Interactive Data Visualization for Decision-making in Government-funded Advanced Research Computing Service

Xiaoyue Cheng [a*], *Yu-Che* Chen [b], *Rich* Knepper [c], *Andrew* Burk [d]

[a] Associate Professor, Department of Mathematical and Statistical Sciences, University of Nebraska at Omaha, USA. Email address: xycheng@unomaha.edu.

[b] Isaacson Professor, Interim School Director, School of Public Administration, University of Nebraska at Omaha, USA. Email address: ychen@unomaha.edu.

[c] Director, Cornell University Center for Advanced Computing, Cornell University, USA. Email address: rich.knepper@cornell.edu.

[d] PhD student, College of Public Affairs and Community Service, University of Nebraska at Omaha, USA. Email address: andrewburk@unomaha.edu.

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 19 May 2025

Abstract. Data-driven decision-making has become more prevalent in the field of digital government. Data visualization as a key tool for data exploration, can help decision makers quickly understand data and grasp insights behind pages of spreadsheets. To better support scientific discovery through advanced research computing, the U.S. National Science Foundation has funded the cyberinfrastructure programs including TeraGrid, XSEDE, and ACCESS for over twenty years. These programs have generated extensive data about the awarded projects along with their resource allocation and usage. However, existing visualization tools for these data were not designed for in-depth analysis or direct decision-making support. This research investigated the needs of resource providers and users, analyzed the data collected from TeraGrid and XSEDE projects between 2003 and 2022, and developed a publicly accessible interactive visualization dashboard. The goal of this visual platform is to enable resource providers and users to independently explore data through graphs and tables and make informed comparisons and decisions. Resource providers can use the platform to study and compare trends in resource allocation and usage across projects funded at various institutions and within different fields, identify more efficient resource users and research directions, and provide personalized services to institutions and principal investigators in need. Resource users can search for potential collaborators and helpers from successful cyberinfrastructure project awardees in similar fields, nearby locations, or related research topics. Additionally, this paper demonstrates the practical application of the visualization dashboard through three examples: a research-intensive institution, an under-represented minority-serving institution, and a junior researcher from a specific field of science, showcasing how each can leverage the platform for decision-making.

Keywords. Data visualization, interactive dashboard, graphical user interface, decision-making, cyberinfrastructure, XSEDE
Research paper, DOI: https://doi.org/10.59490/dgo.2025.946

1. Introduction

In the era of data explosion, making full use of data science can significantly improve comprehensiveness, rationality, timeliness and effectiveness of decision-making. Big data can be employed to depict the comprehensive characteristics of static events and the development trends of dynamic events. This enables decision-makers to more easily identify challenges or issues, determine the most appropriate actions or intervention timings, simulate the potential outcomes and impacts of decisions, and ultimately select the optimal decision under given constraints. However, data science also presents unique challenges, such as how to collect, organize, and store massive amounts of data, how to extract valuable insights from massive data, and how to effectively communicate the summarized data value to policy makers and beneficiaries. Within the chain from data to decision, data visualization serves as a critical link to help stakeholders understand data.

Data visualization uses graphical displays to show data and information (Unwin, 2020). For most people, interpreting information from visual graphs is more direct and efficient than browsing raw data or reading summary tables, without losing accuracy. Data visualization can assist in the decision-making process by presenting important information such as numerical distributions, categorical comparisons, temporal progress, and spatial trends. Statistical graphics can also show comparisons of data distributions under multiple different conditions by selecting correct graphical elements such as graph types, colors, shapes, sizes, and page layouts, etc. Therefore, data visualization can play an essential role in summarizing information and comparing options during decision-making.

In digital government, data-driven decision-making has become more extensive and in-depth (van Ooijen et al., 2019) and the potential of leveraging artificial intelligence (Charles et al., 2022). For instance, statistical data collected by federal, state, and local governments are often publicly available on the government websites like data.gov in the United States. The public can view and understand transparent, data-based government decisions. Meanwhile, individuals may use open data for personal decision-making or to influence government decisions from the bottom up. On government websites, visualization is increasingly being used in the form of data dashboards to provide more convenience to the public.

While significant progress has been made with the growth of data-driven decision-making in digital government, there remain opportunities for further enhancement in some areas like social services (Ryan et al. 2014; van Ooijen et al. 2019; Charles et al. 2022). Although large datasets have been collected in these areas and basic descriptive statistics are often reported, there is a potential to uncover deeper, more valuable insights from the data, such as causal inference, predictive modeling, and equity-focused evaluation. To smooth the decision-making process with data visualization, this study will take the high performance computing services of advanced scientific research projects funded by the U.S. government as a digital government project to showcase the design concept and product examples of interactive visualization platforms; also, to show how the visual products could help both policymakers and potential beneficiaries to explore the value of data and strengthen their analytical and decision-making capabilities.

High performance computing (HPC) is a technology that uses supercomputers or processor clusters to parallelly process massive data and complex calculations at a high speed. HPC is of great significance in accelerating scientific discovery and enhancing technological invention. To provide powerful HPC services for U.S. scientific researchers, the National Science Foundation (NSF) has invested $270 million on the cyberinfrastructure (CI) programs (Chen & Knepper, 2021), including the TeraGrid (Reed & Berman, 2001), the Extreme Science and Engineering Discovery Environment (XSEDE) 1.0 and 2.0 (Towns, et al., 2011; XSEDE, 2016), and the more recent Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) (Boerner, et al., 2023). And NSF's investment has been well paid off. According to Stewart, et al. (2022), the XSEDE program provided a return of more than $1.5 in value for every $1.0 invested by the U.S. Government during 2014 to 2020.

To further develop the HPC CI programs, both the program operation team and users may be interested in addressing questions such as: How to allocate computing resources more efficiently? How to better support inexperienced users to start their HPC projects conveniently? How to advertise the CI programs and encourage more researchers to apply for and utilize these resources and services? How to broaden participation of under-represented individuals and groups? While domain experts may have answers to these questions based on years of experience, insights can also be gained from a data perspective. By analyzing data collected from the TeraGrid/XSEDE projects across almost 20 years, decision makers and scientific users can reveal user behaviors hidden behind the allocation and usage data, offering valuable answers to these critical questions.

The main contribution of this work is to produce an interactive data visualization platform for government-funded HPC cyberinfrastructure services via the user-centered co-design. This platform will provide decision-making support for two groups. For computing service providers, the platform displays and compares the resource allocation and utilization efficiency of research projects using the XSEDE services over the years, aggregated at the institution, field of science, and individual investigator levels, which would assist in finding ways to optimize CI operations. For novice and potential users of HPC services, the visualization platform can help them find potential research collaborators and topics, build the foundation, and boost the confidence to start their HPC projects.

The second section reviews the literature on data visualization and the application of big data in digital government. The third section examines two data visualization tools of the NSF CI programs and introduces the design concept and framework structure of the interactive graphics platform. The fourth section provides a detailed overview of the TeraGrid/XSEDE data and the interactive dashboard, including examples that demonstrate how to utilize the platform. The final section presents a summary of findings and discusses future work.

2. Data Visualization and Decision-making

Data visualization is an important part of Exploratory Data Analysis (EDA). After collecting data, visualizations can be used to examine distributions, outliers, and missing values. This helps assess the discreteness and correlations among variables, handle outliers and missing data, and check the assumptions of statistical models. Moreover, visualizations can uncover key insights or "highlight stories" within the data and effectively communicate statistical results after the analysis is completed.

2.1 Data visualization: types, development, and applications

According to the development stage of data visualization, we divide graphics into static graphics, dynamic graphics, and interactive graphics.

Static graphics are fixed visual displays that do not move or change. They have been used to provide important decision-making information for hundreds of years. For example, in mid-1800s, Florence Nightingale created the polar time series plot to show death rates and causes of British soldiers in the Crimean war (Cohen, 1984). At about the same time, John Snow mapped cholera deaths in London, identifying a water pump as the outbreak's source (Snow, 1856). Modern static graphics usually choose the type of plot based on the types of variables. Univariate distributions can be visualized with histograms or density plots for continuous variables and bar charts for categorical ones; line plots reveal trends in time-series data. For bivariate cases, scatterplots show correlations between continuous variables, mosaic plots display associations between categorical variables, and boxplots or violin plots compare continuous variables across categories. Multivariate relationships can be explored with pairwise plots using univariate visuals on the diagonal and bivariate visuals off-diagonal, or with parallel coordinates plots, where each variable is shown on a vertical axis and observations are connected by lines, revealing patterns such as correlations through line alignment or crossing.

