

Latif, Atif; Tochtermann, Klaus

## Article

# Exploring Scientific Publication and Cross Domain Linked Dataset for Similarity - A Case Study

International Journal of Advancements in Computing Technology

Suggested Citation: Latif, Atif; Tochtermann, Klaus (2013) : Exploring Scientific Publication and Cross Domain Linked Dataset for Similarity - A Case Study, International Journal of Advancements in Computing Technology, ISSN 2005-8039, AICIT, Suwon, Vol. 5, Iss. 11, pp. 179-187

This version is available at:

<http://hdl.handle.net/11108/125>

## Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [info@zbw.eu](mailto:info@zbw.eu)  
<http://zbw.eu/de/ueber-uns/profil/veroeffentlichungen-zbw/>

## Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

## Terms of use:

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.*

# Exploring Scientific Publication and Cross Domain Linked Dataset for Similarity - A Case Study

<sup>1</sup>Atif Latif, <sup>2</sup>Klaus Tochtermann

<sup>1, First Author</sup> ZBW - German National Library of Economics  
Leibniz Information Center for Economics, Kiel, Germany, E-mail:  
[a.latif@zbw.eu](mailto:a.latif@zbw.eu)

<sup>\*2, Corresponding Author</sup> ZBW - German National Library of Economics  
Leibniz Information Center for Economics, Kiel, Germany, E-mail:  
[k.tochtermann@zbw.eu](mailto:k.tochtermann@zbw.eu)

## Abstract

*Linking Open Data Project played a vital role in the realization of structured data at World Wide Web stage by methodically demonstrating the importance of machine understandable data for information linking. It has succeeded in bringing up heap of Linked Open Data ranging from geographic to cross-domain datasets which provide huge opportunities for knowledge discovery and mashup application development. Scientific publication datasets are one of main sources in steering today's research work and has a big share in Linked Data Cloud repository. Besides to it, cross-domain linked data datasets e.g. DBpedia, FreeBase etc. has a huge crowd-sourced open knowledge which proved as good resource for content enrichment and interlinking. Noticing the offered added values of scientific publications and cross-domain datasets it will be great to know; what these datasets has to offer each other in Linked Data settings. We are of a view; if these datasets are interlinked can offer adequate information for enrichment of publication related resources i.e. authors and publications. In addition, this will also help to aggregate information of author in a profile which is currently scattered over different linked data resources. However, currently finding and interlinking with appropriate data is still a challenge in Linked Data Cloud. In this paper we presented a case study by interlinking author from scientific publication dataset (DBLP) with person's record of cross-domain dataset (DBpedia). Moreover, we have investigated to find how much author information is there in DBpedia for indexed DBLP scientific authors and has validated our assumption that meaningful data is present between these datasets.*

**Keywords:** *Linked Open Data, Semantic Web, Linked Data Interlinking*

## 1. Introduction

Linking Open Data project is envisioned to produce structured data which is understandable and process-able by machines. Main objective of this project is to disengage practices of well-gardening data and motivating people to publish their datasets as open and structured data. In a nutshell, it's an effort for bootstrapping Semantic Web Vision at global web scale by creating a global connected data space [1] where related information is better connected. This will make information more reusable and discoverable; as well as leading towards unrestrictive data usage for better data interlinking, querying and intelligent application development. The W3C community project Linking Open Data [2][3] was initiated in 2007. It is based on the Linked Data four principles stated by Tim Berners-Lee [4] which are:

- Use URIs as names for things
- Use HTTP URIs so that people (and machines) can look up those names (see also [5])
- When someone looks up a URI, provide useful information
- Include links to other URIs so that they can discover more things

Basically, these rules provide a set of guidelines for publishing data as Linked Data. Firstly this emphasizes to identify real and abstract concepts within the datasets and then assigning identified resources with unique URIs in Resource Description Framework (RDF) [6] format which are further dereference-able to present more meaningful information. Currently more than

300 data sets consisting of over 31 billion RDF triples which are interlinked by around 504 million RDF links are recorded in September 2011<sup>1</sup>.

