

Hajra, Arben; Pianos, Tamara; Tochtermann, Klaus

**Conference Paper — Accepted Manuscript (Postprint)**

**Linking Author Information: EconBiz Author Profiles**

*Suggested Citation:* Hajra, Arben; Pianos, Tamara; Tochtermann, Klaus (2021) : Linking Author Information: EconBiz Author Profiles, In: Garoufallou, E. Ovalle-Perandones, M. (Ed.): Metadata and Semantic Research 14th International Conference, MTSR 2020, Madrid, Spain, December 2–4, 2020, Revised Selected Papers, ISBN 978-3-030-71903-6, Springer Nature Switzerland, Cham, pp. 180-191, [https://doi.org/10.1007/978-3-030-71903-6\\_18](https://doi.org/10.1007/978-3-030-71903-6_18)

This Version is available at:

<http://hdl.handle.net/11108/486>

**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)  
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/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.*

## Linking Author Information: EconBiz Author Profiles

Arben Hajra<sup>\*[0000-0001-5081-5927]</sup>, Tamara Pianos<sup>[0000-0002-4386-519X]</sup>, Klaus Tochtermann

ZBW - Leibniz Information Centre for Economics  
Kiel/Hamburg, Germany  
[a.hajra, t.pianos, k.tochtermann]@zbw.eu

**Abstract.** Author name ambiguity represents a real obstacle in the world of digital library (DL) information retrieval. The search with an author's name almost always casts doubt on whether all publications in the result list belong to that author or another author sharing the same name. In several other cases, the scholar is interested in having additional information about a selected author, such as a short biography, affiliations, metrics, or co-relations with other authors. The main purpose of this work is the integration and usage of diverse data, based on Linked Data approaches and authority records, to create a comprehensive author profile inside a DL. We are proposing and deploying an approach that provides such author profiles by using the available data, on-the-fly, i.e., harvesting the available sources for this purpose. The proposed approach - developed as a fully functional prototype - has been introduced for evaluation to a group of authors, scholars, and librarians. The results indicate acceptance of such an approach, underlining the benefits and limitations that come with it.

**Keywords:** Authority Files, Identifiers, Researcher Profiles, Wikidata, LOD

### 1 Introduction

Author name ambiguity represents a real obstacle in DL information retrieval. Namely, the search based on the author's name almost always casts doubt on whether the result i.e. all publications in the result list belongs to that author. The problem, therefore, arises with the persons sharing the same name, e.g. Joachim Wagner [1980] vs Joachim Wagner [1954]. On the other side, searching with an author name does not provide a complete list of results either because of different name variations e.g. Judžin F. Fama, Eugene Francis Fama, Gene Francis Fama.

Furthermore, on several occasions when a scholar is searching for a particular author, there is a concern in further details, to ensure a comprehensive overview of the author. The information, such as birth year, affiliations, profession, biography, and metrics would provide a broad picture of the author and help the scholar in that regard. However, retrieving such kind of data will cost the scholar several navigations, clicks, and is thus time-consuming. Besides biographical information, the author profiles would be more inclusive if they displayed different content analyses based on the author's research output. Such kinds of analyses, with a potential visualization of main topics and concepts from their research output e.g. "income distribution", "climate change", "behavioral economics", may provide a clearer outlook of the authors' covered topics.

Another important issue that usually represents the interest of scholars is the list of co-authors with whom a particular author has collaborated, i.e. to know with whom the author has collaborated mostly, and to have an immediate switch to their profile. Moreover, let us assume that a scholar found an author whose research is closely related to their research field. They would be interested in finding other authors working on similar topics, besides the co-author network.

In addition, author metrics are an important indicator for showing the impact of authors' research output. The presence of an h-index or i10 index for measuring the citation impact and productivity may help the scholar to quickly assess the work. The author's opinion on recent topics may also be of interest. For this purpose, the scholar would need to search and browse for any blog post or any other posts on social media.

