Opus
Exploring Publication Data Through Visualizations

Almaha Adnan Almalki
Bachelor in Computer and Information Sciences
King Saud University, 2014

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

June 2018
©Massachusetts Institute of Technology, 2018. All rights reserved.

Signature redacted

Almaha Adnan Almalki
Program in Media Arts and Sciences May 4th, 2018
Massachusetts Institute of Technology

Signature redacted

Certified By

Prof. Cesar A. Hidalgo
Associate Professor in Media Arts and Sciences
Massachusetts Institute of Technology

Signature redacted

Accepted By

Prof. Todd Machover
Academic Head, Associate Professor of Media Arts and Sciences
Massachusetts Institute of Technology

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 27 2018
LIBRARIES
ARCHIVES
Opus
Exploring Publication Data Through Visualizations

Almaha Adnan Almalki

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in Partial Fulfillment of the Requirements for the Degree of Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology

Abstract

Scientific managers need to understand the impact of the research they support since they are required to evaluate researchers and their work for funding and promotional purposes. Yet, most of the online tools available to explore publication data, such as Google Scholar (GS), Microsoft Academic Search (MAS), and Scopus, present tabular views of publication data that fail to put scholars in a social, institutional, and geographic context. Moreover, these tools fail to provide aggregate views of the data for countries, organizations, and journals. Here, we introduce Opus, an interactive online platform that integrates, aggregates, and visualizes publication data from GS to present users with publication data at four different scales (e.g., scholars, countries, organizations, and journals). At each scale, Opus provides benchmarked visualizations that facilitate understanding the work of scholars in a social, generational, geographic, and institutional context. We conducted two user studies with a small group of potential users that show supporting evidence for the benefits of our approach. This design study contributes to the relatively unexplored but promising area of using information visualization to explore publication data.
Opus
Exploring Publication Data Through Visualizations

Almaha Adnan Almalki

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
On May 5th, 2018
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

Signature redacted

Advisor
Prof. Cesar A. Hidalgo
Associate Professor in Media Arts and Sciences
Massachusetts Institute of Technology

Signature redacted

Reader
Prof. Iyad Rahwan
Associate Professor in Media Arts and Sciences
Massachusetts Institute of Technology

Signature redacted

Reader
Dr. Rahul Bhargava
Research Scientist in Media Arts and Sciences
Massachusetts Institute of Technology
Acknowledgments

Cesar Hidalgo my advisor thank you for your endless encouragement, feedback, and support throughout this entire process.

Iyad Rahwan and Rahul Bhargava, my readers, thank you for your support and feedback and for inspiring me over the past year.

To my Media Lab friends and Collective learning group, I am grateful for all of your advice and most importantly your friendship. You have provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. I am fortunate to be a part of such a talented, creative, and intellectually promiscuous society.

Finally, I must express my very profound gratitude to my family, my best friend Lamia, and A for providing me with endless support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.
Contents

Abstract ......................................................... 1
1 Introduction .................................................. 1
2 Literature Review ............................................ 2
3 System ......................................................... 2
   Data .................................................................. 3
      Data Structure .................................................. 3
   Design .................................................................. 3
   Features ................................................................ 4
      Scholar ........................................................... 5
      Country and Organization .................................... 6
      Journal ........................................................... 6
4 Technical Implementation .................................... 6
   Replot .............................................................. 6
5 User Study ........................................................ 7
   Study 1 .............................................................. 7
      Participants ....................................................... 7
      Task and Procedure .......................................... 7
      Results .......................................................... 7
   Study 2 .............................................................. 8
      Participants ....................................................... 8
      Task and Procedure .......................................... 8
      Results .......................................................... 8
6 Conclusion ....................................................... 9

References ......................................................... 9
Opus: Exploring Publication Data Through Visualizations

Almaha Almalki
MIT Media Lab
Cambridge, Massachusetts
almaha@mit.edu

Cesar Hidalgo
MIT Media Lab
Cambridge, Massachusetts
hidalgo@mit.edu

ABSTRACT
Scientific managers need to understand the impact of the research they support since they are required to evaluate researchers and their work for funding and promotional purposes. Yet, most of the online tools available to explore publication data, such as Google Scholar (GS), Microsoft Academic Search (MAS), and Scopus, present tabular views of publication data that fail to put scholars in a social, institutional, and geographic context. Moreover, these tools fail to provide aggregate views of the data for countries, organizations, and journals. Here, we introduce Opus, an interactive online platform that integrates, aggregates, and visualizes publication data from GS to present users with publication data at four different scales (e.g., scholars, countries, organizations, and journals). At each scale, Opus provides benchmarked visualizations that facilitate understanding the work of scholars in a social, generational, geographic, and institutional context. We conducted two user studies with a small group of potential users that show supporting evidence for the benefits of our approach. This design study contributes to the relatively unexplored but promising area of using information visualization to explore publication data.