Scientific publication datasets are one of the major contributors in Linked Data archive. Recently, with the commencement of Library Linked Data Movement [7] many distinguished digital libraries i.e. ZBW – German National Library of Economics<sup>2</sup>, Europeana<sup>3</sup> and LIBRIS – Swedish National Bibliography and authority database<sup>4</sup> has started publishing their data as Linked Open Data. One of the prominent Computer Science digital libraries is DBLP [8] which has been covering computer science literature from last two decades. It provides access to metadata of scientific literature published in well-known workshops, conference proceedings and journals. Moreover, DBLP has a semantic version of its legacy dataset in the form of DBLP L3S dataset which provides publication metadata as linked data. This metadata of publication usually consists of author names, paper titles, paper keywords and venue information.

At present many research studies has been performed to access and interlink DBLP scientific literature with other scholarly communication datasets. One of the recent performed studies [9] has successfully presented the potentials in using DBLP dataset for content enrichment. Correspondingly cross-domain linked data datasets like DBpedia<sup>5</sup> and Freebase<sup>6</sup> has a huge crowd-sourced open knowledge-base which can be used for discovering, enriching and management of semantic information. These popular datasets covers real world and abstract concepts i.e. person, places, organizations, species etc. in the form of structured pages. DBpedia is one of the popular cross-domain dataset and has been proved as a good resource for content enrichment and interlinking. Recent completed case study [10] around DBpedia has shown its utility for interlinking with domain specific datasets i.e. person, places, movies etc.

These mentioned studies motivated us to investigate more on DBLP and DBpedia for similarity. We are of view if DBLP and DBpedia are interlinked can provide related and additional information about the scientific publications i.e. author, papers, keywords which subsequently can be used to enrich any given publication resource. However, for that to happen we need a comprehensive interlinking strategy which can automatically find and validate the relevant results. Currently, the task of interlinking with relevant resources is still a challenge as compared to linked data publishing. The challenges involved in interlinking are: i) understanding of the underlying semantic structures and ontologies ii) accessibility of sparql endpoints with adequate knowledge of sparql querying.

We focused on these challenges in this study and investigated the interlinking possibilities by presenting a concrete case study of interlinking DBLP with DBpedia. To get quality results we searched, identified, queried and disambiguate the records with various heuristics. The expected outcome of this investigation is following:

- Find and aggregate the scattered information of an author over two linked data sets and will link them with owl:sameAs relationship.
- Resulted datasets will help us to organize information of an author in a profile which we can present by already proposed proof of a concept application.
- Will facilitate students and researchers to find information about certain whom they want to collaborate for getting guidance.
- The result set of this study i.e. owl:sameAs mapping file will be made available for Linked Data community. It will help developers to build new mash up applications.

This paper starts with an overview of the related literature. Afterwards test datasets which were investigated for this study are discussed in details. In the next section, similarity finding

---

<sup>1</sup> <http://lod-cloud.net/>

<sup>2</sup> <http://linkeddata.econstor.eu/>

<sup>3</sup> <http://data.europeana.eu/>

<sup>4</sup> <http://libris.kb.se>

<sup>5</sup> <http://dbpedia.org>

<sup>6</sup> <http://freebase.com>

framework with results is discussed in detail. At the end case study is illustrated with proof of concept application result presentation. The paper closes with discussion and summary.

## **2. State of the Art**

A lot of research work has been done on development of new tools and techniques for addressing Linked Data open challenges. As a result concept of Linked Data has matured but still there are some researchable challenges i.e. data interlinking, ensuring data quality, data provenance track and presentation of linked data in simple user interfaces. Regarding automatic generation of similar RDF links for data interlinking; Heath et al. [11] has surveyed many key-based as well as similarity-based approaches for interlinking. They summed up that problem of making links with external data sources is still a challenge and required more case studies along with better tools for interlinking in LOD Cloud. State of the art in this paper is divided into two parts i.e. interlinking services and case studies.