In terms of graphing methods, Cleveland (1984, 1985) discussed the theory of visual perception and the principal elements of graphics. Wilkinson (2005) denoted the grammar of graphics (GoG) system with seven orthogonal class methods to transform data from a raw dataset to a statistical graphic. Wickham (2010) implemented the GoG principles into the R language and developed the well-known graphics package ggplot2. In other software, the design and development of modern graphics packages are also influenced to some extent by GoG and ggplot2 (Morel, 2018; Yim, et al., 2018).

As data volumes grow and software technologies develop, there is a demand for dynamic graphics in the field of visualization. Dynamic graphics are visual displays via a sequence of changing images or animations (Becker, et al., 1987). Like the early film technology, by playing a series of static graphics, users can more intuitively perceive the changing trends or different perspectives in high-dimensional data. Examples of applications include time series, two-dimensional representations of high-dimensional data, visualization on sampling techniques, and so on. For time series data, ordinary static graphics can only display univariate continuous data. If multiple dimensions need to be displayed simultaneously, then each frame can only represent a specific time point. Dynamic graphics, on the other hand, can animate multidimensional data along the time dimension, thereby showing the trends over time. When visualizing high-dimensional data, one challenge is to project high-dimensional data onto a two-dimensional plane. Classic dimension reduction methods have to select optimal projection directions, but dynamic graphics like Grand Tour enable exploration of data by continuously rotating the projection direction (Cook, et al., 1995). Dynamic graphics can also be used to demonstrate sampling techniques by playing static graphics of different random samples, to help students understand concepts more easily in statistical education (Xie, 2013).

Interactive graphics refer to visualizations that allow users to adjust and optimize graphic parameters based on specific needs and highlight interesting areas to gain more effective information perception (McDonald, 1982; Theus & Urbanek, 2008). Modern internet technologies have transformed data visualization from the traditional format where researchers deliver information to users, into a new evolved model that users interactively explore data and customize views from a visualization tool. Common techniques in interactive graphics include selection, deletion, linking, brushing, rotation, scaling, etc. (Cook, et al., 2007) For instance, modifying the bin width of a histogram enables users to shift focus between the overall shape of a numerical variable and local details such as zero values or outliers. Similarly, highlighting a small subset of data points and observing their locations across other variables facilitates deeper understanding of the relationship between variables. An interactive data visualization platform can provide guided graphical options tailored to specific projects and datasets. Users can select datasets, observations, variables, parameters, etc., based on their topics of interest, to generate the optimal static and dynamic graphics that support decision-making. Examples of successful interactive data visualization software include commercial tools like Tableau and PowerBI, as well as open-source software such as R Shiny, Python Dash, and Plotly (Sievert, 2020).

2.2 Data-driven decision-making in digital government

The core values of open government data guide data-driven decision-making in digital government in collaboration with and among organizations and individuals in society. These values are transparency, accountability, and collaboration (Chen, 2017, Ch. 4 on open government, see below). Transparency is about making relevant public policy and service data available to the public. Democratic accountability directs the opening of government data on policies and services to the public. Open government data also creates resources and collaborative opportunities for society, inclusive of organizations and individuals (Janssen, et al., 2015).

Making government data open and available is the first and necessary step to foster transparency, accountability, and collaboration. In the United States, the Data Accountability and Transparency Act of 2014 provides the legal foundation and mandate for the federal government to make data available online. Government can open information and data in three areas to foster transparency: decision-making information such as government procedure, public policy and service data such as budget, and outcome information such as impacts of government expenditure on employment (Grimmelikhuijsen and Welch, 2012).

Government data transparency enhances accountability. Transparency of government policy and service data helps reduce corruption and make the government more accountable (Bertot, et al., 2010). At the local level, the availability of government service information such as the open government data portal for the city of Chicago provides citizenry with information on the variety of government services provided to various districts and neighborhoods. Residents and representatives in these neighborhoods can hold governments accountable for their actions.

Opening government data enables collaboration between government and organizations and individuals in society to create gains in efficiency and effectiveness (Matheus, et al, 2020 and Janssen, et al. 2015). In the United States, the publication of census data and GIS data by the federal government enables citizens and businesses to leverage the government data for economic and efficiency gains. Another mechanism is to have competitions such as hackathons for societal members to create applications leveraging open government data to provide new or improve current public services. The City of New York in the U.S. and the City of Shanghai, China have had annual competition events for such collaboration (Chen, 2017).

Moreover, data science and data visualization for data-driven decision-making enhance the quality and impact of open government data (Matheus, et al. 2020 and Ansari, et al, 2022). The recent advancement in artificial intelligence provides the machine-learning opportunities to analyze a large amount of data to advance public policy-making and public services (Charles, et al., 2022 and Valle-Cruz, et al. 2022). Interactive data visualization enabled by data analytic tools provides organizational and individual users to find answers to public policy and service questions. In addition, such a tool is also used for collaboration between government and citizens in prioritizing government resources such as seen in the area of participatory budgeting at the local level.

Data science and interactive data visualization can advance the values of transparency, accountability, and collaboration that are central to open government data. The use of 311 systems for some major U.S. cities provides residents the ability to track their government service requests and be informed of the local government services for accountability and transparency. The use of data dashboards can enhance collaboration and services such as the 'Waze' app using data from both government and citizens for the City of Rio de Janeiro (Matheus, et al., 2020). The National Science Foundation in the U.S. also has a data dashboard for institutional-based search and analytics to provide useful information.

There are significant opportunities to enhance the usability and use of data visualization to improve the use and usefulness of open government data (Ansari, et al., 2022). These opportunities lie in utilization of qualitative techniques for needs assessment and the involvement of user groups to provide evaluation and input (Ansari, et al., 2022). Acting on these recommendations, our study takes a co-design approach for the development, implementation, and evaluation of the interactive data visualization dashboard.

3. Enabling Visual Exploration on the Cyberinfrastructure Data

3.1 History of the government-funded cyberinfrastructure programs

TeraGrid, XSEDE, and ACCESS are a sequence of CI programs that were funded by NSF. Before TeraGrid, NSF had awarded several programs for large-scale computational infrastructure, including the National Computational Science Alliance and National Partnership for Advanced Computational Infrastructure (NPACI). TeraGrid, operated between 2001-2011, was the earliest distributed terascale facility that coordinated distributed high-performance hardware infrastructures with integrated scalable software and high-speed internet (Reed & Berman, 2001). XSEDE, running between 2011-2022, offered a more powerful virtual organization to provide the dynamic distributed infrastructure and services to scientific researchers (Towns, et al., 2011). ACCESS began in September

2022 and aimed to provide an agile, integrated, robust, trustworthy and sustainable CI ecosystem for computational- and data-intensive research (Boerner, et al., 2023; Parashar, et al., 2022).

3.2 Data platforms for ACCESS and historical programs

To monitor the services provided by the CI ecosystem, the NSF Office of Advanced Cyberinfrastructure funded the development of the XSEDE Metrics on Demand (XDMoD) auditing tools to measure the CI usage and performance metrics on XSEDE allocated resources for the current and historical project data (Furlani, et al., 2013; Palmer, et al., 2015). The XDMoD portal provides graphical user interfaces of data metrics like service units charged, number of jobs, etc. broken down by principal investigator, field of science, institution, resource, and so on (https://xdmod.access-ci.org/)(Palmer, et al., 2015) . For public users, this dashboard provides two tabs: the "Summary" page contains seven plots to display the aggregated usage, measured by computing or cloud resource, CPU hours, job size, etc. The "Usage" page contains more than 120 subpages of plots and data tables, showing the detailed data about jobs, allocations, requests, etc. with the total summary and breakdown numbers at multiple levels. On both pages, users can select a certain time interval and filter the data with 21 service providers.

Besides the XDMoD data platform, the ACCESS program developed its own interactive dashboard to display the geographical distribution of the allocation and usage data on their website (https://allocations.access-ci.org/geo#3.4/38.88/-111.01). On this dashboard, users could see the names and locations of research institutions with the amounts of resource credit allocation and usage, the names and locations of resource providers with the total amounts of used resource credits, and the connections between research institutions and resource providers. By hooking a certain institution, a few arch curves appear to link the institution with the resource providers they used. Reversely, when clicking a resource provider, the dashboard will show linkages to all institutions that are utilizing the service from this provider. With this dashboard, users could learn the usage preference of PIs from given institutions and reveal how the resource providers are supporting the research community.

The two visualization dashboards for the NSF CI programs were developed to facilitate data communication between service providers and resource users. However, both have their own limitations. The XDMoD user interfaces aim to provide a comprehensive view of a large database, requiring it to include and describe every variable, as well as the basic information about each. Even with only the first level of breakdown and without deeper branching, this tool has generated hundreds of pages, which results in an overwhelming amount of information for normal users. For researchers seeking to answer decision-making questions, current visualization platforms offer only basic statistics, like the total allocation units used by a specific institution or field of science. However, they lack the capability to show further interactions between variables. For instance, given a certain group of institutions, what are their strong fields of science regarding the used amount? Do they have some PIs with low usage rate who may need additional help? The ACCESS map platform focuses on showcasing the breadth and impact of the ACCESS program. It displays the total allocation and used amounts by institution and resource but does not offer data details for in-depth analysis.

