): There are many items with overlapping terms for "Groups", "Tags" and "Formats". By the specified number behind the terms, it becomes clear that many terms are used only sporadically and are not relevant to the group as a whole. It can be concluded that the interface allows for an individual exploration of existing sources by hopping from one to another, but a systematic analysis and evaluation of the datasets in the datahub could not be achieved in an easy way. The LOD source candidates from the datahub presented in section 5.2 hence are chosen manually, only the descriptive information about the datasets could be extracted by SPARQL queries.

However, the CKAN datahub is the most important registry of LOD sources. When publishing LOD sources it is recommended to follow the W3C "Guidelines for Collecting Metadata on Linked Datasets in CKAN"[45]. The data provider has to ensure the quality of the description data. A site offering a validator for a datahub entry with step-by-step guidance is available at http://validator.lod-cloud.net/levels.html. In addition, the LATC[46] best practices web page[47] and especially the dataset publishing checklist[48] are very useful.

---

[41] http://datahub.io/dataset/ckan
[42] http://lists.okfn.org/pipermail/ckan-dev/
[43] http://www.sparql.org/query-validator.html
[44] http://semantic.ckan.net/sparql
[45] http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKANmetainformation
[46] http://latc-project.eu (LATC is a Specific Support Action in the context of the FP7 ICT Programme)
[47] http://latc-project.eu/best-practices
[48] https://docs.google.com/document/d/150dJSMZk5W5ucF23hGj62DaoKtTk9qeaEPBN_VCCihI

5 LINKED OPEN DATA SOURCES FOR ATHENAPLUS CONTENT

This chapter presents the data sources identified as suitable to connect with the AthenaPlus content. Section 5.1 describes in detail the sources suggested by the AthenaPlus partners in the survey, including sources published by partners as well as sources to which partners have already connected their content, or plan to connect in the near future. Section 5.2 describes sources from the datahub, identified through an analysis of the Europeana LOD data set. Section 5.3 5.3provides a summary of the findings.

5.1 LOD Sources identified by the Partners

The following tables of LOD sources result from the LOD survey among the AthenaPlus partners. They describe sources which were mentioned by the partners in the survey and which contain both sources of the partners themselves as well as sources to which partners plan to connect with in the future. The structure of the tables follows the common structure used in the datahub to the description of the sources. The descriptions of the LOD sources in the following tables are taken from the datahub in case they are registered, or from the web sites of the publisher of the sources themselves. In both cases the information is provided by the publishers themselves, no further checks were made. By default the information is extracted automatically from the datahub. Only if there is no entry in "Link datahub" the information was manually searched on the web site of the publisher. In case of doubt, the tags [datahub] or [site] identify the different origin of the information.

The following table explains the names:

Table Row	Meaning
Name	Name of the publisher
Author	[datahub:] Creator of published dataset; [site:] Author of information
Maintainer	Contributor or publisher or provider of the dataset
Link LOD source	Link to LOD source
Link datahub	Link to dataset of LOD source in datahub
Description of content	Description of dataset
Licence	Licence information
Amount of resources	Amount of resources
Protocols supported	Protocols supported by the publisher. Can be data dump, HTTP (with content negotiation), OAI-PMH, SRU, SPARQL.
Keywords datahub	Named keywords in datahub.
Formats datahub	Named formats in datahub
Namespaces	Namespace declarations as found in RDF example files from datahub or in examples from the web site of the publisher.
Data and Resources	List of data and resources from datahub with names and URLs. Only relevant resources for our purposes as RDF example files, vocabulary files, sitemap files, SPARQL endpoints or other information is listed.
Additional Info	Listing of additional info from datahub with names and values.

Each of the following tables describes a LOD source identified as candidate to connect with AthenaPlus content. The following tables present an overview of all 18 datasets.

5.1.1 BNE (Spanish National Library)

Name	datos.bne.es
Author	Ontology Engineering Group, Facultad de Informática, Universidad Politécnica de Madrid
Maintainer	Daniel Vila-Suero and Boris Villazón-Terrazas
Link LOD source	http://datos.bne.es/
Link datahub	http://datahub.io/dataset/datos-bne-es
Description of Content	Open bibliographic linked data from the Spanish National Library including 4 million authority records, 2.4 million bibliographic records resulting in over 58 million triples. The Spanish data are modelled using IFLA vocabularies. For the migration and linkage, we used Marimba, a particular tool developed by us and designed for being used by librarians. Data • 4 Million authority records • 2.4 million of bibliographic records • 58 million of RDF triples • 600K links with other resources (VIAF, SUDOC, GND (German national library), Libris, DBpedia)
Licence	http://www.opendefinition.org/licenses/cc-zero
Amount of resources	58,053,215 triples
Protocols supported	Dump, SPARQL
Keywords Datahub	lod, no-license-metadata, published-by-producer, publications, no-proprietary-vocab, bibliographic, cc0, format.ifla, format.rdf, lodcloud
Formats Datahub	api/sparql, application/turtle, meta/void, bz2:nt
Namespaces	
Data and Resources	
Name	URL
Sparql Endpoint	http://datos.bne.es/sparql
Example resource (Turtle)	http://datos.bne.es/resource/XX1718747
bne.ttl	https://commondatastorage.googleapis.com/ckannet-storage/2012-05-23T142816/bne.ttl
Data dump of authority data in (N-Triples)	http://datos.bne.es/datadumps/aut091211.nt.bz2
Data dump of bibliographic data in (N-Triples)	http://datos.bne.es/datadumps/bib121211.nt.bz2
Data dump of owl:sameAs links to viaf.org	http://datos.bne.es/datadumps/links-viaf.nt.bz2
Data dump of owl:sameAs links to DBPedia, GND, Libris and Sudoc	http://datos.bne.es/datadumps/links-others.nt.bz2

5.1.2 *British Museum Collection*

Name	British Museum Collection
Author	Dominic Oldman
Maintainer	The British Museum
Link LOD source	http://collection.britishmuseum.org
Link datahub	http://datahub.io/dataset/british-museum-collection
Description of content	Welcome to this Linked Data and SPARQL service. It provides access to the same collection data available through the Museum's web presented Collection Online, but in a computer readable format. The use of the W3C open data standard, RDF, allows the Museum's collection data to join and relate to a growing body of linked data published by other organisations around the world interested in promoting accessibility and collaboration. The data has also been organised using the CIDOC-CRM (Conceptual Reference Model) crucial for harmonising with other cultural heritage data. The current version is beta and development work continues to improve the service. We hope that the service will be used by the community to develop friendly web applications that are freely available to the community. Licence available at http://collection.britishmuseum.org/Licensing
Licence	Other (Attribution) [see Additional Info: licence_link]
Amount of resources	105,635,648 triples
Protocols supported	SPARQL
Keywords datahub	cultural, culturalheritage, deref-vocab, format-cidoc-crm, format-foaf, format-rdf, format-skos, lod, lodcloud.candidate, no-license-metadata, no-provenance-metadata, no-vocab-mappings, publications, published-by-producer, sparql
Formats datahub	api/sparql, example/rdf+xml
Data and Resources	
Name	URL
SPARQL endpoint	http://collection.britishmuseum.org/Sparql
Object EAF119791	http://collection.britishmuseum.org/description/object/EAF119791.rdf
Additional Info	
Name	Value
licence_link	http://collection.britishmuseum.org/Licensing
namespace	http://collection.britishmuseum.org/id/
triples	105635648

5.1.3   *CLAROS data*

Name	CLAROS data
Author	N/A
Maintainer	University of Oxford's e-research centre, OeRC.
Link LOD source	http://data.clarosnet.org/
Link datahub	N/A
Description of content	Built on the art of ancient Greece and Rome, CLAROS is an international research collaboration, using the latest Information and Communication Technologies to enable simultaneous searching of major collections in university research institutes and museums. The CLAROS Data service provides a RESTful interface for the data of the CLAROS Project, and complements the CLAROS Explorer. This service provides metadata about archaeology and art in machine-readable formats such as RDF, JSON and KML. The data for the CLAROS project are modelled using the Erlangen OWL-DL 1.0 implementation of the CIDOC Conceptual Reference Model using the http://purl.org/NET/crm-owl# namespace.
Licence	ODbL (?) [http://id.clarosnet.org/dataset/claros]
Amount of resources	11,649,742 elements [http://www.clarosnet.org/XDB/ASP/clarosHome/technicalCidoc.html]
Protocols supported	SPARQL
Keywords datahub	N/A
Formats datahub	N/A
Namespaces	foaf,='http://xmlns.com/foaf/0.1/' rdfs='http://www.w3.org/2000/01/rdf-schema#' dc='http://purl.org/dc/elements/1.1/' rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#' dcterms='http://purl.org/dc/terms/' crm='http://purl.org/NET/crm-owl#'
Data and Resources	
Name	URL
SPARQL endpoint	http://data.clarosnet.org/sparql/

5.1.4   *Cultura italia*

Name	CulturaItalia
Author	N/A
Maintainer	Cultura Italia. Ministero per i Beni e le Attività Culturali.
Link LOD source	http://dati.culturaitalia.it/
Link datahub	N/A
Description of content	The pilot project dati.culturaitalia.it contains open data sets of CulturaItalia released by the project partners with the following licence: "CC0 1.0 Universal Public Domain Dedication". Currently dati.culturaitalia.it makes available metadata from:

	Accademia S. Cecelia, Progetto ArtPast, Digibess, ICCU, Internet Culturale, Michael Italia, Polo Museale Fiorentino, Regione Marche and Anagrafe delle Biblioteche Italiane in addition to the Thesaurus PICO. Data of dati.culturaitalia.it are available both through a SPARQL endpoint, and through an OAI Provider.
Licence	CC0 1.0 Universal Public Domain Dedication [http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/linked_open_data/Licenza_CulturaItalia_CC0.pdf]
Amount of resources	Number of statements: 15,048,155 Number of entities: 5,277,737
Protocols supported	OAI-PMH, SPARQL
Keywords datahub	N/A
Formats datahub	N/A
Namespaces	N/A
Data and Resources	
Name	URL
SPARQL endpoint	http://dati.culturaitalia.it/sparql/index.jsp?locale=en

5.1.5 *DBpedia*

Name	DBpedia
Author	DBpedia Team - http://wiki.dbpedia.org/Imprint
Maintainer	DBpedia Team - http://wiki.dbpedia.org/Imprint
Link LOD source	http://dbpedia.org/
Link datahub	http://datahub.io/dataset/dbpedia
Description of Content	DBpedia.org is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia and to link other datasets on the Web to Wikipedia data. The DBpedia knowledge base currently describes more than 3.64 million things, out of which 1.83 million are classified in a consistent Ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 video games, 169,000 organisations, 183,000 species and 5,400 diseases. The DBpedia data set features labels and abstracts for these 3.64 million things in up to 97 different languages; 2,724,000 links to images and 6,300,000 links to external web pages; 6,200,000 external links into other RDF datasets, 740,000 Wikipedia categories, and 2,900,000 YAGO categories. The DBpedia knowledge base altogether consists of over 1.2 billion pieces of information (RDF triples) out of which 335 million were extracted from the English edition of Wikipedia and 865 million were extracted from other language editions.
Licence	http://www.opendefinition.org/licenses/cc-by-sa
Amount of resources	1,200,000,000 triples (datahub)
Protocols supported	SPARQL

Keywords Datahub	deref-vocab, format-rdf, lod, no-license-metadata, no-vocab-mappings, rdf, published-by-producer, ckanupload.esw.200910, provenance-metadata, format-rdfs, access-www, format-owl, access-bulk, access-api, crossdomain, format-xhtml, format-foaf, void-sparql-endpoint, wikipedia, search, format-prov
Formats Datahub	api/sparql, application/x-ntriples, meta/rdf-schema, meta/void, linked data, application/rdf+xml, text/turtle, HTML, meta/sitemap
Namespaces	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:dbpprop="http://dbpedia.org/property/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:ns5="http://dbpedia.org/ontology/PopulatedPlace/" xmlns:dbpedia-owl="http://dbpedia.org/ontology/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" xmlns:ns9="http://www.w3.org/ns/prov#" xmlns:grs="http://www.georss.org/georss/"