### **2.1. Interlinking Services**

Sindice [12] is a state of the art semantic search engine. It provides searching facility to process, consolidate and query the linked data. A huge pool of indexed RDF resources in Sindice can be accessed through its offered searching interfaces, public API access and sparql endpoint. Sindice results often need to be analyzed and refined before it can be directly used for a particular use case. Similar kinds of services are provided by semantic search engines like Falcon [13] or Swoogle [14]. However, we selected Sindice for this study because of its larger indexing infrastructure and easy to use search API.

Similarly, SILK framework [15] provides an interlinking tool which is used to discover relationships of resources within different linked dataset. By using Link Specification Language (LSL) and flexible condition features data publisher can set criteria for matching RDF resources for interlinking. SILK framework works on data sources that are interlinked with the SPARQL specification and already presented in RDF format. This tool is very handy where data publisher already know the certain criteria and conditions which needs to be fulfilled. In a nutshell the set of conditions help in selection of specific type entities inside the data source. For instance, in order to interlink cities in DBpedia, a valid restriction may select all entities with the type `dbpedia:city`. For more complex restrictions, arbitrary SPARQL triple patterns are allowed to be specified. However the dataset where inconsistencies are present and further on required instance data level matching this tool may not produce best of the results. As handling similar situation in our case study we have to come up with technique where we can add some heuristic to cater disambiguation problem.

### **2.2. Interlinking Case Studies**

In this case study [9], authors of open digital journal dataset was interlinked with the DBpedia by extending CAF-SIAL (Concept Aggregation Framework for Structuring Informational Aspects of Linked Open Data) application<sup>7</sup>. It is a proof of concept linked data application which finds and harvests person's relevant information from the LOD cloud [20]. Moreover, this information is organized and presented in informational aspect of person (personal, professional, educational and social) in a profile. This system is currently used by the Journal of Universal Computer Science<sup>8</sup> where authors were successfully interlinked with DBpedia. Moreover, author's information is presented in a well-crafted profile which has proved very beneficial for the journal administration and journal users. This work has motivated us to linkup the DBLP datasets with DBpedia.

---

<sup>7</sup> <http://cafsial.dratiflatif.com>

<sup>8</sup> <http://www.jucs.org/>

Another study [10] depicts a concrete use case where a relational database of scientific authors and publication from I-Know<sup>9</sup> - a research organization was first "RDFized" and then were further semi-automatically linked with DBLP. In this study, first manual selection of the suitable scholarly linked data datasets was performed by querying popular author of research organization in available scholarly linked datasets, i.e. DBLP D2R L3S, CITESEER and ACM RKB Explorer<sup>10</sup>, which resulted in selection of DBLP for interlinking. Secondly, after interlinking data was arranged and presented in an automatic generated profile. This study has shown the potentials and added value of DBLP as a dataset for content (publications and authors) enrichment and resource discovery.

Other recent studies have showcased the wide adaptability of Semantic Web technologies in various domains. For instance in [16] study authors have highlighted the importance of LOD publishing format for government statistical data. It stressed on linked data up taking for better accessibility, reusability, transparency and information linking. In another study [17] authors with the help of using semantic technologies has proposed a personalized recommendation system which can produce more related results for particular user than classic methods.

### **3. Experimented Linked Data Repositories**

For our interlinking study, we have accessed and made use of two linked data repositories i.e. DBLP and DBpedia. These datasets are available openly and can be accessed via their respective sparql endpoint. Given below are the details of these datasets:

#### **3.1. DBLP D2R L3S Server**

The DBLP D2R L3S server<sup>11</sup> is based on the XML dump of the DBLP database. The DBLP [7] database provides bibliographic information on major computer science journals and conference proceedings. The database contains more than 1 million articles and 14,500 authors. To query the DBLP L3S data set, the D2R Server, a semantified version of DBLP bibliography, was accessed via its sparql endpoint.