3.3 Understanding needs from service providers and users

The focus of this study is how to help service providers, i.e., policy makers, understand and utilize data to optimize the operation of the CI ecosystem and solve existing problems. Our team's previous research has explored problems related to the fair allocation of resources and improving resource utilization efficiency. Chen, et al. (2021) built models to identify factors that influenced usage-allocation ratios and compared the ratios across different fields of sciences. In subsequent research, our team interviewed some leaders of the XSEDE program and learned that one of the issues that service providers are interested in is how to raise awareness, interest, and participation among scientific researchers with HPC needs. Hence the target group of this study has expanded from solely service providers to include both service providers and resource users.

From the perspective of service providers, they want to understand whether the training services they provide, such as the Campus Champion program of TeraGrid and XSEDE, are actually effective. Since NSF training projects are usually awarded based on institutions, and different institutions may also offer various computing services independent of NSF, resource utilization based on institutions is more likely to be influenced by decision-making compared to utilization based on a field of science. Therefore, the data visualization platform created by this work will provide service providers with data breakdown and summaries by institution. This will not only help computing centers participating in the NSF CI programs, but also those institutional computing centers to identify users' needs, discover trends, solve problems, and improve service quality.

For resource users, the journey from being unaware of CI projects to generating resource usage involves multiple barriers. First, researchers with HPC needs usually take the initiative to explore existing computing services. They may discover various resources. If NSF's CI program like XSEDE or ACCESS is not the only available option, then researchers can choose the service with the highest input-output ratio for their demands. Therefore, NSF's CI program needs to enhance their promotion of advantages, such as high performance, low cost, and convenient services, to attract users. Next, when researchers become interested in NSF's CI program, they often want to learn about successful cases. If these successful examples are relevant to their own research projects, these potential users can be more motivated. Once they proceed to the application and usage process, when any procedural or technical problems or obstacles arise, whether there exists assistance support from easily accessible helpers to clarify and solve difficulties will significantly impact their user experience.

Moreover, resource users could be divided into the earlier established advantage groups and the emerging, late starter groups. The advantage groups include research teams led by the top experts in their fields, well-funded, research-intensive large institutions, and researchers who joined early in the TeraGrid program or have been using CI resources for years. The starter groups include rising stars in their fields, researchers from small to medium-sized universities with limited support for research activities, first-time CI project applicants, under-represented minority groups, and so on. The advantage groups are often familiar with the application and usage processes, so they do not need much help. However, the starter groups will need extra assistance, otherwise they are more likely to quit the competition when faced with difficulties.

Summarizing the needs of resource users, we believe that they need successful application examples or potential collaborators, as well as helpers to guide them on how to use CI services and hence accelerate their learning process. Traditional decision-making models may select only a few successful cases to inspire all users, but in the data era, the selection of successful examples can be data driven. This approach can filter and match the most suitable success cases for help seekers, based on the same field of science, institution, region, or similar research topics. This study will establish a data visualization platform to provide scientific researchers who need support with recommended potential collaborators and helpers.

3.4 Needs-based visualization design system for decision-making

In this study, we implemented a needs-based design framework to develop an interactive visualization dashboard for decision-making (Fig. 1). This system begins with identifying the specific needs of decision makers. They will then collaborate with domain experts to clearly articulate their needs and the underlying reasons. Domain experts would validate these requirements with decision makers and then find relevant data or collect data by themselves. Once the data is obtained, it is transferred to data experts for filtering, processing, exploration, and visualization. This stage often involves close communication between data experts and domain experts to understand and solve problems in the data. After reviewing a series of static graphics, data experts, guided by domain experts' insights, will gradually refine their ideas to determine which visuals would be most suitable for interactive features, to ensure the final visualization product effectively supports decision makers in exploring the data independently. Finally, data experts complete and deliver the interactive visualization dashboard to decision makers. Decision makers then provide feedback to both data experts and domain experts, to facilitate the debugging, updating, and improvement of the visualization tool.

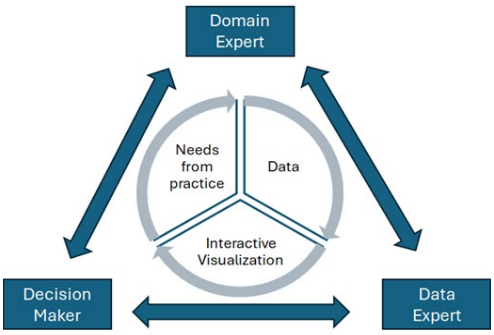


Fig. 1 – Needs-based visualization design system.

4. Data, Application, and Examples

The TeraGrid and XSEDE data are publicly available at the Illinois Data Bank at the University of Illinois at Urbana-Champaign. There are two databases for the allocation award and usage data: the XSEDE (2020) database includes data from the start of TeraGrid in July 2003 to December 2019, while the Towns & Hart (2023) database offers the same data features covering the period from July 2011 to the end of XSEDE in September 2022. Each database consists of two data files: one for awarded CI projects, which includes variables such as grant number, principal investigator (PI) name, institution, field of science, project type, transaction type, project title, abstract, and project start and end dates; and another for detailed resource-level usage information, including variables such as grant number, resource name, transaction type, allocation, used amount, and usage start and end dates.

4.1 Data cleaning and preparation

The two databases were merged as shown in Fig 2. Since the period between July 2011 and December 2019 has overlapped, this segment was excluded from the earlier database. Subsequently, the data from 2003–2011 in XSEDE (2020) was combined with the 2011–2022 data from Towns & Hart (2023). The complete dataset retains two separate files: one for awarded projects and another for resource usage.

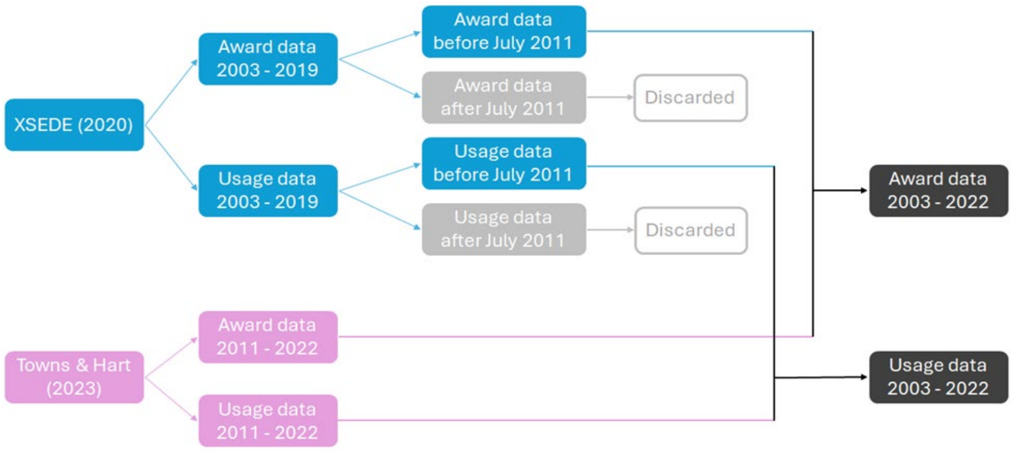


Fig. 2 – Data merging process for two databases, to generate the complete datasets for TeraGrid and XSEDE projects between 2003 and 2022.

The award data comprises 25,798 rows for 16,038 unique grant numbers by 10,277 PIs from 1,044 organizations. The program's peak period occurred between 2016 and 2020, during which nearly 2,000 projects were awarded annually. The usage data includes 85,827 rows covering 16,450 unique grant numbers and 123 unique resource names. Over the 19-year period, the total allocation exceeds 19.5 billion Service Units (SUs), while the total usage surpasses 15.6 billion SUs.

The next step was to merge the award data and the usage data to link the awarded projects with their allocation and usage information (Fig 3). Among the 16,038 unique grant numbers in the award dataset and the 16,450 in the usage dataset, 15,529 grant numbers appeared in both and were retained in the final merged data. The biggest challenge in merging the two data files was compressing information from multiple rows into one row per grant number. While some information about PI and project were duplicated in multiple rows, other variables could have different values for the same grant number. These included fields of science, project types, transaction types, start and end dates, resource names, allocation amounts, used amounts, etc. To resolve this, simple business rules were applied to aggregate the values, such as selecting only one field of science, project type, and transaction type for each project, using the earliest date as the start date, and summing the allocation amount and used amounts from multiple resources. In addition, novel variables were created for each awarded project, such as the number of fields of science, the number of resources utilized, etc.