Author Keywords
Data Visualization; Publication Data; Storytelling; User Interface

INTRODUCTION
Academic research is a global, interdisciplinary, and collaborative endeavor. While the international and interdisciplinary nature of research contributes to the generation and spread of knowledge, managing publication data across geographic and institutional boundaries has become increasingly complex [31]. To search and track the academic literature, for-profit and nonprofit organizations have built software solutions to help organize scholarly publication data. Platforms such as GS, MAS, and Scopus are a few examples of the many ways in which scholars, administrators, and programmer managers access scholarly data to understand the impact of the research they produce and support.

These platforms are popular places for scholars and administrators to find publications and learn about the academic career of scholars. But since the usual format in which the data is presented is through a tabular view, the functionality of these platforms could be extended by providing benchmarked visualizations that acknowledge the social and institutional nature of science. In fact, the use of visualizations is known to enhance the ability of users to understand complex datasets, such as those involved in the overwhelming velocity and volume of publication data [20, 7, 4].

Here, we introduce Opus, an interactive online platform that integrates, aggregates, and visualizes publication data from GS. Opus presents users with publication data at four different scales: scholars, countries, organizations, and journals. Opus provides information about scholars' perceived level of experience and proficiency in social, generational, and geographic perspectives. Opus also provides visualizations that show countries' and organizations' scientific outputs across time and their international research collaboration. For journals, Opus provides functionality to trace their impact and publication activity over time, including their geographical diversity. We conducted two user studies with two
groups of potential users (scientific manager and academic researcher). In the first study, we tested Opus’ performance against GS; we concluded that the use of visualizations and a simple layout helped potential users to retrieve information faster and more accurately. In terms of usability, learnability, and ease of use, we conducted a second user study, and Opus’ simple layout design and clear visualizations performed well.

LITERATURE REVIEW

Data visualization is a powerful technique to encode information using visual representations to discover essential insights from vast complex datasets. Visualizations allow us to quickly record and analyze data visually to find patterns and communicate these findings in a less cognitive effort [2]. The use of visualizations naturally supports the essence of storytelling, which is structuring information to clarify relationships and patterns. Reports also show that the use of visualizations has a potential catalytic effect on storytelling [30]. Historically, visualizations have been used to facilitate analysis and storytelling, such as the famous rose graphs that showed the causes of mortalities during the Crimean War. The graphs were created in 1858 by the English, social reformer and statistician Florence Nightingale to analyze the causes of death in the British army which revealed that most of the British soldiers who died during the Crimean War died of sickness rather than of wounds or other causes. Her use of visualizations helped her to communicate her findings to lobby for sanitary reform and improved conditions in hospitals. She stated the goal of her visualization was: “to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears.”[5]. Her work and others established the principles of a successful data visualization, and these principles remain the same despite the vast change in technology.

In recent years we witnessed the rise of internet-based visualizations, ranging from art projects (e.g., the work of Jonathan Harris, who used a self-organizing particle visualization to explore human emotions, where each particle represents a single emotion posted by a single individual on a global scale [14]) to powerful political stories in the New York Times (Mapping Segregation [27]) [23]. This type of medium is seen by thousands and shared across multiple social media platforms, which has led to creating new opportunities to apply visualizations in data-rich domains such as academia (e.g., publication data). Having such an ecosystem accelerated the use of visualizations for storytelling and allowed us to create both author- and reader-driven experiences by facilitating web-based graphical and interactive elements. Platforms such as the Observatory of Economic Complexity [25] and DataUSA [10] have innovated the way we structure our narratives by using visualization with a high degree of interactivity so that users can ask new questions and discover alternative explanations rather than following the sequential structure of the traditional narrative. These platforms were developed in our group the Collective Learning at MIT Media Lab to enable users to use visualization platforms to create their own stories.