#### **3.2. DBpedia**

DBpedia is a semantic version of Wikipedia<sup>12</sup> - a popular free internet encyclopedia. This project is based on the extraction of structured content from Wikipedia articles which is further made available as Linked Open Data. DBpedia allows querying the properties, relationship and external links of other resources which are associated with Wikipedia pages. DBpedia is considered as a nucleus and famous cross-interlinking hub within Linked Data Cloud as also described by Tim Berners-Lee [18][19]. The English version of the DBpedia knowledge base currently describes 3.77 million things, out of which 2.35 million are classified in a consistent Ontology, including 764,000 persons, 573,000 places, 333,000 creative works, 192,000 organizations, 202,000 species and 5,500 diseases<sup>13</sup>. Resource Description Framework (RDF) is used to represent these records and is accessible in form of RDF triples. For our part of study, we made use of persons records present in DBpedia and accessed these records by querying DBpedia sparql endpoint.

### **4. Similarity Finding Framework**

---

<sup>9</sup> <http://i-know.tugraz.at/>

<sup>10</sup> <http://rkbexplorer.com/>

<sup>11</sup> <http://dblp.l3s.de/>

<sup>12</sup> <http://en.wikipedia.org/>

<sup>13</sup> <http://dbpedia.org/About>

To achieve interlinking between DBLP and DBpedia dataset, a multi-step strategy was devised to find similar resources in form of authors incase validated was then presented in a profile. Further on, these discovered resources were processed and annotated with owl:sameAs relationship for generating a final mapping file. The framework for this strategy is illustrated in Figure 1. The strategy looked as follows:

#### 4.1. Author’s Metadata Collection from DBLP

At first step, metadata of the authors was retrieved by querying DBLP D2R server sparql endpoint with the help of web script. Afterwards, all retrieved author records along with full names and DBLP URIs were stored in a relational database for further processing. In total 14,640 authors records from DBLP was retrieved.