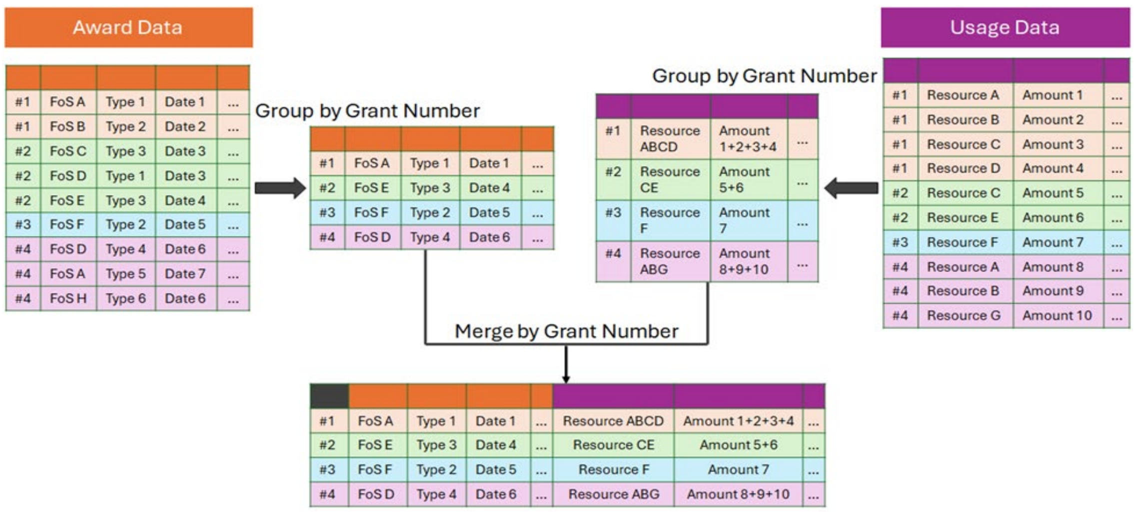


Fig. 3 – Data merging between award and usage datasets. Each row represents one grant project in the merged dataset.

The final step in data cleaning and preparation was to add more useful variables from external data sources, including the longitude and latitude of each searchable organization, generated by the OpenStreetMap API (OpenStreetMap Contributors, 2017), and the Carnegie classification of U.S. institutions published by the Indiana University Center for Postsecondary Research (2021).

4.2 Design and functionality of the interactive dashboard

The interactive visualization dashboard created by this study (https://unodatasci.shinyapps.io/eqci/) was designed with several R packages (R Core Team, 2024) including ggplot2 (Wickham, 2016), shiny (Chang, et al., 2024), flexdashboard (Aden-Buie, et al., 2023), and leaflet (Cheng, et al., 2024). There are two pages for the two groups of decision-makers: an institutional view for the service providers and institutional computing centers, and a researcher's view for the potential service users.

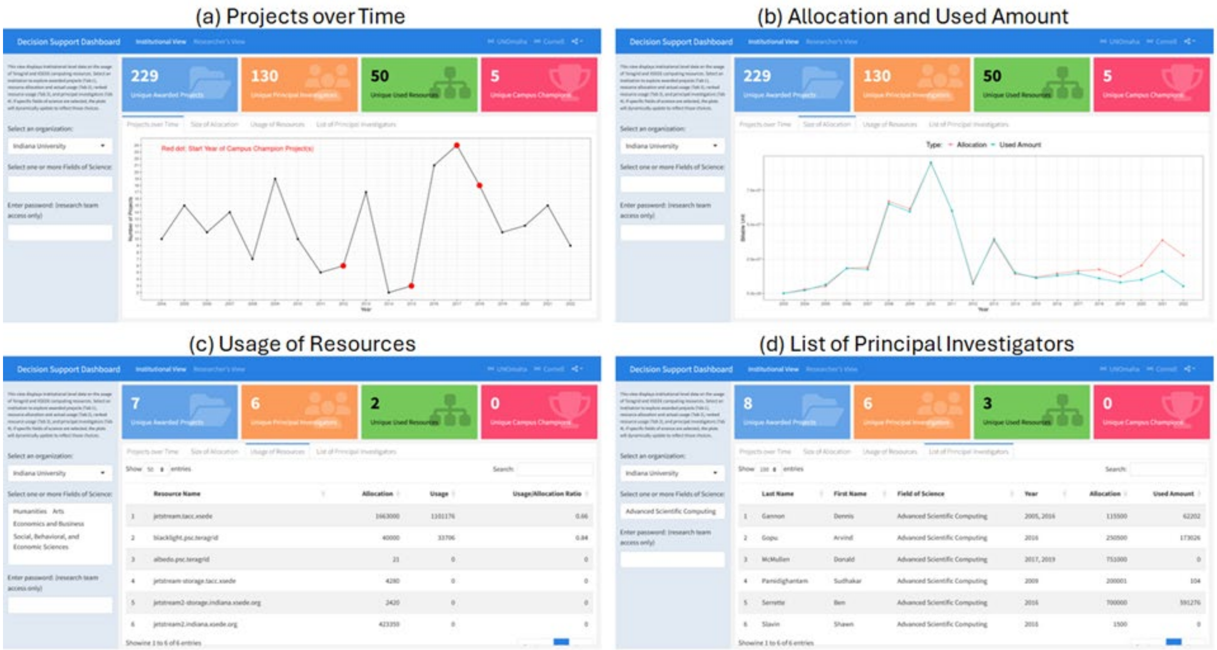


Fig. 4 – Institutional View of the dashboard with Indiana University as an example. Four tabs of visualization can be chosen: (a) Time series plot of the number of awarded projects by year. (b) Time series plot of allocation and used amount comparison. (c) Table of allocation amounts, used amounts, and ratios by resource name for the selected fields of science. (d) Table of the PI's names, fields of science, project start years, allocation, and used amounts.

The design and functionality of the institutional view are shown in Fig 4. The page design includes a title bar, a navigation panel, four data boxes, and a graphical area. The title bar includes the title of the dashboard and the tab buttons of the two views. The left navigation panel contains a page introduction, two data filter boxes, and a password input area. With data filter boxes, users can select the institution and fields of science. This enables users to apply the intersection between the two and to compare results using the generated plots. The password input area is used to protect certain sensitive data, such as the usage rate for each PI. Only the specific research team can enter the password to view sensitive data for further analysis. Above the chart area on the right are four data boxes, representing four key data pieces: numbers of awarded projects, PIs, resource providers, and campus champions. Below the data boxes, users can click to select one chart page. Fig 4(a) shows the number of awarded projects from 2003 to 2022 for the selected institution and field(s) of science. Red dots represent years when the campus champion projects started. This plot helps service providers identify trends in awarded projects for the selected institution and determine whether campus champion projects contribute to an increase in XSEDE projects. Trends of development can be compared by selecting different institutions in the same field of science or the same institution in different fields. Fig 4(b) displays the temporal trends in total allocation (red) and usage (blue). For example, Indiana University's resource usage closely matched allocations from 2003 to 2017. However, starting from 2018 they began to diverge and by 2022, total usage only reached one-fourth of total allocation. These trends may motivate service providers to investigate the reasons behind such discrepancies. Fig 4(c) provides a breakdown of total allocation, used amount, and usage-to-allocation ratio by resource. The table is sorted in descending order of the used amount, but users can sort by allocation or usage-to-allocation ratio by clicking the respective column headers. For instance, four fields of science in Indiana University, including humanities, economics, social, behavioral, etc. had seven awarded projects but only two of the six allocated resources were used. Why these resources were underutilized could be investigated starting from the table. Fig 4(d) presents allocations and used amounts based on PIs. The usage-to-allocation ratio is hidden in this view. By setting thresholds for allocation or used amount, service providers can identify PIs with high or low usage. If further combined with interviews to those PIs, more valuable insights will be obtained to help improve service effectiveness.

The design of researchers' view aims to assist late starters and potential applicants in finding PIs or mentors with successful CI experience in the same field, nearby locations, or with similar research topics. To support this, we provided four search methods: by research field and location, PI name, keywords in project title, and keywords in project abstract. Layout of this page is shown in Fig. 5: the left column contains a page description, search options, and search content settings. The middle column displays the search result of PIs, sorted by project start year from the most recent to the oldest. The upper right column features a map marking the locations of PIs. Clicking on a highlighted dot on the map will pop up the name of the institution, which can then be further used to filter PI search results by institution name. The lower right column provides detailed information about the awarded projects. By clicking on any PI row in the middle column, title and abstract of the corresponding project can be displayed for review.