The use of visualizations to explore scientific data can be considered limited. Few tools use visualization as their primary medium to explore publication data. Tools such as GS and MAS have proven the lack of visualizations in their current design. On the other hand, tools such as VOSviewer [28] focus on using one type of visualization (networks) to construct and visualize bibliometric datasets. These networks show journals, researchers, or individual publications, and they can be generated based on citation, bibliographic coupling, co-citation, or co-authorship relations. Scopus also introduces the visual analysis feature in its tool to provide a better picture of a scholar’s publication history and influence. In Opus, we focused on developing a tool that allows analyses of complex data and the presentation of stories. By integrating a suite of interactive visualizations and providing a diverse range of views to allow the exploration of publication data, we create a vibrant ecosystem for users to engage more with publication data and expand the influence of the stories they create.

SYSTEM

Opus is a web-based platform designed to allow users to visually explore publication data. It provides the power needed to promote additional indicators, including collaboration indicators, so as
to change how researchers and scientific managers view, evaluate, and analyze scientific data.

The platform allows users to explore the publication data on four different scales: scholars, countries, organizations, and journals. On each scale, a set of questions are answered by using suitable interactive visualizations. Unlike other research platforms, Opus is designed to help users explore the work and trajectory of scholars and to investigate their academic impact. It also allows one to discover a scholar’s network of collaborators and identify his/her peers.

The platform also provides comprehensive profiles of countries, organizations, and journals to enable users to analyze nations’, organizations’, and journals’ productivity over time and benchmark their output against peers worldwide.

Data

Opus uses the GS dataset; unfortunately, GS does not provide an API for its data; therefore, HTML pages were scraped to extract all available data up to early 2017. The reason for choosing GS as the primary data source is because it covers more publication data than any other platform [19, 15]. An earlier statistical estimate published in PLOS ONE using a mark and recapture method estimated that GS covers approximately 80-90% of all articles published in English with a rating of 100 million [15].

GS data have the advantage of being disambiguated, but they are not free of limitations. One oft-mentioned limitation is that GS indexes all types of publications, including informal and nonacademic documents. We tackle this limitation by only using publications that matched the list of journals provided by SCImago (ultimately, Scopus). Also, GS overestimates the number of citations per publication [12]. However, it has also been demonstrated that some citations in GS at the article level correlate with the Scopus database [18]. Despite these limitations, GS is a great dataset given the complexity of disambiguating scholarly data at the individual level.

After scraping GS, we started disambiguating the data into five different scales of scholars, countries, organizations, journals, and papers. Then we constructed a graph data model to make querying the data more flexible and fine-grained since it allows access to the data from any possible point of view [29]. After structuring and cleaning the data, Opus’ final dataset contained 732,354 scholars, 247 countries, 5,863 organizations, 49,665 journals, and 53M publications.

Data Structure

The Opus database is a graph data model that includes nodes with unique IDs and properties that vary based on node type. For example, the scholar node contains properties such as scholar name, citation count, publication count, etc. Between different types of nodes, relations are defined; for instance, the scholar node has a connection with the publication node established as Author-of, which links a scholar with all of his/her publications. As shown in Figure 1, there are nine nodes and 15 relation types between the nodes.

Design

The interface design of Opus has evolved through numerous iterations in terms of both graphic design and feature specifications, as shown in Figure 2. Opus follows an index page information architecture (IA) pattern that consists of homepage serves as a jump off point for all the other pages. In the initial proof of the concept, the Opus subpages included four views (e.g., scholar, country, field, and visualization); however, as we evolved, we altered our sub-views to reflect the four main pieces of information found in any paper header (e.g., scholar, organization, country, and journal) (see Figure 2.2).
We used this structure to allow users a diverse range of lenses with which to explore the data across various dimensions.

Highly integrated data can be difficult to display in a simple layout. The Opus interface layout used to follow a navigation tabs layout, as shown in Figure 2.1. The navigation tabs layout allowed users to jump between different layers of information in the same page using tabs. After we tested that approach with potential users, we discovered that, by following such an approach, we were adding more complexity and that users were complaining about the number of tabs they needed to click on to display information on a single page. After multiple iterations, we decided to follow a single-page layout that featured multiple functions on a single page and used scrolling controls to fit multiple contents into a smaller area instead of hiding content behind the tabs [9]. The single page layout removed all the clutter from the previous layout (see Figure 2.1.), leaving a clear and beautiful user interface with concise and focused content.

The Opus system interface (see Figure 3) is divided into four sections. The fixed navigation panel (Figure 3.A) at the top of the page allows users to move between the different views in Opus. We utilized the entire screen to only show one visualization at a time and one legend at the center of the screen (Figure 3.C). On the right side, we added a description about the shown visualization, external information, and any user control elements (Figure 3.D). As we populated the pages, we discovered that one of the single page layout cons is that users could become overwhelmed while scrolling the page. This is why we added the fixed navigation (Figure 3.B) menu on the left side to perform as a spy scroll to help users explore our platform better and to encourage scrolling behavior.