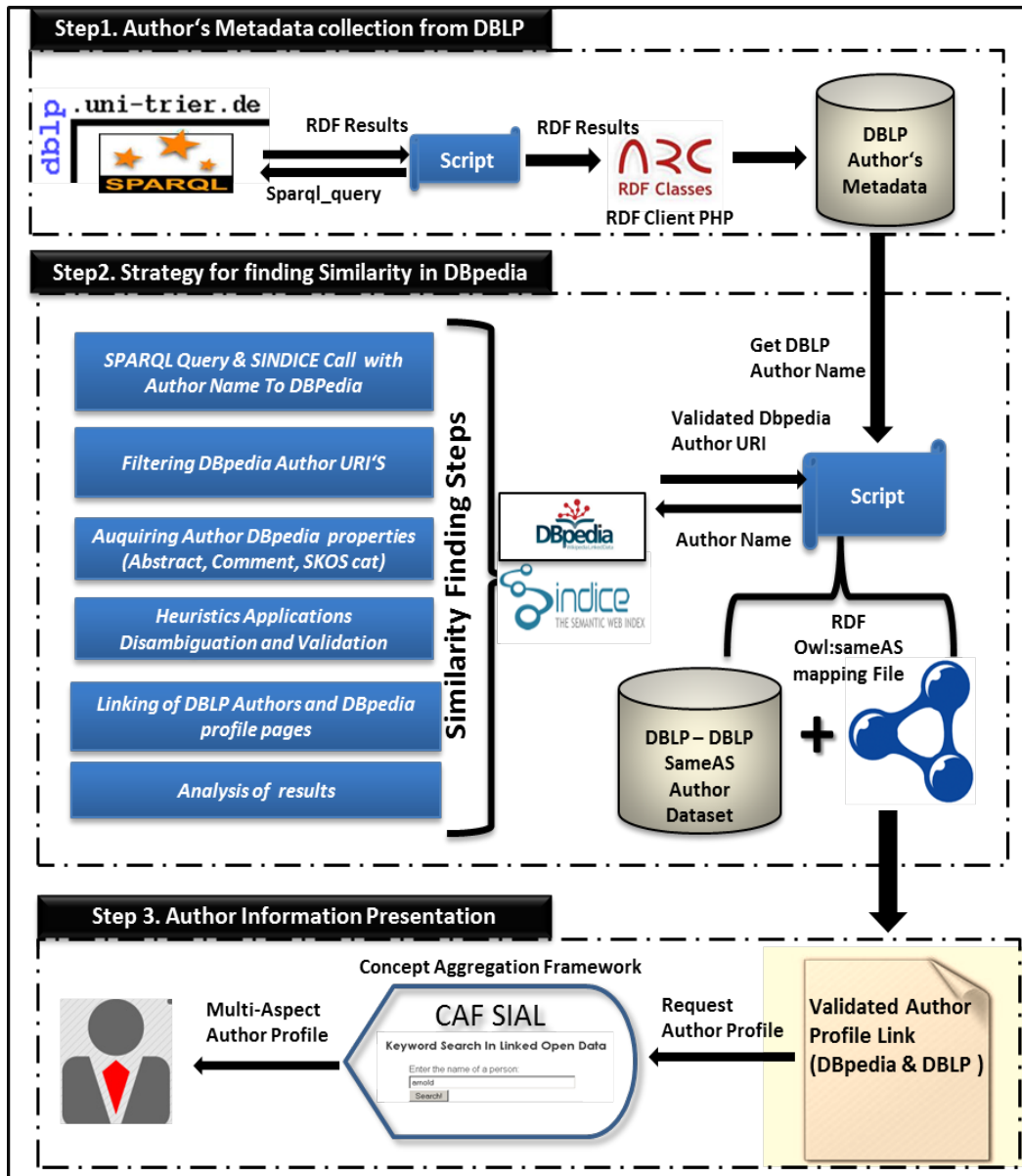


Figure 1. Similarity Finding Framework

## 4.2. Finding Author Similarity from DBpedia

In second step authors of DBLP were searched in DBpedia person data repository for similarity. As a preprocessing step to minimize the inconsistencies in names of authors i.e. umulate and special characters were processed. Afterwards, DBpedia dataset was queried with author names in two ways for producing maximum matched results. Firstly, Sindice Search API was used to search authors for DBpedia URI's. A web script which took the authors name iteratively as an input and automatically called API with formulated search queries was written.

The resulting URIs was then filtered automatically on the basis of heuristics to make sure that it only belongs to the DBpedia. Due to recall and precision problems which every search engine displays we decided to use DBpedia sparql endpoint for improving our results. For that, we employed a string-matching algorithm that compared names of the authors in question with author names from the DBpedia dataset. To accomplish this, we queried sparql endpoint of DBpedia with our web service. This step provided us with the URI's of author whom full name was matched with the person name in DBpedia dataset. This way we found a few additional URIs which was not retrieved by Sindice in the previous step. At the end, we combined the results of both approaches and constructed a relational database with details of each author. URIs of 1184 out of 14640 authors in question was found in this process.

## 4.3. Validation Service

In order to ensure the validity of the matched URIs, a validation service was written which will validate matched DBpedia and DBLP authors URIs. This service iteratively took the author name with matched DBpedia URI and constructed a sparql query for retrieving the associated properties in DBpedia. By manual inspection and de-referencing of the acquired URI's, we discovered some irregularities which are discussed below:

- First, URI's of many authors were present in the DBpedia but providing no additional statements i.e. abstract, comments and categories. This particular limitation made it impossible for predicting the correct match.
- Secondly, a large portion of person had similar name like authors but actually belonged to other professions i.e. sports, celebrity, actor etc. This might have led us to the wrong person identification and interlinking. To negate this we have to think for heuristically approach which can help us to identify persons only belonging to academic or education field.