Fig 5(a) illustrates how to find potential collaborators by the combination of field of science and user's location. The drop-down list for the field of science includes 215 fields from the XSEDE award data, allowing users to choose one area of interest. The search radius drop-down menu has three options: 20 miles, 100 miles, and "everywhere", indicating the same city (20 miles), within a commutable distance (100 miles), or across the entire United States if no local collaborators are available ("everywhere"). Users can enter their location as a city name, institution name, or specific address in the location input box. After clicking the "Confirm" button, the dashboard will search for the location using OpenStreetMap and then mark it with a blue label on the map as the center of the search radius. Yellow dots on the map represent the locations of all PIs in the middle column list, while red dots indicate the locations of selected PIs. Fig 5(b) enables users to search for collaborators by entering a PI's last name, and the results list is sorted by project start year and displayed as yellow dots on the map. To narrow the search result, users can enter an institution name in the search bar of the table. Figs 5(c) and 5(d) allow users to search by keywords in project titles and abstracts. Users can enter one or more keywords, separated by commas, into the keyword input box. Projects that match exactly with all the keywords will be returned, and those that do not match exactly will be sorted by the number of matched keywords, with the top 20 displayed. This keyword search rule ensures users can identify PIs associated with relevant projects from recent years without being overwhelmed by an excessive number of results.

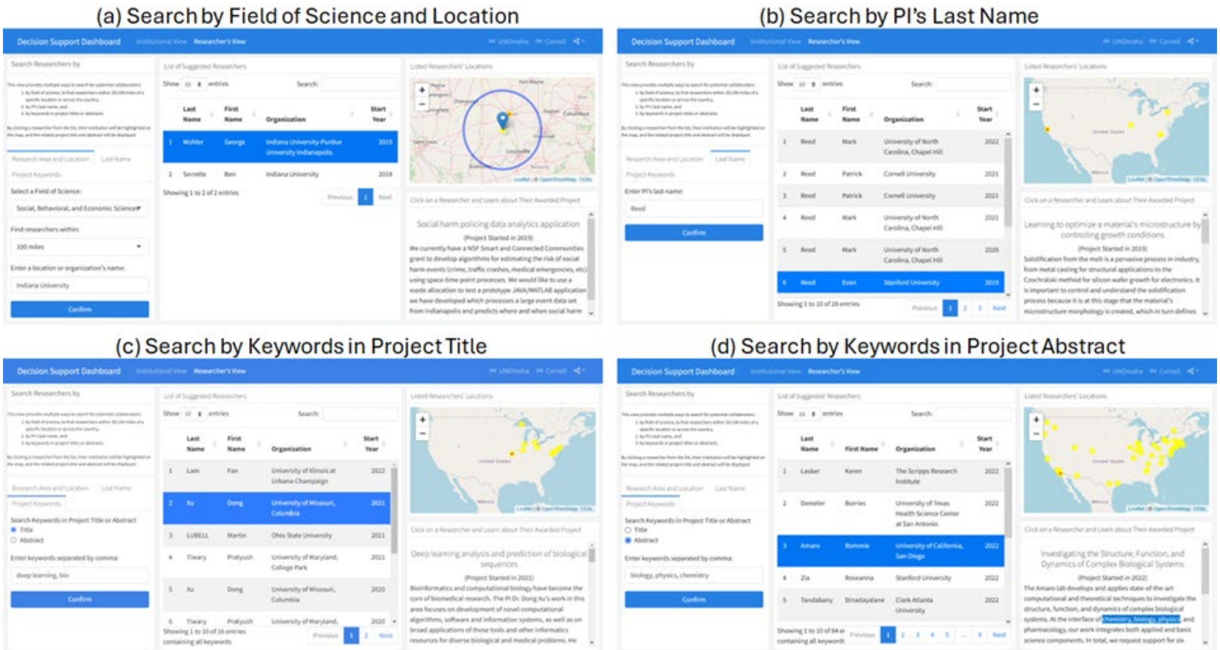
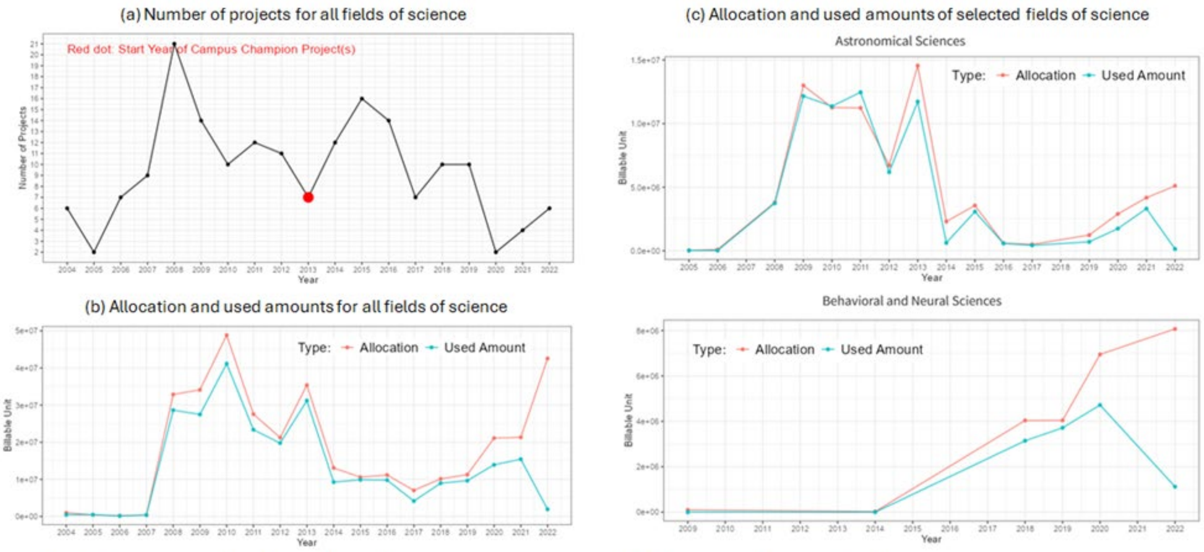


Fig. 5 – Researcher's View of the dashboard. (a) Indiana University is the default searching location. Search radius was set to 100 miles, which generated a blue circle on the map. Clicking the first row of results will display the project title and abstract in the bottom-right panel. (b) By entering the last name "Reed", a list of 28 PIs is obtained with their institutions, locations, and project information. (c) and (d) can match the keywords of interest, separated by commas, with the awarded project titles or abstracts, then corresponding PI names will be shown in the list.

4.3 Example demonstration on the dashboard

Three examples will be demonstrated to explain how the visualization dashboard can assist in decision-making.

Example #1 focuses on a research-intensive institution which also hosts a supercomputing center for the NSF CI program. Researchers at this institution naturally receive early and easy access to training on advanced computing services. We will examine whether any issues need to be addressed here. Example #2 highlights a late-starter institution, specifically an under-represented minority-serving institution. Following the 2021 Carnegie Classification definition on minority-serving institution as "Hispanic Serving Institution, Historically Black College or University, or Tribal College or University only," we matched institution names between the Carnegie Classification and CI data, and identified 96 institutions. One of these, with a moderate level of resource usage, will be explored. Example #3 considers a Midwest junior researcher in Social, Behavioral, and Economic Sciences. If this researcher seeks a potential collaborator or mentor in the field, we will demonstrate how the dashboard can be used to facilitate this search.