As for the visualization, Opus currently contains seven different types of visualizations. We select visualizations based on the data story we wanted to investigate (e.g., relationship, comparison, composition, or distribution) [24]. For example, we used the network to represent the scholar’s coauthor network, since the network shows relationships as a precise combination of the nodes and links. Each visualization presents interactivity by either zooming, hovering, filtering, or clicking on to disclose more information following Ben Shneiderman’s taxonomy for information visualizations [24]. For the graphic design perspective, we chose colors that are visually appealing to the users.

Features
Based on our user and task analysis, we developed a platform that uses visualizations as a medium to reveal trends to provide additional insights. We made sure to involve interfaces that acknowledge the social and institutional nature of science by adding the countries’, organizations’, and journals’ views. We also introduced metrics that can help understand the scholars’ impact on a more narrow context (benchmarked by age and country) to help facilitate comparison.

Scholar
Opus provides information about scholars in a social, generational, geographic, and institutional context. First we presented the user with important numbers about the viewed scholar such as the number of publications and citations to get him/her fa-
familiar with the scholar (see Figure 4.A.1). Users can interact with interactive visualizations to unlock the scholar’s academic career so as to understand his/her main domains through the use of a title cloud visualization that uses the paper’s title to extract the most used words and a treemap that displays the scholar’s fields using Thomson Scientific’s Essential Science Indicators (ESI) database for the field classification of journals [22]. We also displayed generational information about the scholar’s academic output in a simple bar chart that shows the publication count over the years clustered by the fields and another horizontal bar chart to represent the scholar’s publication trajectory based on his/her citation count over time.

Since research is now increasingly done in teams across nearly all fields, available tools lack measurements to quantify that collaboration [31]. Opus provides a way to explore scholars’ social capital by using an intractable network that presents their coauthor network (Figure 4.A.2), where node size represents the number of citation and the link weight represents shared publications. The network also provides information about the diversity of the coauthor network regarding the presented organizations and fields. To explore the scholars’ international research collaboration, we tracked the number of publications coauthored between countries based on coauthor affiliation. We displayed such a relation by using a treemap visualization to display scholars’ collaborators based on organizations and a map graph to display scholars’ collaborators based on countries.

The most used benchmark or metric to measure the success of a scholar is the h-index [1]. However, it can be unfair when you start comparing all scholars without considering their academic age or fields [1]. In Opus, we measured a scholar’s success by comparing his/her citation count or h-index to other scholars that share the same academic age and visualized the output as a histogram (see Figure 4.A.3). Academic age is the latest year that a scholar published a paper in, with at least one citation, minus the year of his/her first published paper. By adding
more controls, users will have a better understanding of the scholar’s performance beyond just a static number.

Country and Organization
Available tools such as GS and MAS tend to focus only on individual-level outputs (presenting a scholar as an individual rather than a part of a country or organization). In Opus, we extend GS and MAS work by introducing new scales (organization and country) to explore publication data in a generational, geographic, and institutional context.

Similar to the scholar profile, we start with a general number to showcase country or organization performance in terms of publication, citation, and the number of scholars. Opus then shows the scientific output of an entity presented in a simple line chart. Based on the publications of an entity, Opus uses a treemap to visualize the diversity of journals colored based on the impact factor based on the SCImago Journal’s 2017 ranking [6]. Users can also explore who the top scholars of an organization are in a country within a scatter plot, where the circle size represents the number of scholars who share the same number of publications or citations based on their academic age.

Opus also provides visualizations (treemap and a map) that show a country’s or an organization’s international collaborators, as shown in Figure 4.B.2. International collaboration creates a broader engagement and status for an organization, helping it to attract more international students and faculties. Also, such information is critical to consider if, for example, an agency wants to fund a research entity. This information helps academic managers to trace the evolution of scientific knowledge at the scale of their country or organization and to understand the growth of specific fields (see Figure 4.C.1), scholars, and publications.

Journal
Opus also acknowledges the importance of visualizing data related to journal performance. We wanted to help researchers quickly find suitable journals that they can be published in, based on, for example, fields, acceptance rate, and impact factor. This is why Opus provides visualizations to show journals’ performance and the number of papers they publish compared to the citations they have received over the years. But, more importantly, users can also explore the top organizations and countries published in a particular journal. In Figure 4.D.2, we show the top organizations published in Nature. This abstract information can be very beneficial for researchers in deciding which journal they need to target.