Hence to disambiguate authors, a set of heuristics were written. After manual inspection, it was noted that there are certain kind of properties in DBpedia for type person those can be exploited to predict correct person. These properties are dbpedia:abstract/ dbpedia:comment and SKOS categories. For example SKOS categories and keywords being used to represent the persons belonging to education profession are: “computer science, computer scientist, professor, informatics, researcher .....” etc. All of these mentioned properties represent a person belonging to scientific community. Thus the persons having same names and belonging to different professions can be filtered out by applying these keywords. For that an automated script was written to check the keywords in the abstract, comments and SKOS categories of the URI. After applying this script on 1184 authors 142 URIs was validated. The remaining URI's were either bad links (showing no data) or representing non-scientific persons.

## 4.4. Author Information Presentation:


After interlinking to highlight the expected benefits of this study, results were presented in an author profile. The presentation of results was possible by making use of our already running proof of concept application CAF-SIAL (see also [20]) for details. Initially DBLP author's whom DBpedia URI was validated are listed at this webpage ([http://cafsial.dratiflatif.com/dbpedia\\_dblp\\_mapping.html](http://cafsial.dratiflatif.com/dbpedia_dblp_mapping.html)). When user clicks on profile hyperlink of any listed author, it initiates a request towards application for creating multi-aspect

profile on a fly. Currently, profile of authors who only had a validated DBpedia URI is listed on this web page. However the analysis of DBLP information showed that there is meaningful information for mining co-authorship network of authors. Both cases have validated our assumption that there lays a valuable data which can be used for content enrichment. An illustration of author profile is shown in figure 2.

CAF-SIAL: Concept Aggregation Framework for Structuring Informational Aspects of Linked Open Data

[Home](#)    [Publications](#)    [News](#)    [License](#)    [Contact](#)

**Iain S. Duff**



**Details (Aspect Classification)**

**Introduction**

Iain S. Duff is a British mathematician and computer scientist, known for his work in numerical methods and software for solving problem with sparse matrices, in particular the Harwell Subroutine Library. From 1986–2009, he was the Group Leader of Numerical Analysis at Harwell Laboratory, which has moved in 1990 to the Rutherford Appleton Laboratory. He is also the Project Leader for the Parallel Algorithms Group at CERFACS in Toulouse.

Personal Aspects	Professional Aspects
<p><b>Name</b> = Duff, Iain S. Iain S. Duff  <b>surName</b> = Duff  <b>givenName</b> = Iain S.  <b>birthdate</b> = 1947-01-05  <b>Birth place</b> = Glasgow, Scotland  <b>residence</b> = United_Kingdom  <b>Short description</b> = Mathematician, computer scientist, numerical analyst  <b>hasPhotoCollection</b> = <a href="http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/Iain_S._Duff">http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/Iain_S._Duff</a></p>	<p><b>Professional Details</b></p> <p><b>field</b> = Computer_science, Mathematics  <b>workPlace</b> = Atomic_Energy_Research_Establishment, Harwell_Laboratory  <b>profession</b> = Professor, Computer Scientist</p>
<p style="text-align: center;"><b>Academics</b></p> <p><b>doctoral Advisor</b> = Leslie_Fox                      Reginald_Prabhakar_Tewarson  <b>almaMater</b> = University_of_Oxford  <b>reference</b> = <a href="http://www.numerical.rl.ac.uk/people/isd/isd.html">http://www.numerical.rl.ac.uk/people/isd/isd.html</a>  <a href="http://www.cse.scitech.ac.uk/nag/hsl/">http://www.cse.scitech.ac.uk/nag/hsl/</a>  <a href="http://history.siam.org/oralhistories/duff.htm">http://history.siam.org/oralhistories/duff.htm</a>  <a href="http://math.nist.gov/MatrixMarket/collections/hb.html">http://math.nist.gov/MatrixMarket/collections/hb.html</a></p>	<p><b>Achievements</b></p> <p><b>Honors</b> = Alumni_of_the_University_of_Oxford, British_computer_scientists</p>
	<p style="text-align: center;"><b>Contributions (DBLP Indexed)</b></p> <p><b>DBLP_autho_link</b> = <a href="http://dblp.l3s.de/d2t/page/authors/Iain_S._Duff">http://dblp.l3s.de/d2t/page/authors/Iain_S._Duff</a></p>
	<p style="text-align: center;"><b>Publication List</b></p> <p><b>Journals</b>= On Computing Inverse Entries of a Sparse Matrix in an Out-of-Core Environment.                      Design, implementation, and analysis of maximum transversal algorithms On the Block Triangular Form of Symmetric Matrices. <a href="#">more</a>  <b>conferences</b> = European Exascale Software Initiative: Numerical Libraries, Solvers and Algorithms.                      A Parallel Matrix Scaling Algorithm. <a href="#">more</a></p> <p style="text-align: right;">[...Section Deleted...]</p>