(d) Computing resources ordered by used amount

	Resource Name	Allocation	Usage	Usage/Allocation Ratio
1	ranger.tacc.teragrid	132786339	116234599	0.88
2	stampede.tacc.xsede	52091302	48247878	0.93
3	jetstream.tacc.xsede	21795054	16528202	0.76
4	kraken.nics.teragrid	16274125	16412450	1.01
5	comet.sdsc.xsede	17497071	14737523	0.84
6	expanse.sdsc.xsede.org	27675098	10822504	0.39
7	lonestar4.tacc.teragrid	14250501	10679908	0.75
8	bridges2-rm.psc.xsede.org	8461353	6198967	0.73
9	stampede2.tacc.xsede	6826629	5499208	0.81
10	staff.teragrid	7929999	3709158	0.47
11	anvil.purdue.xsede.org	24119410	1689822	0.07
12	trestles.sdsc.teragrid	1146915	1022280	0.89
13	lonestar.tacc.teragrid	1526000	947159	0.62
14	lemieux.psc.teragrid	896323	596368	0.67
15	abe-queenbee-steele.teragrid	1249029	497091	0.4
16	bridges.psc.xsede	472603	455650	0.96
17	teragrid_roaming	415020	381142	0.92

(e) Principal investigators ordered by used amount

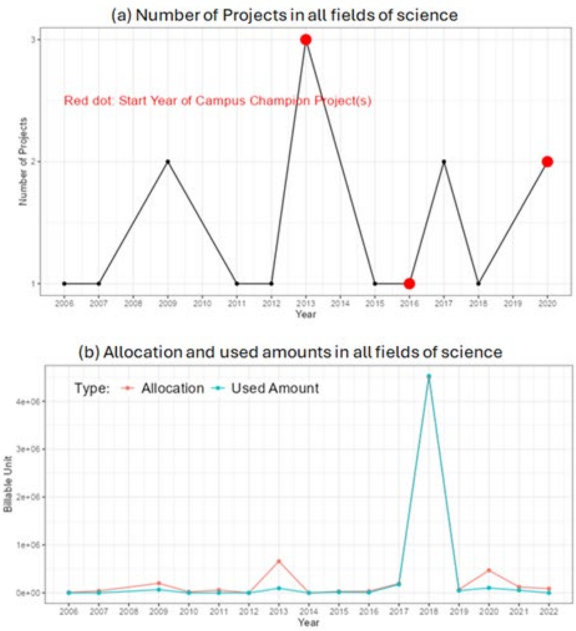
Last Name	First Name	Field of Science	Year	Allocation	Used Amount
Shapiro	Paul	Astronomical Sciences	2006, 2008, 2019, 2020	161948374	151953698
Shapiro	Paul	Astronomy and Planetary Sciences	2021	161948374	151953698
Shapiro	Paul	Extragalactic Astronomy and Cosmology	2009, 2010, 2011, 2012, 2013, 2015, 2017	161948374	151953698
Aluru	Narayana	Advanced Scientific Computing	2019, 2020	101275830	74333977
Aluru	Narayana	Cross-Disciplinary Activities	2013	101275830	74333977
Aluru	Narayana	Interfacial, Transport, and Separations Processes	2016	101275830	74333977
Aluru	Narayana	Materials Research	2015, 2018	101275830	74333977
Aluru	Narayana	Nanotechnology	2021	101275830	74333977
Laguna	Pablo	Astronomy and Planetary Sciences	2022	91376038	66291779
Laguna	Pablo	Gravitational Physics	2012, 2013, 2014, 2016, 2017, 2018, 2019, 2020	91376038	66291779
Laguna	Pablo	Physics	2015	91376038	66291779
Chelikowsky	James	Materials Engineering	2021	63208954	60536849
Chelikowsky	James	Materials Research	2008, 2009, 2010, 2011, 2012, 2013, 2015, 2016, 2017, 2019, 2020	63208954	60536849
Chelikowsky	James	Physics	2008	63208954	60536849
Pestilli	Franco	Behavioral Sciences Biology	2022	46625568	25667256
Pestilli	Franco	Behavioral and Neural Sciences	2018, 2019, 2020	46625568	25667256
Pestilli	Franco	Neuroscience Biology	2016, 2017	46625568	25667256

Fig. 6 – Data exploration for University of Texas at Austin. (a) Time series plot for the number of awarded projects in all fields of science. (b) Time series plot for the allocation and used amounts from all fields of science. (c) Comparison between two fields of science: Astronomical Sciences and Behavioral and Neural Sciences. (d) Top 17 computing resources ordered by their used amounts. (e) PIs with the highest used amounts.

<u>Research-intensive institution</u>. University of Texas at Austin (UTA) is the host of Texas Advanced Computing Center (TACC), which provides big convenience for computing resource users in the university and local area. By selecting the university name in the dashboard, users can obtain the plots shown in Fig 6 (a), (b), (d), and (e) from the Institutional View. Since 2004, UTA has received a total of 180 unique projects, including one campus champion project, involving 119 PIs. These projects have used 44 different resources. According to Fig 6 (a), the number of projects fluctuated over the years, peaking in 2008 before declining, and has remained at a relatively low level since 2020. Fig 6 (b) shows the changes in allocation and usage amounts over time. From 2008 to 2013, both allocation and usage were at their highest levels, followed by a decline between 2014 and 2019. After 2020, allocation increased again, reaching another peak in 2022. However, the used amount did not increase accordingly

but instead declined, which is a concerning trend. Fig 6 (c) compares UTA's two strong research fields: Astronomical Sciences and Behavioral and Neural Sciences. These fields reveal very different allocation and usage patterns. Astronomical Sciences used a large amount of resources from 2009 to 2013. In 2010 and 2011, the used amount was even higher than the allocation. But their usage dropped significantly after 2014. In contrast, Behavioral and Neural Sciences were relatively underdeveloped before 2018, with low project numbers and resource usage. However, after 2018, this field grew rapidly, with resource allocation steadily increasing. Fig 6 (d) compares the usage of different resources. The top three most popular resources all belong to TACC, and the leading resources generally receive high usage/allocation ratios. Fig 6 (e) lists the PIs with the largest allocation and usage amounts, along with their research fields. In some strong fields like Astronomical Sciences, UTA often has multiple PIs with awarded CI projects, and the usage/allocation ratios are relatively high. However, in fields with fewer PIs, such as Biological Sciences and Engineering Systems, the resource usage rate tends to be lower. This pattern suggests that internal collaboration within the research field may influence resource usage significantly.

Minority-serving institution. Florida International University (FIU) is a Hispanic-Serving Institution. Compared to other minority-serving institutions, FIU has a moderate level of experience with CI projects. Since 2006, the university has undertaken a total of 16 unique projects led by 10 PIs. As revealed in Fig 7 (a) and (b), FIU had no more than three new projects per year, and resource usage remained low for most years, except for a significant spike in 2018. This increase was driven by a Nuclear Physics PI who received an award that year (Fig 7 (d)). Most of FIU's allocations and resource usage came from the San Diego Supercomputer Center (SDSC), as shown in Fig. 7 (c). FIU's other strong research fields include computer science, earth science, and networking and communications, as indicated by the list of PIs with the highest resource usage. The usage-to-allocation ratios were below 0.5 for nearly all projects, except for three physics-related projects. Note that FIU received three campus champion projects in 2013, 2016, and 2020, all granted to the same PI. The first two projects were in the field of training, while the third was in physics. However, the allocated resources for these projects were ultimately not utilized, raising questions about the effectiveness of the campus champion projects for this institution.



(c) Computing resources ordered by used amount

	Resource Name	Allocation	Usage	Usage/Allocation Ratio
1	comet.sdsc.xsede	4835728	4768581	0.99
2	stampede.tacc.xsede	167000	67413	0.4
3	abe-queenbee.teragrid	59840	61185	1.02
4	bridges.psc.xsede	50000	51997	1.04
5	supermic.cct-lsu.xsede	50000	48649	0.97
6	expanse.sdsc.xsede.org	121000	25607	0.21
7	gordon.sdsc.teragrid	28001	18454	0.66
8	kraken.nics.teragrid	39672	15571	0.39

(d) Principal investigators ordered by used amount

Last Name	First Name	Field of Science	Year	Allocation	Used Amount
Narayanan	Rajamani	Elementary Particle Physics	2017	4655563	4680743
Narayanan	Rajamani	Nuclear Physics	2018	4655563	4680743
Saeed	Fahad	Algorithm Development	2020	822437	402756
Saeed	Fahad	Computer Science	2021, 2022	822437	402756
Saeed	Fahad	Computer and Computation Research	2015, 2016, 2017, 2019	822437	402756
Liu	Xiaowen	Computer and Computation Research	2012, 2013	583396	94896
Liu	Xiaowen	Networking and Communications Research	2007	583396	94896

Fig. 7 – Data exploration for Florida International University. (a) Time series plot for the number of awarded projects in all fields of science. (b) Time series plot for the allocation and used amounts from all fields of science. (c) Top 8 computing resources ordered by their used amounts. (d) PIs with the highest used amounts.

Junior researchers or potential resource users. A junior researcher in the Midwest may face challenges finding a local collaborator in Social, Behavioral, and Economic Sciences who uses computing services frequently. In states like Nebraska or Iowa, there may be no such collaborators within 100 miles. To address this, the researcher should first select "everywhere" in the Researcher's View (Fig 8(a)) to check the locations of researchers. This reveals some potential collaborators near Chicago, making it a suitable search location (Fig 8(b)). The map highlights three universities: the University of Chicago, Northwestern University, and the University of Wisconsin–Milwaukee. The dashboard would also list corresponding PI names from these institutions. At Northwestern, one researcher specializes in large network modeling and has been awarded four CI projects between 2015 and 2019. The University of Chicago has a researcher in socioeconomic modeling, while Wisconsin-Milwaukee has one studying CEOs outside directorships. If none of these projects meet the junior researcher's interests, she can expand the search to nearby regions, using "Indiana" or "UIUC" as the search location. If choices in the Midwest are still too limited, then keyword-based searches can help. Searching "government" yields two results in project titles (Fig 8(c) left) and 176 in abstracts, with locations shown on the map (Fig 8(c) middle). Clicking on a location, such as the University of California, Los Angeles, allows further exploration of potential collaborators at this institution and details of their research projects.