Data and Resources	
Name	URL
Download page (N-Triples, bz2-compressed)	http://wiki.dbpedia.org/Downloads
SPARQL endpoint	http://dbpedia.org/sparql
Download	http://downloads.dbpedia.org/3.5.1/all_languages.tar
RDF Schema	http://downloads.dbpedia.org/3.5.1/dbpedia_3.5.1.owl.bz2
voiD description	http://dbpedia.org/void/Dataset
Link to an example Linked Data entry	http://dbpedia.org/resource/Berlin
Link to an example data item within the dataset (RDF/XML)	http://dbpedia.org/data/Berlin.rdf
Link to an example data item within the dataset (Turtle)	http://dbpedia.org/data/Berlin.n3
Link to an example data item within the dataset (RDFa)	http://dbpedia.org/page/Berlin
Semantic Web Sitemap	http://wiki.dbpedia.org/sitemap

Additional Info	
Name	Value
Version	2010-09-02 (3.7)
links:2000-us-census-rdf	12529
links:dbtune-musicbrainz	22981
links:education-data-gov-uk	1697
links:eunis	3600
links:flickr-wrappr	8800000
links:freebase	3400000
links:fu-berlin-dailymed	43

links:fu-berlin-dblp	196
links:fu-berlin-diseasome	1943
links:fu-berlin-drugbank	729
links:fu-berlin-eurostat	137
links:fu-berlin-project-gutenberg	2510
links:fu-berlin-sider	751
links:geonames-semantic-web	86547
links:geospecies	11400
links:italian-public-schools-linkedopendata-it	5822
links:linkedgeodata	53024
links:linkedmdb	13800
links:nytimes-linked-open-data	10359
links:opencyc	20362
links:rdf-book-mashup	9078
links:reference-data-gov-uk	22
links:revyu	6
links:tcmgenedit_dataset	904
links:transport-data-gov-uk	3768
links:uk-legislation-api	33
links:w3c-wordnet	467101
links:wikicompany	8348
links:world-factbook-fu-berlin	233
links:yago	18100000
namespace	http://dbpedia.org/resource/
triples	1200000000

5.1.6 *GeoNames*

Name	GeoNames Semantic Web
Author	Bernard Vatant
Maintainer	[N/A]
Link LOD source	http://www.geonames.org/ontology/
Link datahub	http://datahub.io/dataset/geonames-semantic-web
Description of Content	The GeoNames Ontology makes it possible to add geospatial semantic information to the World Wide Web. All over 6.2 million geonames toponyms now have a unique URL with a corresponding RDF web service. Other services describe the relation between toponyms. SPARQL endpoints for this data are made available by third parties. Geonames at Kasabi, and FactForge has it loaded as well.

Licence	http://www.opendefinition.org/licenses/cc-by
Amount of resources	93,896,732 triples
Protocols supported	RDF download
Keywords Datahub	format-rdf, geographic, lod, lodcloud.nolinks, no-license-metadata, no-provenance-metadata, no-vocab-mappings, published-by-producer, format-geo, provenance-metadata, no-deref-vocab, geography, format-cc
Formats Datahub	meta/rdf-schema, application/rdf+xml, example/rdf+xml
Namespaces	xmlns:cc="http://creativecommons.org/ns#" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:gn="http://www.geonames.org/ontology#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#" <gn:Feature rdf:about="http://sws.geonames.org/3020251/"
Data and Resources datahub	
Name	URL
RDF Schema	http://www.geonames.org/ontology/ontology_v2.1.rdf
Download	http://download.geonames.org/all-geonames-rdf.zip
Example (RDF/XML)	http://sws.geonames.org/3020251/about.rdf
Example (Mother Earth)	http://sws.geonames.org/6295630/
Additional Info	
Name	Value
Version	2010-06 (2.1)
namespace	http://sws.geonames.org/
shortname	GeoNames
Triples	93896732

5.1.7 *Gencat Open Data Project*

Name	Gencat Open Data Project
Author	Generalitat de Catalunya
Maintainer	Generalitat de Catalunya
Link LOD source	http://www20.gencat.cat/portal/site/dadesobertes?newLang=en_GB
Link datahub	N/A
Description of content	This is a Government of Catalonia website which publishes public data in order to promote the use and re-use of information generated by the government. The data comes from different bodies of the Government of Catalonia and the data has been sorted into groups in the catalogue.
Licence	Various: How to use the data (terms of use and licences)]
Amount of resources	N/A

Protocols supported	Dump
Keywords datahub	N/A
Formats datahub	N/A
Namespaces	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dct="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:dctype="http://purl.org/dc/dcmitype/" xmlns:dcat="http://vocab.deri.ie/dcat#
Data and Resources	
Name	URL
Terminology repertoires	http://www20.gencat.cat/portal/site/dadesobertes/menuitem.db4d3cf2bccf921baacf3010b0c0e1a0/?vgnextoid=49b19ee9acb42310VgnVCM1000000b0c1e0aRCRD&vgnextchannel=49b19ee9acb42310VgnVCM1000000b0c1e0aRCRD&vgnextfmt=detall2&id=21&newLang=en_GB
Data Formats	XLS, DGN, RSS, XML, CSV, SHP, PDF, JPG, PNG, SVG, KML, BMP, TIFF, GIF, JSON, XLSX, RDF, ZIP, TXT, ICS, DWG, Serialized PHP, RINEX, WMS, ASCII, HTM, DXF, JP2, EXP, MHTML, DOC, RTCM, SID, GeoRSS, MMZ, DBF, GeoPDF, TERM, CSW

5.1.8 *Getty Art & Architecture Thesaurus (AAT)*

Name	Art & Architecture Thesaurus Online
Author	The J. Paul Getty Trust
Maintainer	N/A
Link LOD source	http://www.getty.edu/research/tools/vocabularies/aat
Link datahub	N/A
Description of content	The AAT is a structured vocabulary currently containing around 268,650 terms and other information about concepts. Terms in AAT may be used to describe art, architecture, decorative arts, material culture, and archival materials. The target audience includes museums, libraries, visual resource collections, archives, conservation projects, cataloging projects, and bibliographic projects. Terms for any concept may include the plural form of the term, singular form, natural order, inverted order, spelling variants, scientific and common forms, various forms of speech, and synonyms that have various etymological roots. Among these terms, one is flagged as the preferred term, or *descriptor*. There may be multiple descriptors reflecting usage in multiple languages. Preferences for individual contributors may differ and are noted. The AAT is a *thesaurus* in compliance with ISO and NISO standards. The focus of each AAT record is a concept. Currently here are around 51,470 concepts in the AAT. In the database, each concept's record (also called a *subject*) is identified by a unique numeric ID. Linked to each concept record are terms, related concepts, a *parent* (that is, a position in the hierarchy), sources for the data, and notes. The temporal

	coverage of the AAT ranges from Antiquity to the present and the scope is global.
Licence	Not open: http://www.getty.edu/research/tools/vocabularies/obtain/index.html The AAT is scheduled for release as Linked Open Data in 2013/2014.
Amount of resources	around 268,650 terms, around 51,470 concepts
Protocols supported	N/A
Keywords datahub	N/A
Formats datahub	N/A
Namespaces	N/A

5.1.9 *Getty Thesaurus of Geographic Names (TGN)*

Name	Getty Thesaurus of Geographic Names (Getty Research Institute)
Author	Getty Research Institute
Maintainer	[N/A]
Link LOD source	http://www.getty.edu/research/tools/vocabularies/tgn/
Link datahub	http://datahub.io/dataset/getty-tgn
Description of Content	TGN is a structured vocabulary currently containing around 2,035,195 names and other information about places. Names for a place may include names in the vernacular language, English, other languages, historical names, names and in natural order and inverted order. Among these names, one is flagged as the *preferred name*. TGN is a thesaurus, compliant with ISO and NISO standards for thesaurus construction; it contains hierarchical, equivalence, and associative relationships. Note that TGN is not a GIS (Geographic Information System). While many records in TGN include coordinates, these coordinates are approximate and are intended for reference only. The focus of each TGN record is a place. There are around 1,431,380 places in the TGN. In the database, each place record (also called a *subject*) is identified by a unique numeric ID. Linked to the record for the place are names, the place's *parent* or position in the hierarchy, other relationships, geographic coordinates, notes, sources for the data, and *place types*, which are terms describing the role of the place (e.g., *inhabited place* and *state capital*). The temporal coverage of the TGN ranges from prehistory to the present and the scope is global. [http://www.getty.edu/research/tools/vocabularies/tgn/about.html]
Licence	Not open: http://www.getty.edu/research/tools/vocabularies/obtain/index.html It is planned that the Getty vocabularies will be released as Linked Open Data in 2013/2014.

Amount of resources	around 2,035,195 names, thereof around 1,431,380 places
Protocols supported	[N/A]
Keywords Datahub	geographic, thesaurus, size-xlarge, getty
Formats Datahub	[N/A]
Namespaces	[N/A]

5.1.10 *GND - Integrated Authority File of German National Library (DNB)*

Name	Gemeinsame Normdatei (GND) – Integrated Authority File)
Author	[N/A]
Maintainer	Deutsche Nationalbibliothek (German National Library)
Link LOD source	http://www.dnb.de/EN/gnd
Link datahub	http://datahub.io/dataset/dnb-gemeinsame-normdatei
Description of Content [site]	The Integrated Authority File (GND) contains data records representing on persons, corporate bodies, congresses, geographic entities and works which were entered until April 2012 in the following separate files: Corporate Body Authority File (GKD), Name Authority File (PND), Subject Headings Authority File (SWD) and Uniform Title File of the Deutsches Musikarchiv (DMA-EST file).

The authority records are used in descriptive and subject cataloguing. They constitute a common, conclusive reference system for bibliographic library data and for the cataloguing data of other authority data users such as archives, museums and other scientific and cultural institutions.

GND authority records form part of the Virtual International Authority File (VIAF); they are combined with data from other national authority files. |
Licence	http://creativecommons.org/publicdomain/zero/1.0/
Amount of resources	100,720,519 triples [datahub]
Protocols supported	Dump, OAI, SRU [site]
Keywords Datahub	format-rdf, lod, no-license-metadata, no-provenance-metadata, no-vocab-mappings, rdf, published-by-producer, format-rdfs, lld, crossdomain, no-deref-vocab, bibliographic, buildings, germany, format-foaf, format-skos, persons, corporations, authorities, concepts, places
Formats Datahub	XML, example/rdf+xml
Namespaces	xmlns:gnd="http://d-nb.info/standards/elementset/gnd#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rda="http://rdvocab.info/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:isbd="http://iflastandards.info/ns/isbd/elements/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:marcRole="http://id.loc.gov/vocabulary/relators/" xmlns:lib="http://purl.org/library/" xmlns:umbel="http://umbel.org/umbel#" xmlns:bibo="http://purl.org/ontology/bibo/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

	xmlns:skos="http://www.w3.org/2004/02/skos/core#"
Data and Resources	
Name	URL
Linked Data Service [site]	http://www.dnb.de/EN/lds
Dump [site]	http://datendienst.dnb.de/cgi-bin/mabit.pl?userID=opendata&pass=opendata&cmd=login
OAI [site]	http://www.dnb.de/oai
SRU [site]	http://www.dnb.de/sru
Example (RDF/XML) [datahub]	http://d-nb.info/gnd/118514768/about

5.1.11 *Hungarian National Library (NSZL)*

Name	Hungarian National Library (NSZL) catalog
Author	National Széchényi Library
Maintainer	National Széchényi Library
Link LOD source	http://nektar.oszk.hu/wiki/Semantic_web
Link datahub	http://datahub.io/dataset/hungarian-national-library-catalog

Description of Content	OPAC and Digital Library and the corresponding authority data as Linked Open Data.
	The used vocabularies are
	   * RDFDC for bibliographic data,
	   * FOAF for name authority entries, and
	   * SKOS for subject terms and geographical names.
	   * BIBO for bibliographic terms
	NSZL uses CoolURIs. Every resource has both RDF and HTML representation.
	RDFDC, FAOF and SKOS statements are linked together. The name authority is matched with the DBPedia name files, and URI aliases are handled as owl:sameAs statements. The name authority dataset also contains links to Virtual International Authority File (VIAF) and VIAF links back to NSZL's data.
	NSZL also supports the HTML link auto-discovery.