Finding, harvesting, and consuming the information needed for these purposes is possible only by navigating through many websites and spending a considerable amount of time. The main purpose of this work is the integration and usage of various data, based on Linked Data approaches and authority records, with the aim of creating a comprehensive author profile inside a Digital Library (DL). Such a profile helps the scholarly process, by reducing the time and effort of manually finding and collecting data, and increases the quality of data in the sense of accuracy. Our approach focuses on researchers in business and economics, based on the data sources used in the EconBiz<sup>1</sup> environment. A service like the one presented here can be used by scholars to find more information on specific authors. Authors themselves can use it to see how others make use of named entities or linked data. For this approach, we make use of the available data by harvesting as many sources as possible.

The paper begins by presenting some of the concerns that have motivated us in this regard. Section 2 describes our proposed approach, as well as examining several services and methods that are used in our development, such as authority files and linked open data. The deployment of the proposed approach, the design, and the main features are described in section 3. The collected results, evaluations, and limitations are described in section 4. The paper concludes with the section 5.

## 2 Proposed Approach and Background Research

The goal of this work is to overcome most of the issues tackled in the previous section, by proposing to generate author profiles for every author in a particular DL. Scholars will find the following information aggregated in one place, such as a comprehensive list of publications, co-authorship relations, a short bio about the author, affiliations, professions, topics covered in their research output, and everything else that can be found by linking further sources. Our intention in this regard is to make use of the available data, on the fly, by harvesting the available sources for this purpose. Hence, instead of creating and storing the data locally - creating another isolated silo of data - we intend to use the data by exploiting the existing data or by creating links.

Harvesting the data from other sources is almost impossible if we rely only on the author's name, for the reasons stated in the previous sections. Therefore, we rely on any

---

<sup>1</sup> <https://www.econbiz.de>

available persistent identifier (PID). Nowadays, there are several efforts for generating authority profiles for aggregating and uniquely identifying resources and authors.

## 2.1 Authority control

The benefits of authority controls have been a topic in research and many discussions [1]. Therefore, one of the major benefits is precisely the appearance of a single entity in one name, which can often be presented differently elsewhere, as well as the ability to create interlinks among them [2]. Authority files integrate data from several sources such as name variations, biographical info, affiliations, and represent them into separate clusters. Hence, for each cluster i.e. authority record a unique id is assigned.

As most prominent Authority files we may distinguish the Virtual International Authority File (VIAF) and the Integrated Authority File (GND). VIAF integrates multiple name authority files from many institutions and national libraries, around the world [3]. Initially, the main contributors were the Library of Congress (LC)<sup>2</sup>, the German National Library (DNB)<sup>3</sup>, and OCLC<sup>4</sup>. Currently, the service is hosted by OCLC.

The Integrated Authority File (GND) known as Gemeinsame Normdatei in German, is a service for facilitating the use and administration of authority data. The GND is managed by the German National Library (DNB)<sup>5</sup>, all library networks of the German-speaking countries and their participating libraries, the German Union Catalogue of Serials (ZDB), and many other institutions [4]. Similar to VIAF, it integrates information about different entities such as persons, corporate bodies, conferences, events, geographic information, topics, and works. Each entity is identified with a specific id, known as GND ID. Presently, the GND contains over 15 million authority records.

In addition to classical authority controls, there are several other initiatives for clustering and identifying authors with a persistent identifier. The usage of ORCID, RePEc, SSRN, Google Scholar, Wikidata ID, offers many opportunities in this regard.

## 2.2 Linked Data and linking PIDs

The deployment of Linked Data principles and Semantic Web technologies increases the visibility, interoperability, and accessibility of the data [5, 6], in comparison to the MARC format which does not offer much in that direction [7, 8]. Publishing the data as linked open data (LOD) is getting more and more acceptance [9, 10]. Today we can find the entire catalogs of bibliographical and authority records serialized in any of the resource description frameworks (RDF) formats and accessible as dump files, API, or SPARQL endpoints. Such a case is the German National Library (DNB) that is offering several linked data services for authority or bibliographic records [11, 12]. One of these linked data services is Entity Facts, which provides all the bibliographic data from DNB and the authority data from GND free of charge [13]. Besides the GND data, Entity

---

<sup>2</sup> <https://www.loc.gov/>

<sup>3</sup> <https://www.dnb.de>

<sup>4</sup> <https://www.oclc.org>

<sup>5</sup> <https://www.dnb.de/>

Facts also contain links to other data providers such as VIAF, ISNI, BNF, LoC, Wikipedia, Wikidata, Wikimedia Commons, etc., and in some cases, it harvests information from their side (like the author pictures from Wikimedia Commons).

