Technische Universität Berlin

Quality & Usability Lab, Faculty IV

Development of a frontend for the presentation of heterogeneous research data

Bachelor Thesis

Name	Matr.-Nr.	E-Mail	Studiengang
Bela Katschinski	413287	b.katschinski@campus.tu-berlin.de	Medientechnik B.Sc.

Supervisor: Prof. Dr.-Ing. Sebastian Möller

Submission Date: December 4, 2023

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 04.12.2023 

Ort, Datum Unterschrift

Abstract EN

As part of the project *A Digital Research Space for the BUA* which aims to create a joint research data platform for the Berlin University Alliance a frontend for the platform has been developed and instantiated. To answer the research question *What is a suitable way of presenting heterogeneous research data on a digital platform in a uniform and clear manner?*, the Design Science Research methodology is adopted to derive design requirements, design principles and design features. The instantiated frontend is built on the derived design features. The artifact supports visualization and analyzation features for tabular data. Furthermore other heterogeneous data types can be inspected on-site. Also a new taxonomy for categorizing publications on the new platform is created.

Abstract DE

Im Rahmen des Projekts *A Digital Research Space for the BUA*, die Entwicklungs einer gemeinsamen Forschungsdatenplattform für die Berlin University Alliance zum Ziel hat, wurde ein Frontend für die Plattform entwickelt und instanziiert. Zur Beantwortung der Forschungsfrage *Wie können heterogene Forschungsdaten auf einer digitalen Plattform einheitlich und übersichtlich präsentiert werden?* wird die Methodik des Design Science Research angewendet, um Design Anforderungen, Design Prinzipien und Design Features abzuleiten. Das instanziierte Frontend basiert auf den abgeleiteten Design Features. Das Artefakt unterstützt Visualisierungs- und Analysefunktionalitäten für tabellarische Daten. Darüber hinaus können andere heterogene Datentypen auf der Platform geöffnet werden. Auch eine neue Taxonomie zur Kategorisierung von Publikationen auf der neuen Plattform wird erstellt.

Contents

List of Figures

List of Tables

Acronym List

BUA Berlin University Alliance

BOP Berlin Open Science Platform

UI User Interface

UX User Experience

DSR Design Science Research

DSRP Design Science Research Process

DR Design Requirements

DP Design Principles

DF Design Features

HU Humboldt Universität

TU Technische Universität

FU Freie Universität

DOI Digital Object Identifier

DOM Document Object Model

NPM Node Package Manager

URL Uniform Resource Locator

1 Introduction

"Openness and sharing of information are fundamental to the progress of the effective functioning of the research enterprise [37]." There is an ongoing trend of making data publicly available. More and more governments provide their citizens with open data portals as the topic of open government data is getting ever more discussed [33]. At the same time the phenomenon of open data is disrupting the scientific community [10]. With modern rapid technological advances the global research community and its stakeholders are moving quickly to an open data ecosystem. Nevertheless, current research data repositories of the Berlin University Alliance (BUA) universities confront researchers with difficulties when it comes to discoverability and on-site analysis of research data. Thus, the need for a state-of-the-art joint research data platform for the BUA has been identified by the BUA.

The following thesis is part of the ongoing project *A Digital Research Space for the BUA* [3]. A project aimed at building the curation platform Berlin Open Science Platform (BOP) for research data of the BUA universities [1]. The digital research data platform should provide researchers with a platform housing all electronic research data from participating BUA partners [5]. The platform should enable users with tools to discover and analyze the data. In the course of this thesis, a frontend for this platform is developed. The frontend is populated with demo data to showcase the developed functionalities.