Licence	License Not Specified
Amount of resources	19,300,000 statements [site]
Protocols supported	Dump, SPARQL
Keywords Datahub	format-dc, format-rdf, linkeddata, lod, no-license-metadata, no-provenance-metadata, no-vocab-mappings, rdf, published-by-producer, publications, lld, no-proprietary-vocab, broken_link, format-xhtml, bibliographic, library, format-bibo, format-foaf, format-dbpedia, format-skos, sparql, catalog
Formats Datahub	api/sparql, example/rdf+xml
Namespaces	[N/A]
Data and Resources datahub	
Name	URL
SPARQL endpoint	http://setaria.oszk.hu/sparql
Example OPAC record (RDF/XML)	http://nektar.oszk.hu/resource/manifestation/2645471

Example digital library record (RDF/XML)	http://oszkdk.oszk.hu/resource/DRJ/404
Example name authority (RDF/XML)	http://nektar.oszk.hu/resource/auth/33589
Example subject term	http://nektar.oszk.hu/resource/auth/magyar_irodalom

5.1.12 *ICONCLASS - Multilingual Thematic Classification*

Name	ICONCLASS - Multilingual Thematic Classification
Author	[N/A]
Maintainer	[N/A]
Link LOD source	http://www.iconclass.org/help/lod
Link datahub	http://datahub.io/dataset/iconclass
Description of Content	From the website:
	This is an experimental service that makes the ICONCLASS Iconographic Classification system available as linked-data using the SKOS vocabulary. This service is inspired by the excellent Library of Congress Subject Headings linked data service. It is intentionally copied in spirit and conventions used. The idea is to enable others to make services that could make use of both underlying classification systems using the same toolings. And before you ask, yes I have been in contact with the creator of lcsh.info about this effort. More information about the Linked Data effort can be found on the W3C Wiki.
	Access/re-use
	The service is still experimental (as noted above) and the dataset is incomplete. The dataset can be downloaded as a zip file. There is currently no information about re-using the data.
	"There are about 1.3 million notations in ICONCLASS (with all keys and children full expanded). You can download a dump of the database in RDF N-Triple form from this link: http://iconclass.org/data/iconclass.20121010.ntriple.gz. The file is 19MB compressed and expands to 340MB. This datafile does not contain all the expanded notations including keys. The size would mushroom, so we are trying to figure out a better way to also make the expanded data available. Please feel free to send me suggestions."
Licence	http://www.opendefinition.org/licenses/odc-odbl
Amount of resources	1.300.000 notations
Protocols supported	RDF dump
Keywords Datahub	art, bibliographic, library, ontology, linked-data, semantic, classification, art-history, iconclass, digital-humanities
Formats Datahub	example/rdf+xml
Namespaces	[N/A]
Data and Resources datahub	
Name	URL
ICONCLASS RDF N-Triples	http://iconclass.org/data/iconclass.20121010.ntriple.gz

5.1.13 *Library of Congress- Subject Headings and Name Authority File*

Name	Library of Congress Subject Headings [LCSH] Library of Congress Name Authority File (NAF)
Author	[NAF:] Library of Congress
Maintainer	[LCSH:] Library of Congress
Link LOD source	http://id.loc.gov/authorities/subjects http://id.loc.gov/authorities/names
Link datahub	[LCSH:] http://datahub.io/dataset/lcsh [NAF:] http://datahub.io/dataset/library-of-congress-name-authority-file
Description of Content	[LCSH:] LCSH has been actively maintained since 1898 to catalog materials held at the Library of Congress. By virtue of cooperative cataloging other libraries around the United States also use LCSH to provide subject access to their collections. In addition LCSH is used internationally, often in translation. [NAF:] The Library of Congress Name Authority File (NAF) file provides authoritative data for names of persons, organizations, events, places, and titles. Its purpose is the identification of these entities and, through the use of such controlled vocabulary, to provide uniform access to bibliographic resources. Names descriptions also provide access to a controlled form of name through references from unused forms, e.g. a search under: Snodgrass, Quintus Curtius, 1835-1910 will lead users to the authoritative name for Mark Twain, which is, "Twain, Mark, 1835-1910." Names may also be used as subjects in bibliographic descriptions, so they may be combined with controlled values from subject heading schemes, such as LCSH. Library of Congress Names includes over 8 million descriptions created over many decades and according to different cataloging policies. LC Names is officially called the NACO Authority File and is a cooperative effort in which participants follow a common set of standards and guidelines. [LOD source]
Licence	Other (Public Domain)
Amount of resources	7,332,816 triples, 55,281 links (stitch-rameau) [SH] over 8 million descriptions [NAF]
Protocols supported	RDF dump
Keywords Datahub	[LCSH:] format-rdf, lod, no-license-metadata, no-provenance-metadata, no-vocab-mappings, published-by-producer, publications, lld, no-proprietary-vocab, bibliographic, library, format-skos, authorities [NAF:] lld, bibliographic, size.xlarge
Formats Datahub	application/x-ntriples, application/rdf+xml, example/rdf+xml [SH]; gzip:ntriples, XML [NAF]
Namespaces	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:identifiers="http://id.loc.gov/vocabulary/identifiers/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:ri="http://id.loc.gov/ontologies/RecordInfo#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"

	xmlns:cs="http://purl.org/vocab/changeset/schema#"
Data and Resources datahub	
Name	URL
LCSH SKOS (ntriples)	http://id.loc.gov/static/data/authoritiessubjects.nt.skos.zip
LCSH SKOS (rdf/xml)	http://id.loc.gov/static/data/authoritiessubjects.rdfxml.skos.zip
RDF/XML example	http://id.loc.gov/authorities/subjects/sh87001447.rdf
LC Name Authority File (MADS/RDF only) (size: 2.142GB)	http://id.loc.gov/static/data/authoritiesnames.nt.madsrdf.gz
LC Name Authority File (MADS/RDF only) (size: 2. 261GB)	http://id.loc.gov/static/data/authoritiesnames.rdfxml.madsrdf.gz
Additional Info	
Name	Value
links:stitch-rameau	55281
namespace	http://id.loc.gov/
publishingInstitution	http://lobid.org/organisation/US-DLC
shortname	LCSH
Triples	7332816

5.1.14 *ReLoad (Repository for Linked open archival data)*

Name	ReLoad project (Repository for Linked open archival data)
Author	[N/A]
Maintainer	Archivio Centrale dello Stato; Istituto Beni Culturali Regione Emilia Romagna; regesta.exe
Link LOD source	http://labs.regesta.com/progettoReload/en
Link datahub	[N/A]
Description of Content	The ReLoad project (Repository for Linked open archival data) will foster experimentation with the technology and methods of linked open data for archival resources. Its goal is the creation of a web of linked archival data.

LOD-LAM, which is an acronym for Linked Open Data for Libraries, Archives and Museums, is an umbrella term for the community and active projects in this area [http://lodlam.net/].

The first experimental phase will make use of W3C semantic web standards, mash-up techniques, software for linking and for defining the semantics of the data in the selected databases.

The archives that have made portions of their institutions' data and databases openly available for this project are the Central State Archive, and the Cultural Heritage Institute of Emilia Romagna Region. These will be used to test methodologies to expose the resources as linked open data. |
Licence	[N/A]
Amount of resources	189,710 datasets, 17,493,969 triples
Protocols supported	RDF, SPARQL

Keywords Datahub	[N/A]
Formats Datahub	[N/A]
Namespaces	dc: http://purl.org/dc/elements/1.1/ eac-cpf: http://archivi.ibc.regione.emilia-romagna.it/ontology/eac-cpf/ wgs84_pos: http://www.w3.org/2003/01/geo/wgs84_pos# geo: http://www.w3.org/2003/01/geo/wgs84_pos# foaf: http://xmlns.com/foaf/0.1/ void: http://rdfs.org/ns/void# viaf: http://viaf.org/ontology/1.1/# geonames: http://www.geonames.org/ontology# vcard: http://www.w3.org/2006/vcard/ns# dcterms: http://purl.org/dc/terms/ gn: http://www.geonames.org/ontology# rdfs: http://www.w3.org/2000/01/rdf-schema# bio: http://purl.org/vocab/bio/0.1/ ocsa: http://lod.xdams.org/reload/ocsa/ owl: http://www.w3.org/2002/07/owl# dbpedia: http://localhost:8080/resource/ xsd: http://www.w3.org/2001/XMLSchema# rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# skos: http://www.w3.org/2004/02/skos/core# oad: http://lod.xdams.org/reload/oad/
Data and Resources	
Name	URL
SPARQL endpoint	http://lod.xdams.org/sparql

5.1.15 *Swedish Open Cultural Heritage (K-samsök)*

Name	Swedish Open Cultural Heritage
Author	Börje Lewin
Maintainer	Börje Lewin
Link LOD source	http://www.ksamsok.se/in-english/api/
Link datahub	http://datahub.io/dataset/swedish-open-cultural-heritage
Description of Content	SOCH is a set of 3.4 million (as of december 2010) cultural heritage objects harvested from a large number of museums and other local, regional and national cultural heritage organizations. Different Creative Commons licenses are used for different datasets in SOCH. The main goal is to provide as open licenses as possible
Licence	Other (Open)
Amount of resources	3,400,000 triples (datahub)
Protocols supported	SOCH API
Keywords datahub	format-dc, lod, lodcloud.nolinks, linked-open-data, published-by-producer, openaccess, publications, cultural, culturalheritage, no-deref-vocab, cultural-institutions
Formats datahub	HTML, application/rdf+xml, example/rdf+xml
Namespaces	xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#"

	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ns5="http://kulturarvsdata.se/ksamsok#" xmlns:ns6="http://xmlns.com/foaf/0.1/#" xmlns:dcterms="http://purl.org/dc/terms/"
Data and Resources datahub	
Name	URL
Enterence page for browsing the data	http://kringla.nu
description for how to use the SOCH API	http://www.ksamsok.se/in-english/api/
Demo application showing url priciples more closely	http://www.ksamsok.se/tools/demo/loppan/loppan.html
rdf sample	http://kulturarvsdata.se/ksamsok/api?stylesheet=&x-api=test&method=search&hitsPerPage=50&query=item%3D%22Docka%22+and+create_toTime%3E%3D1700+and+create_fromTime%3C%3D1799
Example (RDF/XML)	http://kulturarvsdata.se/NOMU/object/NM0151866 [N/A]
Additional Info	
Name	Value
Version	1.0
Namespace	http://kulturarvsdata.se/
Triples	3.400.000

5.1.16 *VIAF: The Virtual International Authority File*

Name	VIAF: The Virtual International Authority File
Author	OCLC Online Computer Library Center, Inc.
Maintainer	[N/A]
Link LOD source	http://viaf.org/viaf/data/
Link datahub	http://datahub.io/dataset/viaf
Description of Content	VIAF (Virtual International Autority File) is an OCLC dataset and service -- built in cooperation with national libraries and other partners -- that virtually combines multiple LAM (Library Archives Museum) name authority files into a single name authority service. Put simply it is a large database of people and organizations that occur in library catalogs.
	VIAF is a joint project of 20 national libraries, implemented and hosted by OCLC. The project's goal is to lower the cost and increase the utility of library authority files by matching and linking the authority files of national libraries, and then making that information available on the Web.
	Openness
	The data is released under Open Data Commons Attribution license. Attribution is requested as follows:
	Adherence to ODC Attribution instructions for the correct assertion of attribution is encouraged. The preferred form of attribution for VIAF is:
	"This [title of report or article or dataset] contains information from VIAF (Virtual International Authority File) which is made

	available under the ODC Attribution License."
	Special cases: In circumstances where providing the full attribution statement above is not technically feasible, the use of canonical VIAF URIs is adequate to satisfy Section 4.3 of the ODC Attribution License.
Licence	http://www.opendefinition.org/licenses/odc-by
Amount of resources	[N/A]
Protocols supported	RDF
Keywords Datahub	format-rdf, lod, no-license-metadata, no-provenance-metadata, no-vocab-mappings, published-by-producer, publications, lld, format-owl, no-deref-vocab, bibliographic, library, format-skos
Formats Datahub	gz:txt, gz:xml, gz:iso, example/rdf+xml
Namespaces	xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:void="http://rdfs.org/ns/void#" xmlns:ns2="http://viaf.org/viaf/terms#"
Data and Resources datahub	
Name	URL
VIAF Links (txt)	http://viaf.org/viaf/data/viaf-20120422-links.txt.gz
VIAF Clusters (XML)	http://viaf.org/viaf/data/viaf-20120422-clusters.xml.gz
VIAF Clusters (RDF)	http://viaf.org/viaf/data/viaf-20120422-clusters-rdf.xml.gz
VIAF Clusters (MARC 21)	http://viaf.org/viaf/data/viaf-20120422-clusters-marc21.xml.gz
VIAF clusters (MARC 21 - ISO-2709)	http://viaf.org/viaf/data/viaf-20120422-clusters-marc21.iso.gz
VIAF redirections (RDF)	http://viaf.org/viaf/data/viaf-20120422-persist-rdf.xml.gz
Example (RDF/XML)	http://viaf.org/viaf/40280446/
Example (RDF/XML)	http://viaf.org/viaf/86518157
Additional Info	
Name	Value
Version	1
links:dbpedia	10.000
links:dnb-gemeinsame-normdatei	4.000.000
namespace	http://viaf.org/viaf/
shortname	VIAF
triples	200000000