An interesting fact is the presence of other identifiers as part of the authority files. In this respect, VIAF also provides links to sources such as GND, LC, BNF, Wikidata, etc. The same is the case with the GND, where among other PIDs, identifiers like the ORCID ID can be found. As of June 2019, there were more than 50.000 GND records connected with ORCID IDs [14].

Another example of linked open data implementation in the domain of libraries and authority files is the LOBID<sup>6</sup> services from the North Rhine-Westphalian Library Service Centre (hbz)<sup>7</sup>. Among the services, the “lobid-gnd” provides a search interface for searching the GND and a web API that is based on JSON-LD. The web API application of lobid-gnd is used in our approach [15].

RePEc (Research Papers in Economics) is a bibliographic database of working papers, journal articles, books, book chapters, and software components [16]. It provides several remarkable and up to date ranking services, e.g, rankings of individuals, journals, institutions[17]. Among the several services that are using the RePEc data, CitEc represents an example of showing various metrics, such as a number of citations, h-index, i10 index, for a given author [18].

Wikidata represents a free and open knowledge base that can be accessed and edited by both humans and machines [19]. It is the main data management platform for Wikipedia, and where the community-created knowledge is of essential importance [20]. The growth of Wikidata is enormous, with new records added every minute. Currently, it counts around 90 million data items, i.e. things in human knowledge, including topics, concepts, and objects. Each of the items is identified with a unique id that starts with “Q”, and basically, each item consists of a particular page with the editable details, such as a label, a description and aliases.

Among the tools and services that are using Wikidata, Scholia is an interesting example of how to handle scientific bibliographic information. Scholia is a Web service that aggregates on-the-fly data from Wikidata with the purpose to create scholarly profiles for researchers, organizations, journals, publishers, individual scholarly works, and research topics [21].

In addition to the diverse information provided, Wikidata plays a very important role as the central hub for linking identifiers and authority records [22]. Moreover, there are many advantages of creating and storing such kinds of interlinks in public and community-curated linking hubs like Wikidata, rather than in closed local databases [23].

The presence of authority identifiers and other PIDs’ statements inside the Wikidata items is growing rapidly. At present, around 2.6 million humans in Wikidata contain a VIAF property, and more than one million a GND ID. Other identifiers, such as RePEc, ORCID, and SSRN are also included. Hence, by knowing at least one of the identifiers, the list can be extended with several others through Wikidata.

---

<sup>6</sup> <https://lobid.org>

<sup>7</sup> <https://www.hbz-nrw.de/>

### 2.3 The subject portal EconBiz

EconBiz is a free search portal for economics and business studies, provided by the ZBW - Leibniz Information Centre for Economics<sup>8</sup>. It has a disciplinary focus on business, economics, and related subjects. EconBiz includes more than 10 million records from different databases (ECONIS, OLC EcoSci, RePEc, USB Cologne, EconStor, etc.). The integration of the Standard Thesaurus for Economics (STW)<sup>9</sup> supports researchers by suggesting keywords and related terms. EconBiz content is also accessible through a RESTful API, with the base URL <https://api.econbiz.de/v1/>.

Metadata in EconBiz is partially disambiguated, i.e. around 0.5 Million authors in several bibliographical records are identified with a unique identifier such as a GND ID (figure 1). In addition, through an ongoing project, we aim to extend the number of uniquely identified authors through co-authors analyses. Furthermore, when existing metadata within a DL are incapable to accurately identify an author, the authority file approaches are considered, such as the usage of VIAF clusters.