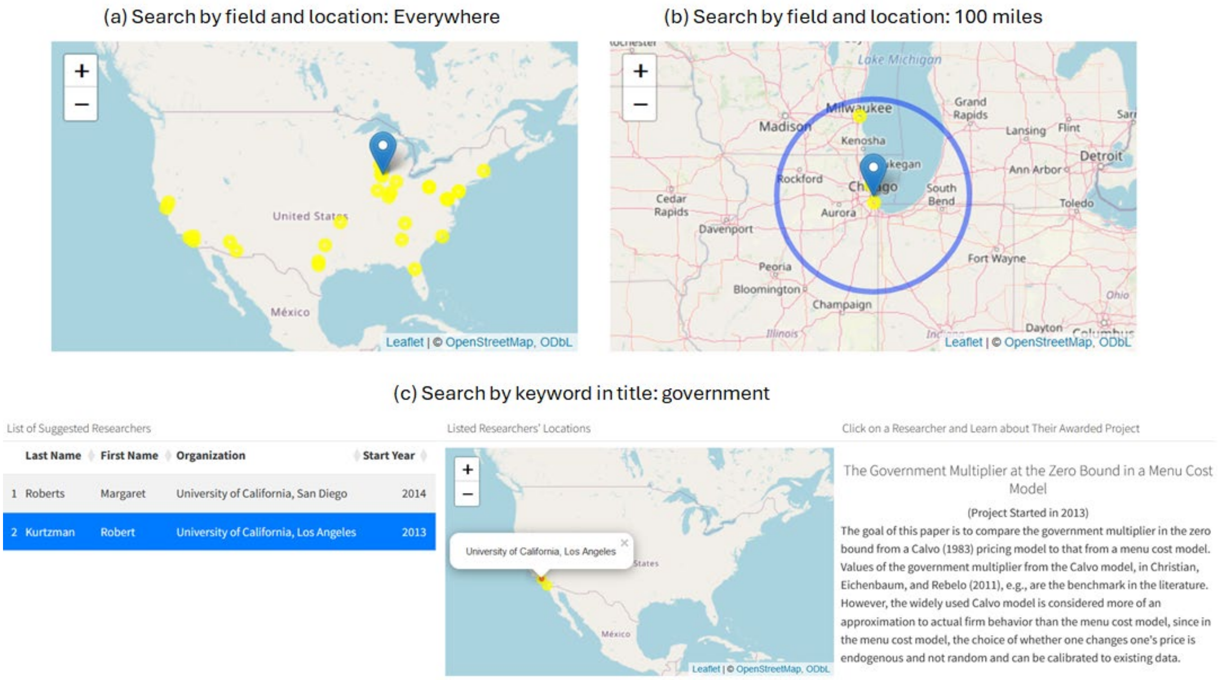


Fig. 8 – Searching for collaborators using multiple approaches. (a) and (b) used Chicago as the location and Social, Behavioral, and Economic Sciences as the field, with different search radius settings. (c) used "government" as a keyword to search project titles, yielding two results.

Through these three examples, we hope the interactive visualization dashboard users can learn how to effectively utilize the graphs and tables to search for useful information, identify potential issues, make comparisons, and support their decision-making process.

5. Conclusions and Future Work

This research reviewed the historical development of data visualization and its applications in the digital government domain, proposed a collaboration model among decision makers, domain experts, and data experts, and applied the model to the field of advanced research computing. Since 2003, the U. S. government has funded cyberinfrastructure programs such as TeraGrid, XSEDE, and ACCESS to promote scientific discovery. These programs have generated a large amount of data on the awarded projects with their resource allocation and usage. However, existing visualization tools for these data were not designed for in-depth analysis or direct decision-making support. This study investigated the needs of resource providers and users, analyzed the data collected from TeraGrid and XSEDE projects (2003-2022), and developed a publicly available interactive visualization dashboard. This dashboard enables resource providers and users to independently explore data through graphs and tables and make informed comparisons and decisions. In addition, this study demonstrates the practical application of the dashboard through three examples: a research-intensive institution, an under-represented minority-serving institution, and a junior researcher from a specific field of science. These examples highlight how data visualization could support and enhance decision-making. A broader impact of this study is the introduction of a collaboration model that does not only apply to the cyberinfrastructure and advanced computing area, but also to other digital government fields that generate and utilize data.

There are some shortcomings in research design and data processing procedures. First, when designing this visualization tool, our perspective mainly focused on the specific needs of CI service providers and potential users: to identify user groups with successful experiences, analyze their models, connect them with the groups who could benefit from these proven models, and improve the service of resource providers within the entire network. However, we may have overlooked the other needs of these two groups, or needs of other groups, such as the long-term, high-ability users of CI services. Second, the data aggregation step combined multiple rows of data into a single row based on the grant number, to merge award data and usage data (Fig. 3). This process resulted in some information loss. For example, a campus champion project may be renewed several times over years after the initial award. However, since the grant number remains the same, the data merging process kept only the most recent renewal and removed all the previous records. As a result, it is impossible to visualize whether this campus champion project has been supporting other projects of this institution in early years or not. If the resource usage pattern of a project changed before and after renewal, the data merging approach would also obscure the information. Third, due to the limitations of the shiny package itself, the user graphical interface of our platform may appear differently on different web browsers or operating systems. The display on mobile devices, such as smartphones, may not be convenient for users to operate.

In the future, the interactive visualization dashboard will be improved in two aspects. From the data perspective, this platform currently only includes data from TeraGrid and XSEDE between 2003 to 2022 and does not involve data from the ACCESS program. We will use the XDMoD API to extract more detailed job-level data generated by ACCESS users when submitting job requests and receiving output from computing resources. This will help identify potential barriers and problems that resource users may encounter during their usage. From the user experience perspective, we plan to interview researchers from different types of institutions to better understand their actual needs. We will have them test the visualization dashboard and improve the design and functionality based on their feedback. Additionally, we will promote the dashboard through workshops and other outreach efforts to expand its user group and serve more people for decision making with cyberinfrastructure data.

Acknowledgement

- **Funding or Grant:** We acknowledge the support of the U.S. NSF Grant Award 2346631.
- **Data/Software Access Statement**: Datasets used for this research can be downloaded at https://databank.illinois.edu/datasets/IDB-0757799 and https://databank.illinois.edu/datasets/IDB-3731847. The visualization dashboard developed by the authors are available at https://unodatasci.shinyapps.io/eqci/.
- **Contributor Statement***:
  - Xiaoyue Cheng: Conceptualisation, Formal analysis, Methodology, Software, Validation, Visualisation, Writing-Original Draft
  - Yu-Che Chen: Conceptualisation, Methodology, Project administration, Supervision, Writing – Review & Editing
  - Rich Knepper: Conceptualisation, Methodology, Supervision, Writing - Review & Editing
  - Andrew Burk: Writing – Review & Editing
- **Use of AI***: During the preparation of this work, the author(s) used ChatGPT in order to improve grammar and refine the language. After using this tool/service, the author(s) reviewed, edited, made the content their own and validated the outcome as needed, and take(s) full responsibility for the content of the publication.
- **Conflict Of Interest (COI)***: There is no conflict of interest.

References

Aden-Buie, G., Sievert, C., Iannone, R., Allaire, J., & Borges, B. (2023). *flexdashboard: R Markdown Format for Flexible Dashboards (Version 0.6.2)* [R]. https://doi.org/10.32614/CRAN.package.flexdashboard

Ansari, B., Barati, M., & Martin, E. G. (2022). Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research. *Government Information Quarterly*, 39(1), 101657. https://doi.org/10.1016/j.giq.2021.101657

Becker, R. A., Cleveland, W. S., & Wilks, A. R. (1987). Dynamic graphics for data analysis. *Statistical science*, 2(4), 355-383.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271. https://doi.org/10.1016/j.giq.2010.03.001

Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., & Towns, J. (2023). Access: Advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: Services & support. In *Practice and Experience in Advanced Research Computing* (pp. 173-176).

Brazil, M., Brunson, D., Culich, A., DeStefano, L., Jennewein, D., Jolley, T., Middelkoop, T., Neeman, H., Rivera, L., Smith, J., & Wernert, J. (2019). Campus champions: Building and sustaining a thriving community of practice around research computing and data. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)* (pp. 1-7).