5.1.17 *Victoria & Albert Museum Collection Records*

Name	Victoria & Albert Museum Collection Records
Author	Victoria & Albert Museum
Maintainer	N/A
Link LOD source	http://www.vam.ac.uk/api

Link datahub	http://datahub.io/dataset/vanda-api
Description of Content	A REST API to search the V&A's collection. Sadly restricted to academic and noncommercial use. Search for items by keyword, artist/maker, collection, year, material/technique, geography. Returns record and images plus thumbnails
Licence	Other (Non-Commercial)
Amount of resources	1,121,275 objects and 366,352 images [site]
Protocols supported	API (RESTful interface)
Keywords Datahub	access-api, museum, art, collection, images, license-noncommercial, art-history, victoria-and-albert
Formats Datahub	json or xml
Namespaces	N/A
Data and Resources datahub	
Name	URL
Document	http://museum-api.pbworks.com/Victoria-and-Albert-Museum-API-Docs

5.1.18 *Wikidata*

Name	Wikidata
Author	
Maintainer	
Link LOD source	http://www.wikidata.org/wiki/Wikidata:Database_download
Link datahub	N/A
Description of content	Wikidata is a free knowledge base that can be read and edited by humans and machines alike. It is for data what Wikimedia Commons is for media files: it centralizes access to and management of structured data, such as interwiki references and statistical information. Wikidata contains data in every language supported by the MediaWiki software.[from MainPage]
Licence	http://www.opendefinition.org/licenses/cc-zero
Amount of resources	13.415.817 items
Protocols supported	Dump
Keywords datahub	N/A
Formats datahub	N/A
Namespaces	N/A
Additional Info	
Name	Value
WikiDataQuery	http://208.80.153.172/wdq/ A service that offers the ability to run complex queries on items

5.2   LOD Source candidates from the datahub

Aiming at the identification of LOD source candidates for AthenaPlus, the Europeana LOD dataset was taken as starting point for the exploration of the datahub. The following table shows the basic information about the Europeana LOD dataset in the datahub.

Name	Europeana Linked Open Data
Author	Bernhard Haslhofer and Antoine Isaac
Link institution	http://data.europeana.eu/
Link datahub	http://datahub.io/de/dataset/europeana-lod
Description of content	data.europeana.eu currently contains open metadata on 2.4 million texts, images, videos and sounds gathered by Europeana. These objects come from data providers who have reacted early and positively to Europeana's initiative of promoting more open data and new data exchange agreements. These collections come from 8 direct Europeana providers encompassing over 200 cultural institutions from 15 countries. They cover a great variety of heritage objects, such as this slovenian version of O Sole Mio from the National Library of Slovenia, or Neil Robson's memories of the herring business from the Tyne and Wear Archives & Museums. Here is an example URI to get started with: http://data.europeana.eu/item/92037/25F9104787668C4B51 48BE8E5AB8DBEF5BE5FE03 The project has been described in a paper presented at the Dublin Core 2011 conference.
Licence	http://www.opendefinition.org/licenses/cc-zero
Groups datahub	Art, Bibliographic Data, Linking Open Data Cloud, Library Linked Data, #BiblioHack, Open Glam, EU Linked Data
Keywords datahub	culturalheritage, culture, eu, europe, europeana, format-dc, format-ore, format-rdf, library, license-metadata, lld, lod, metadata, no-deref-vocab, no-vocab-mappings, provenance-metadata, publications, published-by-producer, void-sparql-endpoint
Formats datahub	application/x-ntriples, meta/void, api/sparql, application/rdf+xml, meta/sitemap

The Europeana LOD dataset contains several items of groups, keywords and formats and is linked via these items with other datasets. From a view focused on cultural heritage content groups and keywords are of more importance than the formats. For a first approach the authors used group and keyword items to find other datasets as LOD source candidates from the datahub via browser interface.

Starting with group items one gets the following information about the groups used by Europeana LOD dataset from the datahub:

Group	Description	Datasets
Art	For data regarding artworks, collections, galleries or museums, state funding or involvement in the arts, art history, and aesthetics.	22
Bibliographic Data	This group comprises open bibliographic datasets according to the Principles on Open Bibliographic Data and a few not yet really open bibliographic datasets. It is maintained by members of the OKFN Working Group on Open Bibliographic Data.	90

Linking Open Data Cloud	This group catalogs data sets that are available on the Web as Linked Data and contain data links pointing at other Linked Data sets. The descriptions of the data sets in this group are used to generate the Linking Open Data Cloud diagram at regular intervals. The descriptions are also used generate the statistics provided in the State of the LOD Cloud document.	340
Library Linked Data	Group for Library Linked Data. Policy described at: http://esw.w3.org/TaskForces/CommunityProjects/LinkedLibraryData/Datasets/CKANmetainformation.	65
#BiblioHack	A collection of datasets for #Bibliohack 13th-14th June	10
Open GLAM	Datasets from GLAM (Galleries, Libraries, Archives and Museums) institutions that are open for (re-)use	29
EU Linked Data	Datasets from European Union institutions, published according to the Linked Data principles. Many of these datasets are re-published as Linked Data by third parties.	33

For further analysis in the context of AthenaPlus those groups are most interesting which provide cultural content, particularly relating to the museum sector. "Art" and "Open GLAM" are obviously such groups. The groups "Bibliographic Data" and "Library Linked Data" are also interesting for linking. The "Linking Open Data Cloud" group and the "EU Linked Data" group are considered too unspecific for the purposes of linking cultural heritage content. The "#BiblioHack" group is only historically interesting since focusing on one specific hackathon.

In the datahub group "Art" 21 datasets are available in addition to the Europeana LOD dataset. Only nine of the 21 datasets provide information about data formats. From these datasets only two provide a data format suitable for LD, the dataset of "ICONCLASS – Multilingual Thematic Classification" and the dataset of "Victoria & Albert Museum Collection Records". Both are already described in the list of LOD sources identified by the partners. So in the group "Art" no additional dataset was found.

In the datahub group "Open GLAM" 28 datasets are available in addition to the Europeana LOD dataset. 24 of the 28 datasets provide information about data formats and from these 18 datasets provide a data format which can be used for LD like "sparql", "rdf", "json" or "xml". 2 of them were duplicated ("British Museum", "British National Bibliography") and 2 of the rest of 16 datasets have been already listed before ("British Museum", "Swedish Open Cultural Heritage"). Among the remaining 14 records, nine datasets were outdated or out of the scope for linking with cultural heritage content, and five datasets were added as LOD source candidates for linking with AthenaPlus content.

Looking at the datahub groups "Bibliographic Data" and "Library Linked Data" it turns out that both mainly contain datasets - authority files, controlled vocabularies and bibliographic datasets - whose scope can be classified as local / regional, national, and international. Concerning bibliographic datasets it is argued here that it is more effective to connect with the respective datasets with national or international scope, e.g. VIAF. The relevant datasets for AthenaPlus content were listed already by the partners so no additional dataset is identified in the "Bibliographic Data" group . In the "Library Linked Data" group four additional datasets are identified as LOD source candidates.

In terms of keywords only those are interesting which suggest a cultural background. Those keywords are "culturalheritage", "culture" and "europeana". The keywords "library" and "lld" are also interesting but their scope is too large in our context. 14 additional datasets which are not mentioned before share one or more of the keywords "culturalheritage", "culture", "europeana" with the Europeana LOD dataset. None of these 16 datasets has been added to the selection, either because they do not support content and formats which can be used for LOD or because they are out-of-date and no activity stream is reported. With the datasets from the groups "Open GLAM" and "Library Linked Data" in total the following nine datasets remain as LOD source candidates for AthenaPlus content linking.

5.2.1 *Amsterdam Museum as Linked Open Data in the Europeana Data Model*

Name	Amsterdam Museum as Linked Open Data in the Europeana

	Data Model
Author	Victor de Boer, Jan Wielemaker, Jacco van Ossenbruggen, Antoine Isaac, Guus Schreiber
Maintainer	Victor de Boer
Link LOD source	http://semanticweb.cs.vu.nl/lod/am/
Link datahub	http://datahub.io/dataset/amsterdam-museum-as-edm-lod
Description of Content	The Amsterdam Museum dataset describes more than 70.000 cultural heritage objects related to the city of Amsterdam described by the museum. The metadata was retrieved from an XML Web API of the museum's Adlib collection database and converted to RDF compliant with the Europeana Data Model (EDM). This makes the Amsterdam Museum data the first of its kind to be officially converted and made available in this format.
Licence	http://www.opendefinition.org/licenses/cc-by-sa
Amount of resources	5,000,000 triples
Protocols supported	Dump, SPARQL
Keywords Datahub	deref-vocab, lod, no-license-metadata, no-provenance-metadata, published-by-third-party, rdf, publications, amsterdam, country-netherlands, crossdomain, cultural, culturalheritage, datagovuk, edm, europeana, museum
Formats Datahub	api/sparql, HTML, -, api/git, example/rdf+xml
Namespaces	N/A
Data and Resources datahub	
Name	URL
SPARQL endpoint	http://semanticweb.cs.vu.nl/europeana/sparql
Public Git repository with RDF (browser version)	http://eculture.cs.vu.nl/git/public/?p=econnect/metadata/AHM.git;a=tree;f=rdf
Public Git repository (use this path for git clone)	http://git://eculture.cs.vu.nl/home/git/econnect/metadata/AHM.git
example: Local view for object "Transom"	http://purl.org/collections/nl/am/proxy-63432
example: Local view for object "Conmemoration plate"	http://purl.org/collections/nl/am/proxy-23182
SPARQL endpoint UI (in HTML form)	http://semanticweb.cs.vu.nl/europeana/user/query
Download	http://eculture.cs.vu.nl/git/public/?p=econnect/metadata/AHM.git;a=tree;f=rdf
Additional Info	
Name	Value
Version	1.0
links:dbpedia	43
links:geonames-semantic-web	658
namespace	http://purl.org/collections/nl/am/
shortname	Amsterdam Museum

triples	5000000
vocab-mappings	skos:exactMatch

5.2.2 *Bibliothèque nationale de France (BNF) - Linked Open Data*

Name	data.bnf.fr - Bibliothèque nationale de France
Author	data.bnf.fr project team - French national Library
Maintainer	data.bnf.fr project team
Link LOD source	http://data.bnf.fr/ [datahub] http://data.bnf.fr/semanticweb-en [site]
Link datahub	http://datahub.io/dataset/data-bnf-fr
Description of Content	data.bnf.fr gathers data from the different databases of the Bibliothèque nationale de France, so as to create Web pages about Works and Authors, together with a RDF view on the extracted data. There are about 2.000.000 RDF triples. There are links to id.loc.gov for languages and nationalities, to dewey.info for subjects, and to DCMI type for types. We use SKOS, FOAF, DC and RDA vocabularies, in a FRBR model.
Licence	Other (Attribution)
Amount of resources	6,330,000 triples; links: 69,000 (lcsh), 5,000 (ddc), 9,383 (viaf), 4,312 (dbpedia)
Protocols supported	Dump [datahub], RESTful API [site]
Keywords Datahub	format-dc, lod, published-by-producer, publications, lld, no-deref-vocab, format-foaf, format-skos
Formats Datahub	example/rdf+xml, example/ntriples, application/rdf xml
Namespaces	xmlns:bnf-onto="http://data.bnf.fr/ontology/bnf-onto/" xmlns:dc="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:rdarelationships="http://rdvocab.info/RDARelationships WEMI/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
Data and Resources datahub	
Name	URL
Example resource	http://data.bnf.fr/12417486/alexandre_dumas_la_reine_margot/rdf.xml
Example resource	http://data.bnf.fr/12417486/alexandre_dumas_la_reine_margot/rdf.nt
Full dump of data.bnf.fr	http://echanges.bnf.fr/PIVOT/databnf_all_rdf_xml_2012-11-15.tar.gz?user=databnf&password=databnf
Additional Info	
Name	Value
Version	2
http://dbpedia.org	4312
http://viaf.org/	9383

links:dewey_decimal_classification	5000
links:lcsh	69000
namespace	http://data.bnf.fr/
publishingInstitution	http://lobid.org/organisation/FrPBN
triples	6330000