TECHNICAL IMPLEMENTATION
Opus architecture includes a low-level database layer, a middle-level API layer, and a high-level user interface layer. The database layer consists of a Neo4j-implemented graph database [26]. The API layer queries data from the database layer using GraphiQL [13], and a Ruby connector is implemented to link both layers. When the API layer queries data from the database, it calculates various agreements and visualization metrics and sends them to the interface layer. The user interfaces request data through the API layer upon user interaction. Figure 5 shows an abstract view of the Opus architecture.

The UI is implemented in the JavaScript React framework [11] to enable declaration of the user interfaces and to quickly model the state of those interfaces. It allows the use of reusable, composable, and stateful components, as well as the immediate state changes of a component, and it reflects these changes in the user interface based on user interaction. To visualize information in a treemap, network, bar chart, scatter plot, and word cloud, various visualization libraries are used such as d3plus-react [17] and Replot [21], which is a visualization library that was built in the course of developing Opus.

Replot
Using React as a front-end framework provides a lot of flexibility; however, the problem appears when integrating React and D3 (data-driven documents) seamlessly. Both libraries are built upon data-driven DOM manipulation. Without careful precautions, React will not accept any changes made to the DOM. One solution is creating a DOM node and passing it to D3 on each render. It mutates the detached DOM and is then converted to React elements. However, this means building and mutating a full real DOM
tree on every render and then removing it. This solution is incredibly inefficient because the DOM is not lightweight. As a solution, we created our own native SVG visualizations for React called Replot. Replot is built by only using React and SVG, allowing the instant update of data and customization to almost every feature in a chart. Currently, Replot includes a bar chart, line chart, scatter plot, box plot, treemap, network, and map.

USER STUDY

We conducted two user studies with the objective to assess the potential usability and usefulness of Opus in visualizing and analyzing publication data. The first study had participants answer questions about scholars and their perceived level of experience and proficiency using Opus or GS. In the second study, participants had to answer questions about scholars’, countries’, organizations’, and journals’ scientific performance over time by only using Opus. We also gathered feedback about potential capabilities to add to Opus. We used an approach that is appropriate for studying information visualization-based projects [8, 16, 3] that include four phases (e.g., tutorial, practice, evaluation, and collecting feedback). In our case, we only included the evaluation and gathering feedback phases.

Study 1

Participants
Nine users (5 women and 4 men) participated in the study. Participants’ ages ranged from 25 to 44, and the majority of participants had a graduate or professional degree, except three with a bachelor’s degree. All participants were potential users of Opus for their current work as project managers at MIT International Science and Technology Initiatives. They were familiar with tools similar to Opus such as GS and used them on a regular basis to evaluate the scholars. However, none of them had seen or tried Opus before the study. They did not receive any compensation for participating in the study.

Task and Procedure

Participants were asked to randomly select a number from 1 to 10 to represent their user ID and to determine which tool they would use (Opus or GS). If participants picked an odd number, they would conduct the study using Opus and vice versa. The study included two steps: gathering information and participants filling out a survey to state their demographics. The second step was the evaluation step, where all participants needed to answer the same six questions about a specific scholar and his/her academic performance using the assigned tool. The study was conducted with all participants on one occasion for one hour.

Results

The evaluation phase revealed that participants who used Opus (5 participants) were able to successfully answer most of the questions accurately compared to the participants who used GS (4 participants). Table 1 summarizes participants’ task performance in both tools.

For each evaluation question, there was only one participant who did not answer the question correctly when using Opus. This was compared to GS, where the average number of participants who answered the questions correctly was less than half. We can assume that the use of interactive visualizations to display complex information helped participants to understand better.

Regarding the time, GS and Opus had a similar duration, even though participants didn’t explore Opus before the study. This shows that Opus’ simple page layout helped participants to easily navigate the platform without any significant struggles to find the right information. We believe that if we conducted a study with the tutorial practice phase before the
evaluation phase, users would be able to find answers faster than when using GS. After the study, it was clear that some information was not as easy to find as other information (e.g., most cited paper), and this helped us to further improve the design.

Study 2

Participants

Seven users (5 women and 2 men) participated in the study. Participants were aged from 18 to 44, and the majority of participants had a bachelor’s, graduate, or professional degree. The majority of participants were researchers at the MIT Media Lab, except one who was a relation coordinator and one who was a program specialist. All of the participants were potential users of Opus for their current work, and they were familiar with tools similar to Opus such as GS. However, none of them had seen or tried Opus before the study. Participants received a $10 Amazon gift card as a small gesture of our gratitude for their participation.