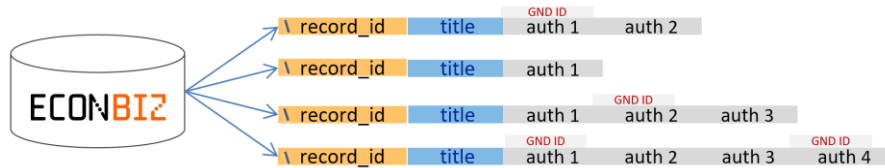


Fig 1. Author's disambiguation state in EconBiz records

## 3 Author Profiles - EconBiz as a use case

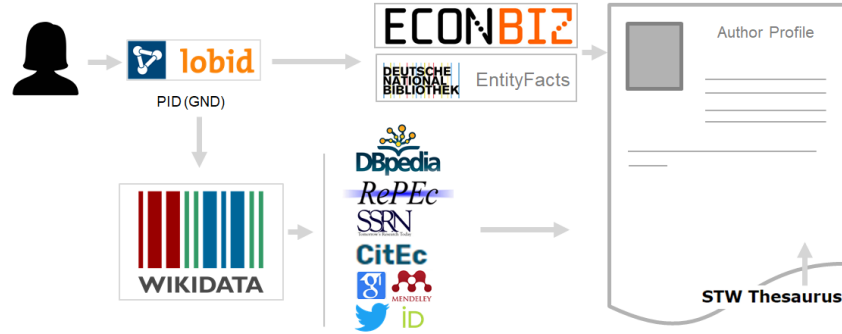
The proposed approach is developed as a prototype for generating author profiles for the subject portal (DL) EconBiz. For each author in EconBiz, where a GND ID is assigned, a comprehensive profile is created. As explained in the previous section, especially in part 2.3, the GND ID is accepted as the central identifier for this purpose.

### 3.1 The approach

In the intention to create profiles for authors, the main idea is to make use of the data available from other sources. Instead of creating and storing the data locally, which will require additional effort on creation, collection, and curation, we are harvesting the data from several sources, on the fly. The ever-increasing publication of open data, as well as the adaptation of linked data principles, provides some relief in this regard. Hence, today metadata or entire catalogs are accessible without any cost or manual input. In our approach, we make use of several sources and services for different purposes. The harvesting process is made through the consumption of already provided APIs and various SPARQL queries. An example of such a query is shown in Listing 1, for retrieving all the prizes that a particular author has won in economics, based on Wikidata input. Figure 2 provides an overview of the flow and data sources used for this approach.

<sup>8</sup> <https://www.zbw.eu>

<sup>9</sup> <https://zbw.eu/stw/version/latest/about.en.html>



**Fig 2.** The flow chart for creating EconBiz author profiles

Initially, we are targeting one of the LOBID services (lobid-gnd) for searching the appropriate GND ID, given that searching by name is easier for the scholar than with a particular ID. With the GND ID, the Entity Facts service can be consumed to retrieve diverse information about the author, e.g. alternative names, profession, picture, profession, affiliations, and some external identifiers. In all the cases when other identifiers like RePEc ID or ORCID are available but the GND ID is missing, Wikidata is used as a hub for extending the range of identifiers. More identifiers increase the possibility to target more sources i.e. DBpedia, Wikidata, ORCID, or RePEc.

The usage of EconBiz content is of crucial importance for further data processing and visualizations. The EconBiz API is consumed for listing the results, author's topics, and co-authors network. Besides, for further terms and concepts enrichment, the STW thesaurus is considered.

**Listing 1.** SPARQL query to Wikidata for getting author's prizes in economics

```
SELECT DISTINCT ?label WHERE {
  wd:{WID} p:P166 ?statement.
  ?statement ps:P166 ?award.
  ?award wdt:P31 wd:Q17701409.
  ?award rdfs:label ?name.
  OPTIONAL {?statement pq:P585 ?date.BIND(YEAR(?date) AS ?year) }
  SERVICE wikibase:label {bd:serviceParam wikibase:language "en"}
  BIND(CONCAT(STR(?year), " - ", str(?name) ,"") AS ?label ).
  FILTER (LANG(?name) = "en").
}ORDER BY ASC(?awardLabel) ASC(?year)
```