5.2.3 *British National Bibliography (BNB) - Linked Open Data*

Name	British National Bibliography (BNB) - Linked Open Data
Author	The British Library Metadata Services
Maintainer	The British Library Metadata Services
Link LOD source	http://bnb.data.bl.uk/
Link datahub	http://datahub.io/dataset/bluk-bnb
Description of Content	British National Bibliography (BNB) published as Linked Data by the British Library, linked to external sources including VIAF, LCSH, Lexvo, GeoNames, MARC country, and language, Dewey.info, RDF Book Mashup. Published to this data model for books and this data model for serials.
Licence	http://www.opendefinition.org/licenses/cc-zero
Amount of resources	89,733,617 triples, 132,391 links lcsh, 889,690 links viaf
Protocols supported	Dump, SPARQL
Keywords Datahub	format-rdf, lod, published-by-producer, publications, lld, no-proprietary-vocab, crossdomain, format-xml, format-xhtml, bibliographic, library, british-library, cc0
Formats Datahub	api/sparql, api/describe, api/search, example/rdf+xml, example/rdf+ttl, example/rdf+json, HTML, example, XML, html | pdf, application/x-zip-compressed, nt/zip-compressed, nt, ttl
Namespaces	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dct="http://purl.org/dc/terms/" xmlns:bibo="http://purl.org/ontology/bibo/" xmlns:blterms="http://www.bl.uk/schemas/bibliographic/blterms#" xmlns:elements="http://iflastandards.info/ns/isbd/elements/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:linked-data="http://purl.org/linked-data/api/vocab#" xmlns:void="http://rdfs.org/ns/void#"
Data and Resources datahub	
Name	URL
SPARQL Endpoint	http://bnb.data.bl.uk/sparql
Search Service	http://bnb.data.bl.uk/search
Example resource - RDF	http://bnb.data.bl.uk/doc/resource/006893251.rdf
Example resource - Turtle	http://bnb.data.bl.uk/doc/resource/006893251.ttl
Example resource - JSON	http://bnb.data.bl.uk/doc/resource/006893251.json
Example resource - html	http://bnb.data.bl.uk/doc/resource/006893251.html

Example resource RDF - ttl/json/html - default html	http://bnb.data.bl.uk/doc/resource/006893251
British Library Terms RDF Schema	http://www.bl.uk/schemas/bibliographic/blterms
Current Information on releases and documentation	http://www.bl.uk/bibliographic/datafree.html
Data model (Book)	http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf
Data model (Serial)	http://www.bl.uk/bibliographic/pdfs/bldatamodelserial.pdf
BNB Books dataset	http://BNB Books dataset
SPARQL Editor	http://bnb.data.bl.uk/flint
VoID Descriptions for LOD BNB	http://www.bl.uk/bibliographic/downloads/Void_201308_ttl.zip
Bulk downloads	http://www.bl.uk/bibliographic/download.html#lodbnb
Additional Info	
Name	Value
links:ddc	139472
links:lcsh	132391
links:viaf	889690
namespace	http://bnb.data.bl.uk/id/
shortname	BNB
Triples	89733617

5.2.4 Deutsche Nationalbibliografie (DNB)

Name	Deutsche Nationalbibliografie (DNB)
Author	Deutsche Nationalbibliothek (German National Library)
Maintainer	Deutsche Nationalbibliothek (German National Library) [site]
Link LOD source	http://www.dnb.de/EN/datendienste/linkedData [datahub] http://www.dnb.de/EN/nationalbibliografie [site]
Link datahub	http://datahub.io/dataset/deutsche-nationalbibliografie-dnb
Description of Content	The Linked Data Service of the German National Library (Deutsche Nationalbibliothek, DNB) has expanded and includes bibliographic data since January 2012. As a first step, the bibliographic data of the DNB's main collection (apart from the printed music and the collection of the Deutsches Exilarchiv) and the serials (magazines, newspapers and series of the German Union Catalogue of serials (ZDB)) have been converted.
Licence	http://www.opendefinition.org/licenses/cc-zero
Amount of resources	N/A
Protocols supported	Dump, OAI-PMH
Keywords Datahub	lod, rdf, lld, bibliographic, germany, bibliographic data, cc0
Formats Datahub	rdf/turtle, RDF
Namespaces	xmlns:gnd="http://d-nb.info/standards/elementset/gnd#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rda="http://rdvocab.info/" xmlns:foaf="http://xmlns.com/foaf/0.1/"

	xmlns:isbd="http://iflastandards.info/ns/isbd/elements/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:marcRole="http://id.loc.gov/vocabulary/relators/" xmlns:lib="http://purl.org/library/" xmlns:umbel="http://umbel.org/umbel#" xmlns:bibo="http://purl.org/ontology/bibo/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
Data and Resources datahub	
Name	URL
Deutsche Nationalbibliografie (DNB)	http://datendienst.dnb.de/cgi-bin/mabit.pl?cmd=fetch&userID=opendata&pass=opendata&mabheft=DNBTitel.rdf.gz
Example periodical	http://d-nb.info/944912621/about/rdf
Example book	http://d-nb.info/994738471/about/rdf
Example collection	http://d-nb.info/1015411169/about/rdf
Example article	http://d-nb.info/1014234816/about/rdf
Example series	http://d-nb.info/011078677/about/rdf
Additional Info	
Name	Value
Version	2013-04-26
links:culturegraph	10659659
links:doi	127028
links:gnd	25124413
links:handle	2029
links:iso639-2	3108164
links:urn	858098
links:zdb	786791
triples	167642465

5.2.5 *Dewey Decimal Classification (DDC)*

Name	Dewey Decimal Classification (DDC)
Author	OCLC Online Computer Library Center, Inc.
Maintainer	Michael Panzer
Link LOD source	http://dewey.info/
Link datahub	http://datahub.io/dataset/dewey_decimal_classification
Description of Content	Dewey.info is an experimental space for linked DDC data. The intention of the dewey.info prototype is to be a platform for Dewey data on the Web. Included as linked data are the DDC Summaries (the top three levels of the DDC) of Edition 22 in 11 languages and all assignable numbers of Abridged Edition 14 in three languages. Classification semantics are encoded in RDF using SKOS and other element vocabularies.

	Every class also has an HTML representation (XHTML+RDFa) and several RDF serializations (RDF/XML, Turtle, JSON).
Licence	http://creativecommons.org/licenses/by-nc/2.0/
Amount of resources	402,288 triples
Protocols supported	Dump, SPARQL
Keywords Datahub	format-rdf, linkeddata, lod, lodcloud.nolinks, published-by-producer, publications, lld, schemata, license-metadata, no-proprietary-vocab, classification, publication
Formats Datahub	api/sparql, example/rdf+xml, example/x-turtle, example/xhtml+xml
Namespaces	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:xhv="http://www.w3.org/1999/xhtml/vocab#" xmlns:cc="http://creativecommons.org/ns#" xmlns:dct="http://purl.org/dc/terms/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
Data and Resources datahub	
Name	URL
Dewey Decimal Classification (DDC)	http://dewey.info/sparql.php
	http://dewey.info/class/641/about.rdf
	http://dewey.info/class/641.5/a14/about.ttl
	http://dewey.info/class/6/2009/03/about.de.html
Additional Info	
Name	Value
license_link	http://creativecommons.org/licenses/by-nc-nd/3.0/
namespace	http://dewey.info/
shortname	DDC
triples	402288

5.2.6 *Freebase*

Name	Freebase
Author	Google
Maintainer	Shawn Simister
Link LOD source	http://freebase.com/ [datahub] https://developers.google.com/freebase/ [site]
Link datahub	http://datahub.io/dataset/freebase
Description of Content	"Freebase is an open database of the world's information. It is built by the community and for the community—free for anyone to query, contribute to, built applications on top of, or integrate into their websites." […] Triple count and link statistics provided by Freebase contributor Tom Morris […] Freebase has an RDF service that exposes URIs and generates RDF descriptions for all Freebase topics.
Licence	http://www.opendefinition.org/licenses/cc-by
Amount of resources	337,203,427 triples; links: 3,348,530 (dbpedia), 350,110 (bbc-music), 120,626 (sec-rdfabout), 100,000 (geospecies),

	9,930 (nytimes-linked-open-data)
Protocols supported	Dump, MQL
Keywords Datahub	deref-vocab, format-rdf, lod, no-vocab-mappings, rdf, published-by-producer, provenance-metadata, lld, access-www, license-metadata, access-bulk, broken_link, access-api, crossdomain, size-large, format-dcterms, encylopaedia
Formats Datahub	tsv, api/search, example/rdf+xml
Namespaces	key: <http://rdf.freebase.com/key/>. ns: <http://rdf.freebase.com/ns/>. owl: <http://www.w3.org/2002/07/owl#>. rdfs: <http://www.w3.org/2000/01/rdf-schema#>. xsd: <http://www.w3.org/2001/XMLSchema#>.
Data and Resources datahub	
Name	URL
Directory containing all downloads	http://download.freebase.com/datadumps/
A full dump of Freebase assertions as tab separated utf8 text size:3.9GB	http://download.freebase.com/datadumps/latest/freebase-datadump-quadruples.tsv.bz2
A tab-separated file for each type in Freebase size:1.3GB	http://download.freebase.com/datadumps/latest/freebase-datadump-tsv.tar.bz2
Directory of browseable subsets by domain	http://download.freebase.com/datadumps/latest/browse/
Topic dump with basic info for each topic Size:1.0 GB	http://download.freebase.com/datadumps/latest/freebase-simple-topic-dump.tsv.bz2
Search api for freebase	http://freebase.com/api/service/search
Example RDF resource	http://rdf.freebase.com/ns/en.blade_runner
Additional Info	
Name	Value
links:bbc-music	350110
links:dbpedia	3348530
links:geospecies	100000
links:nytimes-linked-open-data	9930
links:sec-rdfabout	120626
namespace	http://rdf.freebase.com/
triples	337203427

5.2.7 Muninn World War I

Name	Muninn World War I
Author	Rob Warren
Maintainer	Rob Warren
Link LOD source	http://rdf.muninn-project.org [site]
Link datahub	http://datahub.io/dataset/muninn-world-war-i
Description of Content	The Muninn Project is a multi-disciplinary, multi-national, academic research project investigating millions of records

	pertaining to the First World War in archives around the world.
Licence	http://www.opendefinition.org/licenses/cc-by
Amount of resources	
Protocols supported	SPARQL
Keywords Datahub	deref-vocab, format-rdf, lod, rdf, published-by-producer, provenance-metadata, vocab-mappings, format-rdfs, access-www, format-owl, license-metadata, access-bulk, access-api, crossdomain, format-foaf, format-skos, void-sparql-endpoint, search
Formats Datahub	api/sparql, compressed tarfile containing n-triples, meta/rdf-schema, mapping/owl, example/rdf+xml, application/x-nquads, meta/sitemap, meta/void
Namespaces	N/A
Data and Resources datahub	
Name	URL
SPARQL endpoint	http://rdf.muninn-project.org/sparql
Download page (N-Triples, bz2-compressed)	http://rdf.muninn-project.org/Downloads
Mappings	http://rdf.muninn-project.org/Downloads/muninn-ontologies-latest.tar.bz2
Link to an example Linked Data entry	http://rdf.muninn-project.org/modern/2011/11/11/Cenotaph/ca12581df8bc9fc3c1535de9daa1814a
Download	http://rdf.muninn-project.org/Downloads/Muninn-Dump-2013-01-15.nq
Semantic Web Sitemap	http://rdf.muninn-project.org/sitemap.xml
voiD description	http://rdf.muninn-project.org/Muninn.rdf
Additional Info	
Name	Value
Version	2013-01-15
Note	Project data is added on an ongoing basis
license_link	http://www.opendatacommons.org/licenses/by/
links:dbpedia	691
links:geonames-semantic-web	31968
links:linkedgeodata	10
namespace	http://rdf.muninn-project.org/ww1
shortname	Muninn
sparql_graph_name	http://rdf.muninn-project.org/
triples	1000000