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R (Version 1.10.0)* [R]. https://doi.org/10.32614/CRAN.package.shiny

Charles, V., Rana, N. P., & Carter, L. (2022). Artificial Intelligence for data-driven decision-making and governance in public affairs. *Government Information Quarterly*, 39(4), 101742. https://doi.org/10.1016/j.giq.2022.101742

Chen, Y.-C. (2017). *Managing digital governance: Issues, challenges, and solutions*. Routledge, Taylor & Francis Group.

Chen, Y.-C., & Knepper, R. (2021). US cyberinfrastructure for scientific innovation: adaptation, management, and performance. In *Research Handbook on E-Government* (pp. 222-241). Edward Elgar Publishing.

Chen, Y.-C., Cheng, X., & Knepper, R. (2021, June). Performance of US Scientific Research Cyberinfrastructure: Structural and Relational Factors for Usage. In *DG. O2021: The 22nd Annual International Conference on Digital Government Research* (pp. 424-436).

Cheng, J., Schloerke, B., Karambelkar, B., & Xie, Y. (2024). *leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library (Version 2.2.2)* [R]. https://doi.org/10.32614/CRAN.package.leaflet

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), 531-554.

Cleveland, W. S. (1985). *The elements of graphing data*. Wadsworth Advanced Books and Software.

Cleveland, W. S. (1993). A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4), 323-343.

Cohen, I. B. (1984). Florence nightingale. *Scientific American*, 250(3), 128-137.

Cook, D., Buja, A., Cabrera, J., & Hurley, C. (1995). Grand tour and projection pursuit. *Journal of Computational and Graphical Statistics*, 4(3), 155-172.

Cook, D., & Swayne, D. F. (with Buja, A., Lang, D. T., Hofmann, H., Wickham, H., & Lawrence, M.). (2007). *Interactive and Dynamic Graphics for Data Analysis: With R and GGobi* (2007 edition). Springer.

Edwards, P. N., Jackson, S. J., Bowker, G., & Knobel, C. (2007). Understanding Infrastructure: Dynamics, Tensions, and Design. National Science Foundation. https://deepblue.lib.umich.edu/handle/2027.42/49353

Furlani, T. R., Schneider, B. L., Jones, M. D., Towns, J., Hart, D. L., Gallo, S. M., DeLeon, R. L., Lu, C.-D., Ghadersohi, A., Gentner, R. J., Patra, A. K., von Laszewski, G., Wang, F., Palmer, J. T., & Simakov, N. (2013). Using XDMoD to facilitate XSEDE operations, planning and analysis. *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*, 1–8. https://doi.org/10.1145/2484762.2484763

Grimmelikhuijsen, S. G., & Welch, E. W. (2012). Developing and Testing a Theoretical Framework for Computer-Mediated Transparency of Local Governments. *Public Administration Review*, 72(4), 562–571.

Indiana University Center for Postsecondary Research (2021). The Carnegie Classification of Institutions of Higher Education, 2021 edition, Bloomington, IN. https://carnegieclassifications.acenet.edu/

Janssen, M., Matheus, R., & Zuiderwijk, A. (2015). Big and Open Linked Data (BOLD) to Create Smart Cities and Citizens: Insights from Smart Energy and Mobility Cases. In E. Tambouris, M. Janssen, H. J. Scholl, M. A. Wimmer, K. Tarabanis, M. Gascó, B. Klievink, I. Lindgren, & P. Parycek (Eds.), *14th International Conference on Electronic Government (EGOV)*: Vol. LNCS-9248 (pp. 79–90). https://doi.org/10.1007/978-3-319-22479-4_6

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, 37(3), 101284.

McDonald, J. A. (1982). INTERACTIVE GRAPHICS FOR DATA ANALYSIS (Technical Report No. SLAC-253). Stanford Linear Accelerator Center, Stanford University. https://purl.stanford.edu/tg751vg9206

Morel, P. (2018). Gramm: grammar of graphics plotting in Matlab. *Journal of Open Source Software*, 3(23), 568.

National Science Foundation's Cyberinfrastructure Council. (2007). *Cyberinfrastructure Vision for 21st Century Discovery*. https://www.nsf.gov/pubs/2007/nsf0728/

NSF. (2016). NSF awards $110 million to bring advanced cyberinfrastructure to nation's scientists, engineers. Retrieved from https://www.nsf.gov/news/news_summ.jsp?preview=y&cntn_id=189573

OpenStreetMap Contributors. (2017). Planet dump retrieved from https://planet.osm.org [Dataset]. https://www.openstreetmap.org/

Palmer, J. T., Gallo, S. M., Furlani, T. R., Jones, M. D., DeLeon, R. L., White, J. P., Simakov, N., Patra, A. K., Sperhac, J., Yearke, T., Rathsam, R., Innus, M., Cornelius, C. D., Browne, J. C., Barth, W. L., & Evans, R. T. (2015). Open XDMoD: A Tool for the Comprehensive Management of High-Performance Computing Resources. *Computing in Science & Engineering*, 17(4), 52–62. https://doi.org/10.1109/MCSE.2015.68

Parashar, M., Friedlander, A., Gianchandani, E., & Martonosi, M. (2022). Transforming science through cyberinfrastructure. *Communications of the ACM*, 65(8), 30-32.

R Core Team. (2024). R: A Language and Environment for Statistical Computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Reed, D. A., & Berman, F. (2001). The TeraGrid: Cyberinfrastructure for 21st Century Science and Engineering. National Science Foundation. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=58fba1297dfcd11783f37674770dff10c80435a7

Ryan, G. W., Bloom, E. W., Lowsky, D. J., Linthicum, M. T., Juday, T., Rosenblatt, L., Kulkarni, S., Goldman, D., & Sayles, J. N. (2014). Data-driven decision-making tools to improve public resource allocation for care and prevention of HIV/AIDS. *Health Affairs*, 33(3), 410-417.

Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and shiny (1st edition)*. Chapman and Hall/CRC.

Snow, J. (1856). On the mode of communication of cholera. *Edinburgh medical journal*, 1(7), 668.

Stewart, C. A., Costa, C. M., Wernert, J. A., Hancock, D. Y., McMullen, D. F., Blood, P., Sinkovits, R., Mehringer, S., Knepper, R., Fischer, J., Bland, M., Rogers, G., Couvares, P., Campbell, T., Jankowski, H., Snapp-Childs, W., & Towns, J. (2022). Metrics of financial effectiveness: Return On Investment in XSEDE, a national cyberinfrastructure coordination and support organization. *Practice and Experience in Advanced Research Computing 2022: Revolutionary: Computing, Connections, You*, 1–9. https://doi.org/10.1145/3491418.3530287

Theus, M., & Urbanek, S. (2008). *Interactive graphics for data analysis: principles and examples*. CRC Press.

Towns, J., Roskies, R., Boisseau, J., Kovatch, P., & Andrews, P. (2011). XSEDE: eXtreme Science and Engineering Discovery Environment. *National Science Foundation*.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., & Wilkins-Diehr, N. (2014). XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering*, 16(5), 62–74. https://doi.org/10.1109/MCSE.2014.80

Towns, J., & Hart, D. (2023). XSEDE: Allocations Awards and Usage for the NSF Cyberfrastructure Portfolio, 2004-2022. https://doi.org/10.13012/B2IDB-3731847_V1

Unwin, A. (2020). Why is data visualization important? What is important in data visualization? *Harvard Data Science Review*, 2(1), 1.
Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, 37(4), 101509. https://doi.org/10.1016/j.giq.2020.101509
van Ooijen, C., Ubaldi, B., & Welby, B. (2019). *A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance* (No. No. 33; OECD Working Papers on Public Governance). OECD Publishing. https://doi.org/10.1787/09ab162c-en
Wickham, H. (2010). A layered grammar of graphics. *Journal of computational and graphical statistics*, 19(1), 3-28.
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis (2nd ed.)*. Springer.
Wilkinson, L., Wills, D., Rope, D., Norton, A., & Dubbs, R. (2005). *The Grammar of Graphics (2nd edition)*. Springer.
Xie, Y. (2013). animation: An R package for creating animations and demonstrating statistical methods. *Journal of Statistical Software*, 53, 1-27.
XSEDE. (2016). XSEDE PY5 Annual Report: July 1, 2011 through June 30, 2016. Washington D.C.: XSEDE Office.
XSEDE-Extreme Science and Engineering Discovery Environment. (2020). XSEDE: Allocations Awards for the NSF Cyberinfrastructure Portfolio, 2004-2019 [Dataset]. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-0757799_V1
Yim, A., Chung, C., & Yu, A. (2018). *Matplotlib for Python Developers: Effective techniques for data visualization with Python*. Packt Publishing Ltd.