### 3.2 The implementation of EconBiz Author Profiles

The developed prototype, the EconBiz Author Profiles<sup>10</sup>, integrate various sets of data from several sources. The users have the possibility to initiate a search from a search bar, which integrates the LOBID search service and enables the search with the author's name with a combination of several parameters such as the birth year. In addition to the

<sup>10</sup> <https://authors.econbiz.de>

previous functionally, users have the possibility to select an author from already generated lists of Nobel laureates in economics and the top 1,000 RePEc economists. The list of Nobel laureates in Economics is generated through the Nobel API<sup>11</sup> and Wikidata, while the list of the top 1,000 economists is generated through the RePEc API<sup>12</sup>. It is worth mentioning, that the RePEc API provides the name of the author, the ranking position, and the short RePEc ID. Since our system is relying on GND IDs, we are using the Wikidata hub for assigning the GND ID to the authors. During our first check at Wikidata for this kind of mapping (RePEc to GND), August 2018, around 97% of the authors from RePEc top 1,000 were already part of Wikidata. Hence, with just a minimal effort all the top 1,000 RePEc authors are now mapped with a GND ID. If the ranking changes, as usually happen, the system automatically updates the list with the most recent names and positions.

The display of information is organized in three main areas, as shown in figure 3.

**Area 1:** displays the main details about the author. By harvesting the data from sources such as Entity Facts, DBpedia, Wikidata, etc., we are able to retrieve essential information about the author. Through Entity Facts the author profile is enriched with the information such as the academic title, alternative spellings, living years and places, short bio, a picture (if available), professions, affiliations, and part of the external links. Authors' pictures, - if available - are provided by Entity Facts in most of the cases, while in other cases we are checking Wikidata with a simple SPARQL Query, where "12365291X" is the GND ID of the selected author.

In addition to the Entity Facts, information from other sources are retrieved. The abstract below the picture originates from DBpedia (denoted as c.), is retrieved by querying the DBpedia SPARQL endpoint<sup>13</sup>. At the same time, DBpedia offers the possibility for other data enrichments. The second element is the citation metrics at the beginning of the author's details. The number of citations (5.480 in this case), the h-index, and i10 are derived from the CitEc service, based on RePEc content (see a.).

Moreover, through Wikidata the author profile integrates several other links and pages related to the author, such as ORCID, Google Scholar, SSRN, and RePEc. By including the Twitter ID, the author's Twitter timeline shows the tweets which in many cases represent the most up-to-date information on recent topics (fig. 4).

**Area 2:** the second area shows the terms and concepts used in the author's research output (a.) and the co-authorship network (part b). The data in this area is generated only from the EconBiz content, i.e. all the publications where the GND ID of the corresponding author is present. The view of terms/concepts is extracted and calculated from titles, abstracts and subjects/keywords of publications by the respective author, as a result of term frequencies methods [6]. Through the usage of a word tag cloud, the representation is visualized by providing an instant overview of the main terms/topics used in the author's research output and indicating the fields in which the author's contribution is most prominent. In addition, by clicking the "Single terms" button, the view is changing to the terms that contain just a single word. Hence, the scholar may perform quick analyses through the offered visualizations.

---

<sup>11</sup> <https://nobelprize.readme.io>

<sup>12</sup> <https://ideas.repec.org/api.html>

<sup>13</sup> <https://dbpedia.org/sparql>



The screenshot displays the EconBiz author profile for William D. Nordhaus. The interface is divided into several sections:

- Top Section:** Includes tabs for 'Terms', 'Co-authors', and 'Related authors'. A word cloud visualizes terms associated with the author, with 'climate change' and 'greenhouse gas emissions' being prominent. Below the word cloud is a search filter section with 'INCLUDE' and 'EXCLUDE' buttons, and a list of terms like 'gases', 'air pollution', 'climate change', 'emissions trading', and 'carbon capture'.
- Author Information Section:** Displays the author's name 'William D. Nordhaus', his title 'Prof. Dr.', and his GND ID '12365291X'. It also shows citation metrics: 'Total Citations: 5,480', 'h Index: 34', and 'i10: 74'. A list of alternative spellings is provided, including 'W. D. Nordhaus', 'U. Nordchautz', 'W. Nordhaus', 'William Nordhaus', 'Wellian Nuodehaosi', and '威廉 诺德豪斯'.
- Biography Section:** States 'B: 1941 Albuquerque, NM' and provides a detailed biography: 'Biblio: Seit 1970 an der Yale University tätig; Associate Professor, 1973-1978 Professor, 1979-1991 John-Musser-Professor, dann A-Whitney-Grisswold-Professor und seit 2001 ist er Sterling-Professor für Ökonomie; Nobelpreis (Ökonomie) 2018'.
- Image Section:** Features a portrait of William D. Nordhaus, labeled 'b.'.
- Publications Section:** Shows a list of publications, with the first three entries highlighted:
  1. Projections and uncertainties about climate change in an era of minimal climate policies (2018, Aufsatz in Zeitschrift, Article in journal, Availability: Link)
  2. Projections and uncertainties about climate change in an era of minimal climate policies (2016, Cowles Foundation discussion paper, Type: Arbeitspapier; Working Paper; Graue Literatur; Non-commercial literature; Availability: Link)
  3. Projections and uncertainties about climate change in an era of minimal climate policies

**Fig 3.** Displaying author information on EconBiz author profiles [https://authors.econbiz.de]

Moreover, this kind of visualization is used in the next steps for narrowing down the list of results and effortlessly finding the publications needed, as presented in part d. of the same area. More details about this feature are provided in the next section, while a general overview can be found in our previous research about visual search in DLs [24].

Part b. of this area two is related to the list of co-authors associated with the selected author, based on EconBiz content where the author's GND ID is used. The collaboration with others is visualized by calculating the frequency of co-authorship. In addition to providing a snapshot, this also provides possible navigation to other authors. The tab option in c., denoted as "Related authors", triggers the view to authors working in related fields by analyzing the entire authors' list.

**Area 3:** lists authors' publications indexed in EconBiz, accessed through the API. As mentioned in the previous section, the proposed approach makes it possible to narrow the result list through different query formulations. In this regard, the use of the word tag cloud from the area 2a provides comfortable search options. Hence, just by clicking on the term/concept, the results will be limited to the records matching the criteria, as for example the selection of concepts "climate change" and "greenhouse gas emission" results in a list of 21 records. In addition to this option, i.e. formulating the search query from the tag-cloud, the scholar is able to include and exclude terms by using the text-box below the tag-cloud. Excluding a term, by inserting a term through the "exclude" button, means to list all the publications where that term does not appear on the indexed fields.

All search terms, added from the tag-cloud or manually through the text-box, appear in a panel below the tag-cloud. From there, the scholar is able to extend the panel with new entries or remove any of them. In addition, the already inserted terms may be extended with related terms through the usage of external thesauri [24]. This feature in our model is set to the hover event, by prompting a box with the most related terms based on the STW thesaurus (see area 2e.). Moreover, the list may be extended with additional machine learning generated terms through word embedding approaches, as we have exposed in our previous studies [6, 24].

Besides the presented features, the list of results could be filtered and sorted using several criteria, through the interactivity in the area 3a.



**Fig 4.** Other author profiles and Twitter timeline

In addition to the content indexed in EconBiz, the approach makes use of different other services to provide more related and up-to-date information. Such a scenario represents the usage of Harvard Business School Working Knowledge<sup>14</sup> for recommending similar articles based on the query that the scholar has formulated.

## 4 Evaluations

Up to the time of writing this text, the proposed approach- developed as fully functional prototype- has been introduced for evaluation to a group of persons reaching 50. The group consists of experts in the field of economics and business, authors that can find themselves in the service, students and other scholars. In addition, a number of subject librarians had a look at the service. All the subjects who participated in the evaluation

<sup>14</sup> <https://hbswk.hbs.edu/>

have provided overall feedback about the service, including the approach, the selection of the data, and the way how the information was presented and visualized. For this purpose, we have followed some of the UX methods for collecting the data, such as interviews, focus groups, and questionnaires. In general, the provided feedbacks reflect the overall design and implementation, while on several occasions the feedbacks are very specific, related to a particular design or functionality.