5.2.8 *Public Library of Veroia*

Name	Public Library of Veroia

Author	Sotiris Karampatakis
Maintainer	Sotiris Karampatakis
Link LOD source	http://libver.math.auth.gr/
Link datahub	http://datahub.io/dataset/libver
Description of Content	Public Library of Veroia Linked Open Data Project. This is the first effort of a Greek Library to expose it's Bibliographic Data as Linked Open Data.
Licence	http://www.opendefinition.org/licenses/cc-by-sa
Amount of resources	N/A
Protocols supported	SPARQL
Keywords Datahub	format-rdf, lod, no-license-metadata, no-provenance-metadata, published-by-third-party, publications, format-rdfs, lld, no-proprietary-vocab, bibliographic, media
Formats Datahub	api/sparql, example/rdf+xml, api/search, HTML, application/rdf+xml, meta/void
Namespaces	N/A
Data and Resources datahub	
Name	URL
SPARQL Endpoint	http://libver.math.auth.gr/sparql
Link to an example data item within the dataset (RDF/XML)	http://libver.math.auth.gr/data/agent3911.rdf
Virtuoso Faceted Search	http://libver.math.auth.gr/fct
Link to an example data item within the dataset (RDFa)	http://libver.math.auth.gr/page/agent3911
2012-09-06T043619/record1_10.rdf	https://commondatastorage.googleapis.com/ckannet-storage/2012-09-06T043619/record1_10.rdf
2012-09-25T143058/void.ttl	https://commondatastorage.googleapis.com/ckannet-storage/2012-09-25T143058/void.ttl
Additional Info	
Name	Value
Version	1.0beta
links:dbpedia	2915
links:dbpedia-el	4197
namespace	http://libver.math.auth.gr/resource
triples	1285417

5.2.9 *RAMEAU subject headings (STITCH)*

Name	RAMEAU subject headings (STITCH)
Author	Antoine Isaac <http://www.few.vu.nl/~aisaac>, Rameau committee <http://rameau.bnf.fr>
Maintainer	N/A
Link LOD source	http://www.cs.vu.nl/STITCH/rameau/

Link datahub	http://datahub.io/dataset/stitch-rameau
Description of Content	Data exposed: SKOS representation of the RAMEAU book indexing vocabulary, maintained by the French National Library (BnF) Size of dump and data set: 130 MB uncompressed Notes: aligned to LCSH and SWD via SKOS mapping links Published as SKOS/RDF by the STITCH project. As of February 2012, URIs of the RAMEAU concepts redirect to equivalent descriptions at data.bnf.fr, see http://data.bnf.fr/semanticweb-en
Licence	License Not Specified
Amount of resources	1,619,918 triples; links: 60,000 (lcsh), 20,000 (gnd), 700 (agrovoc-skos)
Protocols supported	Dump
Keywords Datahub	format-rdf, linkeddata, lod, no-license-metadata, no-provenance-metadata, no-vocab-mappings, published-by-third-party, rdf, ckanupload.esw.200910, publications, lld, no-deref-vocab, library, format-skos, vu, authorities, decommissioned
Formats Datahub	application/rdf+xml, example/rdf+xml
Namespaces	N/A
Data and Resources datahub	
Name	URL
Subject headings (SKOS)	http://www.cs.vu.nl/STITCH/rameau/dump/Rameau.skos.gz
LCSH mappings (SKOS)	http://www.cs.vu.nl/STITCH/rameau/dump/MACSMapping-Rameau_LCSH.skos.gz
SWD mappings (SKOS)	http://www.cs.vu.nl/STITCH/rameau/dump/MACSMappings-Rameau_SWD.rdf.zip
Example resource	http://stitch.cs.vu.nl/vocabularies/rameau/autorites_matieres [N/A]
Dump at data.bnf.fr	http://echanges.bnf.fr/PIVOT/databnf_all_rdf_xml.tar.gz?user=databnf&password=databnf [N/A]
Additional Info	
Name	Value
links:agrovoc-skos	700
links:dnb-gemeinsame-normdatei	20000
links:lcsh	60000
namespace	http://stitch.cs.vu.nl/vocabularies/rameau/
triples	1619918

5.3   LOD Sources at a glance

The following table provides an overview of the LOD sources which are described in detail in the previous two sections. The table can be used for a comparison of the LOD sources with regard to its support of the specifications from the description of "good" LOD sources from section 2.1. However, the criteria therein should be applied not too rigid. The properties of LOD sources may change. Some sources of considerable importance are not yet available under an open license or in a suitable technical format. The most relevant example are the Getty thesauri, their release as LOD is already scheduled so that the AthenaPlus project can likely benefit. [49]. The partner survey has shown that reliability, quality, and sustainability of the data and data suppliers are as important as technical and quantifiable requirements.

No	Name	Link	Description	Licence	Resources	Protocols
1	Amsterdam Museum	http://semanticweb.cs.vu.nl/lod/am/	The Amsterdam Museum dataset in Eurpeana Data Model RDF	CC-BY-SA	5,000,000 triples	Dump SPARQL
2	BNB (British National Bibliography)	http://datahub.io/dataset/bluk-bnb	British National Bibliography (BNB) published as Linked Data by the British Library	CC0 1.0	approximately 2.8 million descriptions (93,583,853 triples) of books and serials	Dump SPARQL
3	BNE (Spanish National Library)	http://datos.bne.es/	Open bibliographic linked data from the Spanish National Library	CC0 (1.0)	58,053,215 triples 4 Million authority records 2.4 million bibliographic records	DUMP, SPARQL
4	BNF (Bibliothèque nationale de France)	http://data.bnf.fr/semanticweb-en	Open bibliographic linked data from the French National Library	Various (http://data.bnf.fr/licence)	6,330,000 triples; links: 69,000 (lcsh), 5,000 (ddc), 9,383 (viaf), 4,312 (dbpedia)	DUMP, RESTful API
5	British Museum Collection	http://collection.britishmuseum.org	Linked Data access to the collection data of the British Museum's Online Collection	British Museum's Open Data Licence 1.0	105,635,648 triples	SPARQL
6	CLAROS	http://data.clarosnet.org/	CLassical Art Research Online. Service of the universities of Oxford, Cologne, and Paris, for the art of ancient Greece and Rome.	ODbL (?)	11,649,742 elements	SPARQL

---

[49] http://www.getty.edu/research/tools/vocabularies/lod/#status

No	Name	Link	Description	Licence	Resources	Protocols
7	Cultura italia	http://dati.culturaitalia.it/	Portal of Ministero per i Beni e le Attività Culturali. Metadata from museums and other local, regional and national cultural heritage organizations.	CC0 (1.0)	Statements: 15,048,155 Entities: 5,277,737	OAI-PMH, SPARQL
8	DBpedia	http://dbpedia.org/	Community effort to extract structured information from Wikipedia to make this information available on the Web.	CC-BY-SA 3.0 GNU Free Documentation License	about 4.0 million things, whereof 3.22 million are classified (832,000 persons, 639,000 places, 372,000 creative works, 209,000 organizations, 226,000 species)	SPARQL
9	Dewey Decimal Classification (DDC)	http://dewey.info/	DDC Summaries (the top three levels of the DDC) of Edition 22 in 11 languages and all assignable numbers of Abridged Edition 14 in three languages.	CC-BY-NC 2.0	402,288 triples	Dump, SPARQL
10	DNB (Deutsche Nationalbibliografie)	http://www.dnb.de/EN/nationalbibliografie	The Linked Data Service of the German National Library (Deutsche Nationalbibliothek, DNB)	CC0 1.0	167,642,465 triples 10,659,659 links culturegraph 25,124,413 links gnd	Dump
11	Freebase	http://freebase.com/	Freebase is a big collection of structured data (knowledge graph), and a Freebase platform for accessing and manipulating that data via the Freebase API.	CC-BY 2.5	about 1.2 billion triples, about 40 million entities (26m Music, 6m Media, 4m Books, 3m People ...)	Dump, API
12	Gencat	http://data.clarosnet.org/	Portal of Government of Catalonia. website which publishes public data. from different bodies of the Government.	Various	N/A	Dump
13	GeoNames	http://www.geonames.org/	Geographical database which covers all countries and contains over eight million placenames	CC BY 3.0	over 8.3 million toponyms	Dump RESTful web service

No	Name	Link	Description	Licence	Resources	Protocols
14	Getty AAT	http://www.getty.edu/research/tools/vocabularies/aat	Structured vocabulary to describe art, architecture, decorative arts, material culture, and archival materials	scheduled for release as LOD mid-January 2014	around 268,650 terms and other information about around 51,470 concepts	N/A
15	Getty TGN	http://www.getty.edu/research/tools/vocabularies/tgn/	Structured vocabulary containing names and other information about places	scheduled for release as LOD in 2014	around 2,035,195 names and other information about around 1,431,380 places	N/A
16	GND	http://www.dnb.de/EN/gnd	Integrated Authority File (GND) of German National Library (Deutsche Nationalbibliothek). Contains data records representing on persons, corporate bodies, congresses, geographic entities	CC0 1.0	100,720,519 triples	Dump SRU
17	ICONCLASS	http://www.iconclass.org/help/lod	Multilingual classification system for cultural content	ODbL 1.0	about 1.3 million notations	RDF dump, HTTP content negotiation
18	LOC Subject Headings	http://id.loc.gov	Authority Files provide authoritative data for subject headings (LCSH) and for names (LCNAF) of persons, organizations, events, places, and titles.		over 8 million descriptions (NAF)	Dump
19	Muninn World War I	http://rdf.muninn-project.org/	The Muninn Project is a multi-disciplinary, multi-national, academic research project investigating millions of records pertaining to the First World War in archives around the world.	ODC-By 1.0	1,000,000 triples 31,968 links geonames 691 links dbpedia	SPARQL
20	NSZL	http://nektar.oszk.hu/wiki/Semantic_web	Authority Files of the Hungarian National Library (National Széchényi Library)	N/A	19,300,000 statements	Dump SPARQL
21	Public Library of Veroia	http://libver.math.auth.gr/	Bibliographic Data of the Public Library of Veroia Linked Open Data Project	CC-BY-SA	1,285,417 triples 2,915 links dbpedia 4,197 links dbpedia-el	SPARQL

No	Name	Link	Description	Licence	Resources	Protocols
22	RAMEAU subject headings (STITCH)	http://www.cs.vu.nl/STITCH/rameau/	SKOS representation of the RAMEAU book indexing vocabulary, maintained by the French National Library (BnF)	N/A	1,619,918 triples; links: 60,000 (lcsh), 20,000 (gnd), 700 (agrovoc-skos)	Dump
23	ReLoad / LODLAM	http://labs.regesta.com/progettoReload/	Italian Repository for Linked Open Archival Data (Reload)	N/A	17,493,969 triples 189,710 datasets	SPARQL
24	SOCH (K-samsök)	http://www.ksamsok.se/in-english/api/	Portal of Swedish National Heritage Board. Objects harvested from a large number of museums and other local, regional and national cultural heritage organizations.	Various, Other (Open)	3,400,000 triples	SOCH API
25	VIAF	http://viaf.org/viaf/data/	OCLC dataset and service - built in cooperation with 20 national libraries and other partners - that virtually combines multiple LAM name authority files into a single name authority service	ODC-By 1.0	200,000,000 triples	Dump
26	V&A	http://www.vam.ac.uk/api	RESTful interface to the collections of the Victoria and Albert Museum	N/A	1,121,275 objects 366,352 images	RESTful API
27	Wikidata	http://www.wikidata.org	Database project of the Wikimedia Foundation to provide support for Wikipedia, Wikimedia projects, and others.	CC0 1.0	13,422,057 pages	Dump

6 CONCLUSIONS

The analysis of LOD sources carried out in this deliverable lead to 27 possible data sources as candidates for linking with AthenaPlus content. 18 of these data sources were suggested by partners in the survey, additional nine sources were found through exploration of the datahub.