Task and Procedure

Participants were asked to randomly select a number from 1 to 10 to represent their user ID. The study included three steps: gathering information and participants filling out a survey to state their demographics. In the second evaluation, we asked participants to answer questions about academic performance (e.g., scholar, country, organization, and journal). Finally, we asked participants to complete a short 12-question survey about their impressions of the system. Each participant received an individual session of 30 minutes.

Results

The overall assessment from the participants was that Opus is a useful platform that helped them publish data in an interesting way.

As a part of the feedback survey, we asked participants to mention at least three aspects they liked about Opus. Their answers can be clustered into three main features: first, the simplicity of Opus’ layout and the appealing graphic design; second, the clarity of visualization and visualization choices; and, third, all participants liked how Opus displays data in four views (e.g., scholar, country, organization, and journal).

To measure the Opus net score, we asked participants how likely it was that they would recommend Opus to a friend or colleague. For the beta version, Opus received a 14.30; however, the net score of GS and MAS was not publicly available. Thus, it is difficult to make sense of the number (see Figure 6).

To measure Opus’ usability, feasibility, and learnability, we asked three questions (see Figure 7). All seven participants found Opus UI to have a simple design and to be intuitive and easy to use/learn. Four
Q1: Was it difficult to find the information you needed in Opus?
Q2: Was Opus easy to use?
Q3: Was Opus easy to learn how to use?

Figure 7. Results of the post-study survey.

out of seven participants particularly liked the fixed navigation menu, which helped them to explore the page content smoothly. Other participants liked the consistency of the design across the multiple views. They also mentioned how the colors and type of visualization used were beautiful and easy to read. Because of these design aspects, participants were able to use the system and learn about the basic functionalities quickly. Overall, Opus falls into the easy to learn/use range (see Figure 7). Responses for the ease of use, learnability, and feasibility statements had a group mean value of 4.38 and 4.71 in question 1, 4.30 in question 2, and 4.14 in question 3, on a scale from 1 to 5, with 5 indicating extremely easy.

To expand Opus' functionality, we asked participants to list features they would like to see on the Opus platform. One participant wanted Opus' search engine to have filters to make the search easier. Another participant wanted to see the information presented in visualization as data tables by adding a switch view. Three out of seven participants wanted to generate a dynamic visualization of the direct comparison of two entities without having to open two tabs for two entities and to compare them manually. All of these ideas were very helpful and thoughtful, and we plan to implement them in the future version of Opus.

CONCLUSION
We have described the development process, design, instantiation, and evaluation of a platform, Opus, for innovatively visualizing publication data. The platform emphasizes that displaying such a dataset needs to be renovated and that it's necessary to packet in such a clear manner. We believe that Opus has significant practical relevance for researchers and academic managers since they lack the tools that allow them to make sense of complex datasets and conduct competitively intelligent decisions.

The contributions of this work include some design ideas that could be employed by others working on a similar dataset, even in different domains. Also, we proposed a new way to disambiguate such a dataset in four different scales (e.g., scholar, country, organization, and journal) to explore the data with acknowledgment of the social nature of science. In the course of developing Opus, we also developed a new visualization library, Replot, to help speed up the developing process of comparable tools using a similar front-end framework.

Our design process and system evaluation uncovered some promising directions for future research, a few of which were described in the previous section. Incorporating other sources of data such as MAS and Scopus, etc., would provide further analytic power and could add a new scale to explore the data (e.g., paper view). The challenge in adding such additional data would be to do so in a manner that does not introduce significantly greater visual complexity and make analytically more difficult.

As with any research, our study does have some limitations. A more extensive study with more participants can help us better understand Opus’ flaws. We also feel there is a benefit to providing more capabilities to Opus (e.g., link publications to real papers and add more ways to filter those data). We also see there is a tremendous opportunity to add one page to Opus that allows users to combine data from different views to enable them to answer multiple questions on the same page by creating dynamic visualizations. All of these suggestions can improve the user experience and enhance Opus’ analysis capabilities. While our platform can easily handle several thousand endpoints, we acknowledge that scalability may be a potential concern. Each of these limitations provides interesting future research opportunities.

REFERENCES
1. Sergio Alonso, Francisco Javier Cabrerizo, Enrique Herrera-Viedma, and Francisco Herrera. 2009. h-Index: A review focused in its variants, computation and standardization for