#### **4.1 Results**

The collected feedbacks are divided into two groups, according to the authors who find themselves in the service, and the others who see the service from users' point of view. All evaluators consider that an approach, as we have presented here, represents a useful instrument to facilitate scientific communication. A positive assessment is given to the integration and display of information, by linking a large number of data about the author. The idea of building a comprehensive profile in a single place, without having to navigate different locations and without input by the author, is one of the benefits in almost every received feedback. In addition to biographical information, the word tag-cloud generated from the most frequently used terms in the author's work is also an important element that brings a new dimension to the profile. Thus, they estimate that the information provided through such visualizations gives good overviews of the areas covered by the author. The co-authorship network is also considered a very important service in order to understand in greater detail the cooperation between authors and to navigate to the relevant author easily. The functionalities to narrow down the results, through query formulations or even within the publication years, represent a crucial component in regard to the interactivity with the service.

There have been several occasions where through the usage of the tool various metadata mistakes were identified (i.e. false authorship, duplicate names, etc.). Hence, such a service can also serve as a tool for quality control of bibliographical records.

Despite the positive reflections from all participants, there are some limitations evident which are emphasized in the following section.

#### **4.2 Limitations**

One of the main limitations that continue to be an obstacle in many cases is the incompleteness of the author's disambiguation with relevant identifiers in all publications where the author's name appears. The application of this approach would be much more productive in an environment where all authors in the corresponding publications and in the entire collection are uniquely identified. In any situation where there is a lack of identifiers, or even when some of the authors have identifiers, the following problems are observed: the list of results may be incomplete i.e., the list of publications may not include the publications where the author is not connected to the identifiers, the visualization of concepts that reflect the author's contribution to the relevant fields and co-authorship networks are not comprehensive. This also makes it difficult to set a default for the EconBiz-search because a search for a name may retrieve too many results not belonging to the author while a GND-search may retrieve few results.

Given the fact that our approach utilizes existing data from different sources, despite many advantages, some deficiencies are present in several situations. Since we do not own the data, we do not have complete control over it, hence the process of updating or adding new facts is complex and update-cycles may take long. Such kind of complexity is present when the Entity Facts data are considered, which are based on GND authority records. Hence, we can positively respond to part of the requests addressed by the authors for changes in the data obtained from that service, however, some detailed changes can only be done by specific GND editors. Data from Wikidata is handled in a completely different manner, where anyone can add or update any statement, normally by following the set rules.

## 5 Conclusion and Future Work

In our approach, the creation of author profiles consists of steps such as: proper identification of an author and a persistent identifier, harvesting data from various sources, presentation of that data in a comprehensive and structured way, as well as the analysis of publications (mainly indexed in EconBiz) to produce visualizations. Such a profile is a useful tool in the hands of scholars and provides a number of benefits, by making numerous clicks and navigation to several other sites superfluous. To this end, scholars can explore the research output of individual authors, get quick overviews on the research interests of an author, navigate to more information like co-authors or related authors, identify other authors who work in a similar field, find publications relevant to their field of interest, etc.

The data harvesting process relies on the linked data principles that attempt to make use of already existing data, instead of requesting from the author/librarian to re-create the same again. This reduces the effort of manually finding and collecting data. If more authority files are used and openly linked in the future, this approach will be even more beneficial to authors and researchers alike.

In order to provide authors with greater control over the presented data, the use of ORCID can be done analogously to GND authority records. In fact, such an idea has been introduced since the beginning of our approach, however, despite the advantages of using ORCID data - created by the authors themselves - it has been found that most authors do not possess an ORCID profile yet, or their profile has a lack of information.

In addition, as future work, we are considering to increase the functionality in terms of adjustment of some visualization. One example would be the adjustment of the word tag-cloud to different time segments, that only the topics of that timeframe would be displayed, e.g. only early publications by the author or only the most recent years.

The creation of topic pages is an idea that can be explored in the future. Such an approach is already applied for generating the COVID-19 page<sup>15</sup>, an experimental proof of concept that allows finding publications or authors publishing on this topic.