The selection of candidates has been based on the recommendations and criteria of the W3C, in particular on the five star rating system for LOD publishing. The starting point to include only data sources that support a SPARQL endpoint and that are published with a CC0 license could not be maintained: If these two criteria were strictly applied, a considerable number of candidates would have been excluded at an early point of the project though further developments in the field are expected in the course of the project (e.g. the release of the Getty thesauri as LOD[50]). Therefore the criteria were applied softly, the state-of-the-art of LOD publishing is in progress.

In terms of the five star rating system the LOD sources identified as candidates will mostly range from three to five stars at the time of writing this deliverable. In this context it is worthwhile to mention that protocols such as REST may turn out to be suitable for linking processes while e.g. not fully complying with the fourth star "standards from W3C (RDF and SPARQL)". In addition to this, the release of SPARQL 1.1[51] in March 2013 by the W3C is expected to play an important role in the linking procedure, as the available data source will evolve to support it. The SPARQL 1.1 is more expressive than SPARQL supporting more operators for text manipulation as well as federated queries, functionalities that can prove very useful for linking.

With regard to the evolving LOD sector, the final selection of the sources for linking with AthenaPlus content within the work on D4.6 will require a revisit and further evaluation of the sources presented in this deliverable. Technical evaluation of datasets which are in the datahub can be done with several tools suitable for this purpose: The uptime of SPARQL endpoints is measured by Mondeca[52]. A statistic of the datahub is available in LODStats[53]. The DataHub LOD validator[54] provides an overview of the completeness of LOD sources.

Subsequent analysis within D4.6 on actual linking opportunities from the content point of view will be based on the underlying terminologies as well as the prerequisites for the envisaged linking processes. Since quality criteria turn out to be highly important for the partners the criteria for linking resources will need to be carefully developed.

In general, it can be expected that establishing links will be the more reliable the closer the linking process is tied to the source data in terms of its semantics. It is therefore a highly recommended strategy to include links, e.g. URIs for resources in LOD datasets, already in the actual metadata production phase. So partners may use the list of LOD source candidates presented in this deliverable as an inspiration for their own metadata production process and figure out opportunities to include such links from the outset.

---

[50] The target date for the release of AAT as LOD is mid-January 2014 (http://www.getty.edu/research/tools/vocabularies/lod/)
[51] http://www.w3.org/TR/sparql11-query/
[52] http://labs.mondeca.com/sparqlEndpointsStatus/
[53] http://stats.lod2.eu/ and http://aksw.org/Projects/LODStats.html
[54] http://validator.lod-cloud.net/

APPENDIX 1: REFERENCES

Books

Linked Data: Evolving the Web into a Global Data Space

http://linkeddatabook.com
Tom Heath and Christian Bizer (2011). Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. DOI: 10.2200/S00334ED1V01Y201102WBE001. ISBN: 9781608454303 (paperback). ISBN: 9781608454310 (ebook). Copyright © 2011 by Morgan & Claypool. All rights reserved.
This book gives an overview of the principles of Linked Data as well as the Web of Data that has emerged through the application of these principles. The book discusses patterns for publishing Linked Data, describes deployed Linked Data applications and examines their architecture.

Linked Data Search Engines

http://linkeddatabook.com/editions/1.0/#sec:searchEngines
A number of search engines have been developed that crawl Linked Data from the Web by following RDF links, and provide query capabilities over aggregated data.

Publishing and Using Cultural Heritage Linked Data on the Semantic Web

http://dx.doi.org/10.2200/S00452ED1V01Y201210WBE003
Eero Hyvönen (2012). Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Synthesis Lectures on the Semantic Web: Theory and Technology. Lecture #3. Morgan & Claypool. DOI: 10.2200/S00452ED1V01Y201210WBE003. ISBN: 9781608459971 (paperback). ISBN: 9781608459988 (ebook). Copyright © 2012 by Morgan & Claypool. All rights reserved.
This book gives an overview on why, when, and how Linked (Open) Data and Semantic Web technologies can be employed in practice in publishing CH collections and other contents on the
Web. This book is targeted to computer scientists,museum curators, librarians, archivists, and other
CH professionals interested in Linked Data and CH applications on the Semantic Web.

W3C Sites

W3C SWEO Linking Open Data community project

http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.

Semantic Web Search Engines

http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines

This page collects links to Semantic Web Search Engines. Semantic Web Search Engines use robots to crawl RDF data from the Web and provide search and navigation facilities over crawled data. It is part of the W3C SWEO Linking Open Data community project.

Government Linked Data (GLD) Working Group

The mission of the Government Linked Data (GLD) Working Group is to provide standards and other information which help governments around the world publish their data as effective and usable Linked Data using Semantic Web technologies.

Best Practices: Vocabulary Selection

http://www.w3.org/2011/gld/wiki/222_Best_Practices_for_Vocab_Selection
The group will provide advice on how governments should select RDF vocabulary terms (URIs), including advice as to when they should mint their own. This advice will take into account issues of stability, security, and long-term maintenance commitment, as well as other factors that may arise during the group's work.

Linked Data Cookbook

http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

The approach in writing this document has been to collate and present the most relevant engineering practices prevalent in the Linked Data development community today and identify those that: a) facilitate the exploitation of Linked Data to enable better search, access and re-use of open government information; or b) are considered harmful and can have non-obvious detrimental effects on the overall quality of data publishing on the Web.

W3C Library Linked Data Incubator Group

http://www.w3.org/2005/Incubator/lld/

The mission of the W3C Library Linked Data Incubator Group, chartered from May 2010 through August 2011, has been "to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities — focusing on Linked Data — in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future."

VoID Vocabulary

http://www.w3.org/TR/void/

VoID is an RDF Schema vocabulary for expressing metadata about RDF datasets. It is intended as a bridge between the publishers and users of RDF data, with applications ranging from data discovery to cataloging and archiving of datasets.

Open Data Directory

http://dir.w3.org/directory/pages/landing-page.xhtml?view

The Community Directory contains a listing of organizations creating and/or publishing data on the Web. The Directory is open and does not require W3C affiliation.

EU Projects

LATC (Linked Open Data Around-The-Clock)

http://latc-project.eu/

LATC is a Specific Support Action (SSA) in the context of the FP7 ICT Challenge 4; it is a publicly funded project, with partners all over Europe and clients around the world. In the LATC project data sources of Institutions and Bodies of the European Union are published as Linked Open Data [http://latc-project.eu/datasets] to seed the EU data cloud: [https://raw.github.com/LATC/EU-data-cloud/master/diagram/eu-data-cloud.png]

EUCLID (EdUcational Curriculum for the usage of LInked Data)

http://www.euclid-project.eu/
http://www.slideshare.net/EUCLIDproject/querying-linked-data

EUCLID is a European project facilitating professional training for data practitioners, who aim to use Linked Data in their daily work. EUCLID delivers a curriculum implemented as a combination of living learning materials and activities (eBook series, webinars, face to face training), validated by the user community through continuous feedback.

The Datahub

CKAN datahub

http://datahub.io
the free, powerful data management platform from the Open Knowledge Foundation

Virtuoso SPARQL Query Editor

http://semantic.ckan.net/sparql/
The query editor for the datahub

LODStats

http://stats.lod2.eu/

LODStats is a statement-stream-based approach for gathering comprehensive statistics about datasets adhering to the Resource Description Framework (RDF). LODStats is integrated into the CKAN dataset metadata registry and obtained a comprehensive picture of the current state of the Data Web.[55]

SPARQL Endpoints Status of CKAN datahub

http://labs.mondeca.com/sparqlEndpointsStatus/index.html
http://labs.mondeca.com/sparqlEndpointsStatus/stats/stats.html
Public SPARQL endpoints list is fetch dynamically using CKAN SPARQL Endpoint. Supported by Mondeca.

The Linking Open Data cloud diagram

http://www.lod-cloud.net/
This web page is the home of the LOD cloud diagram.

Data Hub LOD Validator and Datasets Completeness Levels

http://validator.lod-cloud.net/
This website gives an overview of Linked Data sources catalogued on Data Hub and their completeness level for inclusion [http://validator.lod-cloud.net/levels.html] in the LOD cloud.

Tools

OpenLink Data Explorer Extension

http://ode.openlinksw.com/
The OpenLink Data Explorer (ODE) is a browser extension (currently available for Firefox, Safari, Chrome, Opera, and Internet Explorer with additional browser support to follow) that adds a new option to the realm of Web User Agent functionality, in the form of new menu options for viewing Data Sources associated with Web Pages

SPARQL Query Validator

http://www.sparql.org/query-validator.html

---

[55] cited from: http://aksw.org/Projects/LODStats.html

APPENDIX 2: DEFINITION OF TERMS AND ABBREVIATIONS

API   Application programming interface which allows the use of an application (a software program) as service provider in other applications, extend it with new functionality or simply customize and control it.

CIDOC   International Committee for Documentation (in French: Comité International pour la DOCumentation) of ICOM.

CIDOC CRM   The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

CKAN   The Comprehensive Knowledge Archive Network (CKAN) is a web-based open source data management system maintained by the Open Knowledge Foundation. The system is used as a public platform on datahub.io.

DC   Dublin Core. Metadata standard of the Dublin Core Metadata Initiative (DCMI).

EDM   The Europeana Data Model for Cultural Heritage. Major improvement on ESE

ESE   Europeana Semantic Elements. Europeana's first data model.

FRBR   Functional Requirements for Bibliographic Records. A conceptual entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA) that relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user's perspective.

FRBRoo   FRBR-object oriented. A formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information. Approach to concatenate FRBR and CIDOC CRM.

GLAM   An acronym for Galleries, Libraries, Archives, and Museums, the cultural heritage institutions.

GUI   Graphical user interface. Part of a computer operating system or an application which enables a person to communicate with a computer through the use of symbols, visual metaphors usually on a two-dimensional display.

HTTP   Hypertext Transfer Protocol. Application layer protocol used for the Internet and similar networks.

ICOM   International Council of Museums. Non-governmental organization created in 1946 by and for museum professionals committed to the promotion and protection of natural and cultural heritage.

IRI   Internationalized resource identifier. Generalization of the URI. While URIs are limited to a subset of the ASCII character set, IRIs may contain characters from the Universal Character Set (Unicode/ISO 10646).

LAM   see GLAM.

LIDO   Lightweight Information Describing Objects. XML harvesting schema intended for delivering metadata, for use in a variety of online services, from an organization's online collections database to portals of aggregated resources, as well as exposing, sharing and connecting data on the web.

MQL   Metaweb Query Language. An API and non-standard query language for making programmatic queries to the Freebase database.

OWL   Web Ontology Language. Extends RDFS by adding more advanced constructs to describe semantics of RDF statements. It allows stating additional constraints, such as for example cardinality, restrictions of values, or characteristics of properties such as transitivity. It is based on description logic and so brings reasoning power to the semantic web.

RDF   Resource Description Framework. A W3C standard model for data interchange on the Web. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. This linking structure forms a directed, labelled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.

RDF-S   RDF Schema, the RDF's vocabulary description language provides basic vocabulary for RDF. Using RDF-S it is for example possible to create hierarchies of classes and properties.

REST   Representational State Transfer. Defines a set of lightweight architectural principles by which Web services can be designed that focus on a system's resources. Resource states can be addressed and transferred over HTTP by a wide range of clients written in different languages.

RIF   Rule Interchange Format. W3C recommendation designed for exchanging rules among rule systems, in particular among Web rule engines.

SKOS   Simple Knowledge Organization System. W3C recommendation designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary.

SPARQL   SPARQL Protocol and RDF Query Language. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports aggregation, sub-queries, negation, creating values by expressions, extensible value testing, and constraining queries by source RDF graph. The results of SPARQL queries can be result sets or RDF graphs.