**Figure 2. Author Profile PowerPoint Illustration**

## 5. Observations

The analysis of results led us to the following observations:

- In interlinking process we have identified many person pages which may lead us to ambiguous results mainly due to similar name issues. We propose that some additional measure can be taken to cater this issue. For instance, editors or contributors can introduce a special category or profession property netted with URI for a person page.
- Not many of computer scientist's profiles present in DBpedia dataset. By keeping in mind the huge usage of social media it came out as a surprise. However, we are of a view if more profiles can be maintained by the editors or contributor can help in: a) more valuable interlinking and b) scientist to increase their visibility. This practice may also be valuable in increasing interlinking scores between DBpedia and DBLP and further on interlinking results can be used in development of various mash-up applications.
- Over all, success of Linked Open Data project depends upon the heavy interlinking between external datasets, so that we can have a single connected global data space where simple and sophisticated queries are possible. If it happens can lead us to more knowledge discoveries.



However currently, the lack of adequate tool, quality data and few interlinking case studies impeding the way to achieve basic vision of Linked Open Data. Still there are needs of more efforts for data creation, publishing tools and interlinking mechanisms. We hope this study can be useful for future interlinking studies.

## 6. Conclusions

Linked Open Data provides a heap of structured data ranging from geographic to cross-domain datasets which provide huge opportunities for knowledge discovery and mashup application development. Scientific publication datasets are one of main sources in steering today's research work and has a big share in Linked Data Cloud repository. Besides to it, cross-domain linked data datasets has a huge crowd-sourced open knowledge which proved as good resource for content enrichment and interlinking. Noticing the offered added values of scientific publications and cross-domain datasets it will be great to know; what these datasets has to offer each other in Linked Data settings. However, the process of interlinking data with external datasets stills quite a challenge. In this paper, we investigated the interlinking challenge and proposed a multi-step strategy for finding similar records for interlinking between DBLP and DBpedia. Surprisingly, we have found only small number of similar records in both repositories. However, this study has shown that the discovered information is very handful for content enrichment. Especially the biographical information of a person with other relevant properties in DBpedia can be very useful first in disambiguation of persons and secondly in improving author information with in scientific publication dataset. At the end we have generated a mapping file of validated interlinking links as of owl:sameAs relationship and presented the results in a multi aspect author profile. However, we assume if more information of authors can be created by editors or contributors in DBpedia can help in achievement of critical mass. This can lead to development of new interesting mashup applications hence playing a key role in increasing authors' visibility to outside datasets.