---

<sup>15</sup> <https://authors.econbiz.de/1206288906>

## References

1. Tillett, B.B.: Authority Control: State of the Art and New Perspectives. 51–98 (2012). doi:10.4324/9780203051092-8
2. Taylor, A.G.: Authority files in online catalogs : An investigation of their value. *Cat. Classif. Q.* 4, 1–17 (1984). doi:10.1300/J104v04n03\_01
3. Loesch, M.F.: VIAF (The Virtual International Authority File) – <http://viaf.org>. *Tech. Serv. Q.* 28, 255–256 (2011). doi:10.1080/07317131.2011.546304
4. DNB - The Integrated Authority File (GND), <https://www.dnb.de/EN/gnd>
5. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. *Synth. Lect. Semant. web theory Technol.* 1, 1–136 (2011). doi: 10.2200/S00334ED1V01Y201102WBE001
6. Hajra, A., Tochtermann, K.: Linking science: approaches for linking scientific publications across different LOD repositories. *Int. J. Metadata, Semant. Ontol.* 12, 121–141 (2017). doi:10.1504/IJMSO.2017.10011833
7. Gonzales, B.M.: Linking libraries to the web: Linked data and the future of the bibliographic record. *Inf. Technol. Libr.* 33, 10–22 (2014). doi:10.6017/ital.v33i4.5631
8. Zhu, L.: The Future of Authority Control: Issues and Trends in the Linked Data Environment. *J. Libr. Metadata.* 19, 215–238 (2019). doi:10.1080/19386389.2019.1688368
9. Downey, M.: Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context. *J. Libr. Metadata.* 19, 117–136 (2019). doi:10.1080/19386389.2019.1590936
10. Hallo, M., Luján-Mora, S., Maté, A., Trujillo, J.: Current state of Linked Data in digital libraries. *J. Inf. Sci.* 42, 117–127 (2016). doi:10.1177/0165551515594729
11. DNB: Linked Data Service, <https://www.dnb.de/EN/lds>
12. Hannemann, J., Kett, J.: Linked data for libraries. In: *Proc of the world library and information congress: 76 Int’l Federation of Library Associations and Institutions (IFLA)*. 1–11. (2010)
13. DNB - Entity Facts, <https://www.dnb.de/EN/entityfacts>
14. Vierkant, P.: Mehr als 50.000 Personendatensätze der Gemeinsamen Normdatei (GND) mit ORCID-Records verknüpft, <https://www.orcid-de.org/mehr-als-50-000-personendatensatze-der-gemeinsamen-normdatei-gnd-mit-orcid-records-verknuepft/>
15. Adrian Pohl, Fabian Steeg, Pascal Christoph: lobid – Dateninfrastruktur für Bibliotheken. *Informationspraxis.* 4, (2018). doi:10.11588/ip.2018.1.52445
16. RePEc: Research Papers in Economics, <http://repec.org/>
17. Zimmermann, C.: Academic rankings with RePEc. *Econometrics.* 1, 249–280 (2013). doi:10.3390/econometrics1030249
18. CitEc: Citations in Economics, <http://citec.repec.org/>
19. Wikidata, [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)
20. Vrandečić, D., Krötzsch, M.: Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM.* 57, 78–85 (2014). doi:10.1145/2629489
21. Nielsen, F.F., Mietchen, D., Willighagen, E.: Scholia and scientometrics with Wikidata. In: *European Semantic Web Conference*. pp. 237–259. Springer, Cham (2017). arXiv:1703.04222
22. Van Veen, T.: Wikidata: From “an” identifier to “the” identifier. *Inf. Technol. Libr.* 38, 72–81 (2019). doi:10.6017/ital.v38i2.10886
23. Neubert, J.: Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings. In: *NKOS@TPDL 2017*. pp. 14–25. , Thessaloniki, Greece (2017)
24. Hajra, A., Tochtermann, K.: Visual Search in Digital Libraries and the Usage of External Terms. In: *2018 22nd International Conference Information Visualisation (IV)*. pp. 396–400. IEEE (2018). doi:10.1109/iv.2018.00074