URI   Uniform resource identifier. String of characters with a precise syntax used to identify a web resource.

XML   Extensible Markup Language. A meta markup language developed by W3C to create individual markup languages (as an application of XML) which can be served, received, and processed on the Web in an easy way. XML is now widely used for communicating data between applications.

APPENDIX 3: The AthenaPlus survey questions

I. Partner Information

Please give the following information about your organisation:

1. Country

2. Organisation name

3. Organisation's short name in the project (e.g. UNIMAR)

4. Is the provider a (tick x for all that apply)?
☐ Museum
☐ Library
☐ Archive
☐ Sound archive
☐ Publisher
☐ Aggregator
☐ Other

5. If you ticked 'Other' please give organisation type

Contact (who filled in the form)
6. Name

7. Job title

8. Telephone number

9. E-mail address

II. Using Linked Open Data

10. Are you or your organisation familiar with the concept of Linked Open Data (LOD)?

☐ Yes

☐ No

11. Have you or your organisation had experience of using LOD in connection with your collections?

☐ Yes

☐ No

11a. If you answered 'Yes' please give details. Include for each LOD source(s) you use (e.g. "GeoNames"):
- **The URL which gives access to the LOD source;**
- **Why you use it (e.g. "it is a trusted source");**
- **How you are consuming the data (e.g. "Data dump", "SPARQL endpoint");**
- **How you are linking your own data to the LOD source;**

11b. Does the quality of the LOD sources your organisation is using match your expectations? Please rank for each of the source the quality of the data from 1 (very good) to 5 (bad) according to your opinion.

12. Do you or your organisation know of any LOD projects or initiatives in your country in the field of cultural heritage?

☐ Yes

☐ No

12a. If you answered 'Yes' please give details. Include the URL which gives access to the project or initiative you know about.

III. Publishing Linked Open Data

13. Have you or your organisation had experience of publishing LOD in connection with your collections?

☐ Yes
☐ No

13a. If you answered 'Yes' please give details. Include for each LOD source you have published:
- **The URL which gives access to the LOD source;**
- **What type of data you publish (e.g. "full records", "only basic information");**
- **Which ontologies (namespaces) you are using (e.g. "Dublin Core"; "SKOS"; "CRM"; "own ontology");**
- **How you are serving the data (e.g. "Data dump", "SPARQL endpoint");**
- **What kind of licence for reuse do you give (e.g. "any use"; "non-commercial", "Creative Commons [type]", "no license given");**

14. Does your organisation plan to publish LOD in the near future?

☐ Yes
☐ No

14a. If you answered 'Yes' please give details:
- **Which ontologies (namespaces) are you planning to use (e.g. "Dublin Core"; "SKOS"; "CRM"; "own ontology");**
- **How are you planning to serve the data (e.g. "Data dump", "SPARQL endpoint");**

15. Does your organisation plan to connect with new LOD sources in the near future?

☐ Yes
☐ No

15a. If you answered 'Yes' please give details with which LOD sources you want to connect with.

16. How do you notice that there are new LOD sources as candidates to connect with?

```
[empty text box]
```

17. Which language(s) should a LOD source support so that you would consider connecting with?

```
[empty text box]
```

18. Please describe the expectations and criteria (if there are some) for connecting with LOD sources.

```
[empty text box]
```

19. Do you check the quality of LOD sources?

☐ Yes
☐ No

19a. If you answered 'Yes' please give details in which way you are doing this.

```
[empty text box]
```

APPENDIX 4: DATAHUB VIEW OF "OPEN GLAM" GROUP

no-proprietary-vocab (6)

lld (6)

bibliographic (6)

library (5)

crossdomain (5)

rdf (4)

provenance-metadata (4)

no-license-metadata (4)

format-skos (4)

deref-vocab (4)

cultural (4)

cc0 (4)

published-by-third-... (3)

no-provenance-metadata (3)

license-metadata (3)

heritage (3)

format-xml (3)

europeana (3)

bibliographic data (3)

void-sparql-endpoint (2)

no-vocab-mappings (2)

no-deref-vocab (2)

museum (2)

lodcloud.nolinks (2)

images (2)

format-xhtml (2)

format-rdfs (2)

format-foaf (2)

format-dc (2)

edm (2)

british-library (2)

bibliography (2)

datagovuk (1)

culture (1)

cultural-institutions (1)

country-uk (1)

country-netherlands (1)

collection (1)

authority (1)

archives (1)

archive (1)

amsterdam (1)

Polymath Virtual Library (Authority data) - Fundación Ignacio Larramendi

The Polymath Virtual Library aims to bring together the works of the most important Hispanic polymaths and to establish semantic relationships between them, expressing the...

`application/rdf xml` `example/rdf+xml`

UK Web Domain Historical Format Profile

The UK Web Domain Historical Format Profile contains information on the MIME types of over 2.5 billion web resources from 1996-2010.

`HTML`

Public Library of Veroia

Public Library of Veroia Linked Open Data Project. This is the first effort of a Greek Library to expose it's Bibliographic Data as Linked Open Data.

`api/sparql` `example/rdf+xml` `api/search` `HTML` `application/rdf+xml` `meta/void`

Europeana Linked Open Data 🔥

data.europeana.eu currently contains open metadata on 2.4 million texts, images, videos and sounds gathered by Europeana. These objects come from data providers who have reacted...

`application/x-ntriples` `meta/void` `api/sparql` `application/rdf+xml` `meta/sitemap`

Rijksmuseum API

The Rijksmuseum Application Programming Interface (API) is a Rijksmuseum service for partners and application developers. Register for an API key for access to structural...

`XML`

Cologne based libraries

Cologne-based libraries in Germany released 5.4 million bibliographic records to the public via the CC0 public domain dedication.

`HTML`

Open Beelden API

Open Images is an open media platform that offers online access to audiovisual archive material to stimulate creative reuse. Footage from audiovisual collections can be...

`XML`

British Museum Semantic Web Collection Online

It provides access to the same collection data available through the Museum's web presented Collection Online, but in a computer readable format. The use of the W3C open data...

`sparql`

SMK (Statens Museum for Kunst), The National Gallery of Denmark

Collection of 159 high res images

`ZIP`

Wikimedia Commons

Wikimedia Commons is a media file repository making available public domain and freely-licensed educational media content (images, sound and video clips) to everyone, in their...

`XML`

WW1 Gazettes 🔥

This dataset has no description

`xml, rdf+xml, turtle and json`

access-www (1)

access-bulk (1)

access-api (1)

Show Only Popular Tags

⊤ Formats   Clear All

example/rdf+xml (10)

api/sparql (9)

HTML (9)

XML (5)

meta/void (4)

application/rdf+xml (4)

meta/sitemap (3)

application/x-zip-c... (2)

api/search (2)

PDF (2)

xml, rdf+xml, turtl... (1)

uri (1)

sparql (1)

rdf/turtle (1)

rdf/nt (1)

nt/zip-compressed (1)

meta/rdf-schema (1)

mapping/owl (1)

image/jpeg (1)

html | pdf (1)

example/rdf+ttl (1)

example/rdf+json (1)

sparql (1)

rdf/turtle (1)

rdf/nt (1)

nt/zip-compressed (1)

meta/rdf-schema (1)

mapping/owl (1)

image/jpeg (1)

html | pdf (1)

example/rdf+ttl (1)

example/rdf+json (1)

example (1)

compressed tarfile ... (1)

application/x-ntriples (1)

application/x-nquads (1)

application/rdf xml (1)

EUropeana meta-archive sources

Europeana http://www.europeana.eu/ is made our of other archive and gathers the informations giving an API to access it from remote apps. This dataset is about the...

`CSV` `HTML`

JISC Open Bibliography British National Bibliography dataset 🔥

British Library British National Bibliography package, with records of all new books published in the UK since 1950. Use of Data: This data is being made available under a...

`application/rdf+xml` `api/sparql` `PDF` `HTML` `example/rdf+xml`

British Museum Collection 🔥

Welcome to this Linked Data and SPARQL service. It provides access to the same collection data available through the Museum's web presented Collection Online, but in a computer...

`api/sparql` `example/rdf+xml`

Theatricalia

Theatricalia is a database of past and future theatre productions. It contains information about nearly 20,000 productions involving over 60,000 people at over 1,500 theatres....

English Heritage Places

This dataset contains metadata for about 400,000 nationally important places as recorded by English Heritage, the UK Government's statutory adviser on the historic environment....

`HTML` `api/sparql` `example/rdf+xml` `meta/void`

Swedish Open Cultural Heritage 🔥

SOCH is a set of 3.4 million (as of december 2010) cultural heritage objects harvested from a large number of museums and other local, regional and national cultural heritage...

`HTML` `application/rdf+xml` `example/rdf+xml`

Tyne and Wear Museums Collections (Imagine)

Part of the Europeana Linked Open Data, this is a collection of metadata describing (and linking to digital copies where appropriate) items in the Tyne and Wear Museums...

`XML` `rdf/nt`

Flickr - The Commons

About The key goals of The Commons on Flickr are to firstly show you hidden treasures in the world's public photography archives, and secondly to show how your input and...

Open Plaques

Open Plaques is an open source project (both code and data) has data on over three and a half thousand commemorative 'plaques' (often blue and round) that can be found across...

`JSON`

Bundesarchiv - Wikimedia Commons

About From overview Starting on Thursday Dec 4, 2008, Wikimedia Commons witnessed a massive upload of new images. We received nearly 100,000 files from a donation from the...

« 1 2

APPENDIX 5: SPARQL QUERY FOR THE DATAHUB

The following SPARQL query proved to return the results presented in chapter 5. It is worthwhile to mention that the Default Data Set Name (Graph IRI) must be empty before running the query. The following example query retrieves information about the dataset of the "Bavarian State Library, Bavarian Library Union, Cooperative Library Network Berlin-Brandenburg".

```
PREFIX  rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX  rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX  dcat: <http://www.w3.org/ns/dcat#>
PREFIX  dc:   <http://purl.org/dc/elements/1.1/>
PREFIX  dcterms: <http://purl.org/dc/terms/>

SELECT  ?p ?o
WHERE
  {
      { SELECT DISTINCT ?p ?o
        WHERE
          { {   SELECT DISTINCT  ?p ?o
                WHERE
                  { <http://datahub.io/dataset/b3kat> ?p ?o1 .
                    ?o1 <http://xmlns.com/foaf/0.1/name> ?o
                    FILTER ( str(?p) IN
                          ("http://purl.org/dc/terms/creator",
                           "http://purl.org/dc/terms/contributor") )
                  }
            }
          UNION
            { SELECT  ?p ?o
              WHERE
                { <http://datahub.io/dataset/b3kat> ?p ?o
                  FILTER regex(str(?o), "^(?!node).+")
                }
            }
          UNION
            { SELECT DISTINCT  ?p ?o
              WHERE
                { <http://datahub.io/dataset/b3kat> dcat:distribution ?o0 .
                  ?o0 ?p1 ?o1 .
                  ?o1 ?p2 ?o2 .
                  ?o2 rdf:type dcterms:IMT .
                  ?o0 rdf:type dcat:Distribution .
                  ?o2 ?p ?o
                  FILTER regex(str(?p), "^(?!http://www.w3.org/1999/02/22-rdf-syntax-ns#).+")
                }
            }
          UNION
            { SELECT DISTINCT  ?p ?o
              WHERE
                { <http://datahub.io/dataset/b3kat> dcat:distribution ?o0 .
                  ?o0 ?p1 ?o1 .
                  ?o1 ?p ?o
                  FILTER regex(str(?p), "^(?!http://www.w3.org/1999/02/22-rdf-syntax-ns#).+")
                  FILTER regex(str(?p), "^(?!http://www.w3.org/2000/01/rdf-schema#label).+")
                  FILTER regex(str(?p), "^(?!http://xmlns.com/foaf/0.1/page).+")
                  FILTER regex(str(?o), "^(?!node).+")
                }
            }
          }
      }
    UNION
      { SELECT DISTINCT  ?p ?o
        WHERE
          { GRAPH <http://datahub.io/dataset/b3kat>
              { { ?s rdfs:label ?p .
                  ?s rdf:value ?o
                  FILTER ( ?p != ?o )
                }
              }
          }
        ORDER BY ?p
      }}
```