## 7. References

- [1] Heath, T. and Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space* (1st edition). *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan and Claypool 1:1., 2011
- [2] W3C community project *Linking Open Data*, 2007. <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- [3] Bizer, C., Heath, T., Ayers, D., Raimond, Y., "Interlinking Open Data on the Web", *Demonstrations Track at the 4th European Semantic Web Conference*, Innsbruck, Austria, 2007
- [4] Berners-Lee, T. (2006). *Linked Data -- Design Issues*, 2006. <http://www.w3.org/DesignIssues/LinkedData.html>
- [5] Sauer mann, L., Cyganiak, R., Ayers, D., Völkel, M., "Cool URIs for the Semantic Web", W3C Interest Group Note, 2008. <http://www.w3.org/TR/2008/NOTE-cooluris-20081203>
- [6] *Resource Description Framework*, 2004 <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>
- [7] *Library Linked Data Incubator Group: Use Cases*. <http://www.w3.org/2005/Incubator/ldd/XGR-ldd-usecase-20111025/>
- [8] Michael, L., "DBLP - Some Lessons Learned", *PVLDB*, vol. 2, no. 2, pp. 1493—1500, 2009
- [9] Latif, A., Afzal, M.T., Helic, D., Tochtermann, K., Maurer, H., "Discovery and Construction of Authors' Profile from Linked Data (A case study for Open Digital Journal)", *Linked Data on the Web (LDOW'10)*, 2010
- [10] Latif, A. and Tochtermann, K., "Webbing Semantified Scholarly Communication Datasets for Improved Resource Discovery", *Journal of Digital Information Management*, vol. 10, no. 4, pp. 245-253, 2012
- [11] Bizer, C., Heath, T., Berners-Lee, T., "Linked data-the story so far", *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5 (3) , pp. 1-22, 2009

- [12] Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G., "Sindice.com: A Document-oriented Lookup Index for Open Linked Data", *International Journal of Metadata, Semantics and Ontologies*. vol. 3, no. 1, pp. 37-52, 2008.
- [13] Cheng, G., Ge, W., Qu, Y., "Falcons: Searching and Browsing Entities on the Semantic Web", In *Proceeding(s) of 17th international conference on World Wide Web, WWW '08*, ACM, pp. 1101-1102, Beijing, China, 2008
- [14] Ding, L., Finin, T., Joshi, A., Pan, R., S. Cost, R., Peng, Y., Reddivari, P., C. Doshi, V., Sachs, J., "Swoogle: A Search and Metadata Engine for the Semantic Web", In *Proceeding(s) Thirteenth ACM Conference on Information and Knowledge Management*, ACM, pp. 8-13, Washington, D.C., USA, 2004
- [15] Volz, J., Bizer, C., Gaedke, M., Kobilarov, G., "Silk—A Link Discovery Framework for the Web of Data", In *Proceeding(s) of CEUR-WS Vol-538 of 2nd Linked Data on the Web Workshop (LDOW2009)*, CEUR-WS, vol. 538, Spain, 2009.
- [16] Bunyakiati, P. , Voravittayathorn, P., "Towards Linked Open Statistical Data: Dissemination Formats and Major Statistic Data Sets of the AEC Countries", *JCIS: Journal of Communications and Information Sciences*, Vol. 3, No. 2, pp. 49 ~ 58, 2013
- [17] Mu, X., Chen, Y., Cao, Y., Li, Y., "Personalized Recommendation System Modeling in Semantic Web", *AISS: Advances in Information Sciences and Service Sciences*, Vol. 5, No. 2, pp. 278 ~ 284, 2013
- [18] Sir Tim Berners-Lee Talks with Talis about the Semantic Web, Talis, 7 February 2008.
- [19] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z., "DBpedia: A Nucleus for a Web of Open Data", In *Proceeding(s) of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007*, *Lecture Notes in Computer Science* 4825 Springer 2007, ISBN 978-3-540-76297-3, 2007
- [20] Latif, A., Afzal, M.T., Us Saeed, A., Hoefler, P., Tochtermann, K., "CAF-SIAL: Concept Aggregation Framework for Structuring Informational Aspects of Linked Open Data". In *Proceeding(s) of of International Conference on Networked Digital Technologies*, pp. 100-105, Ostrava, Czech Republic, 2009