The BUA is a partnership of the four Berlin-based universities Freie Universität (FU) Berlin, Humboldt Universität (HU) zu Berlin, Technische Universität (TU) Berlin, and Charité – Universitätsmedizin Berlin. The alliance aims to turn Berlin into an integrated research environment and a leading science hub [2]. It is building an ecosystem of scientific institutions and aims to strengthen the cooperation of Berlin-based researchers. The Alliance has identified 5 main objectives to achieve its goals [2]. Focusing on Grand Challenges, Fostering Knowledge Exchange, Advancing Research Quality and Value, Promoting Talent and Sharing Resources. The BOP is part of Objective 5, Sharing Resources, as it tries to collect all scientific output of the BUA Universities in one digital portal [3]. Responsible for the development of the BOP is a group of research teams of FU Berlin and TU Berlin. The Human-Centered Computing (HCC) research group at FU Berlin defines requirements of the platform in participatory design workshops together with direct and indirect stakeholders namely researchers or librarians [3]. The identified requirements are given to the TU Berlin for development [4]. The BOP aims to improve transparency, openness and cooperation of Berlin's Universities [47].

1.1 Problem Identification

The BUA boasts a significant output of scientific publications. However, these publications are scattered on the respective data repositories of their host university. The TU Berlin hosts publications on their repository DepositOnce, the FU Berlin and Charité host their publications on the repository Refubium and the HU Berlin provides publications on their edoc server. This fragmentation of data makes it hard for researchers of the BUA to find data located at the individual institutions. This fragmentation is enhanced by the fact that all repositories have been developed individually and offer a completely different User Experience (UX). TU Berlin re-

searchers accustomed with the TU Berlin's repository DepositOnce have no intuitive understanding of the User Interface (UI) when using FU Berlin's repository Refubium or HU Berlin's edoc server. The repositories have no similarities except their functionality of hosting scientific publications. They use different UIs and employ different data taxonomies to categorize publications. The process of finding a publication differs from repository to repository. Whereas DepositOnce relies mainly on a full text search for finding items, a Refubium user must select a community first before having access to a full text search. The repository is based on the categorization of publications in communities with sub-communities and collections as part of overlying communities. To untangle this fragmentation of publications and UXs and enhance the discoverability of digital research data of the BUA the BOP aims to provide a unified environment to explore research data. As the main advantage of a platform is the combination of various data sources to have them available in an aggregated form [12] it makes sense to bundle all BUA publications on a platform to increase its value provided to users.

Additionally, non of the existing data repositories of the BUA offer the functionality of data visualization or data analysis. Making research data publicly available and easily discoverable brings various advantages to the scientific community and the publishing researchers themselves [13]. Considering the UX, it is important to provide a visualization functionality as data visualization potentially eases interpretation of abstract data [53]. As visualization takes fuller advantage of the capabilities of the human brain it can shift the balance between perception and cognition as patterns can be detected easier [20]. Complex data can be better understood by users when visualized [49]

A challenge that comes with the interactive visualization of data is ensuring a stable performance [54]. Earlier studies have indicated that heightened latency worsens the UX when exploring datasets [45]. Executing real-time large-scale data visualization can impose challenges regarding the performance due to communication delays and processing overhead [30]. Ensuring a stable performance of visualization tools is particularly important as the value of visualization becomes more evident when dealing with large datasets [21].

Furthermore, the current BUA research data repositories are strongly limited in their file inspection functionality. There is now way for users to inspect the content of hosted files on-site. Some PDFs can be opened with the browser native PDF viewer when hosted directly on-site. Files of any other type like tabular data or media have to be downloaded and opened locally on the user's machine. This further limits the usability of the current BUA research data repositories.

1.2 Objectives

The objective of this thesis is the development of a frontend for the research data platform of the BUA as an IT artefact and its evaluation. The central research question of this thesis is: *What is a suitable way of presenting heterogeneous research data on a digital platform in a uniform and clear manner?*

The developed frontend should accommodate scientific publications from BUA universities with heterogeneous data types attached. Various data types should be able to be inspected on-site. Data types which can be previewed should include tabular data, images, text, audio, video and PDF documents. Tabular data should

be able to be visualized and provide rudimentary data analysis on-site. The UI of the platform should be inspired by state-of-the-art data platforms. The platform should provide a full text search as well as filter possibilities for Provider, Generic Data Type, Department, Specific File Type and Tags. As part of the categorization of publications from the different BUA repositories a new data taxonomy should be developed deriving from the currently existing data taxonomies of the BUA repositories Refubium, DepositOnce and edoc server.

The development of the platform should follow the guidelines of the Design Science Research (DSR) methodology to ensure the scientific credibility of this thesis. The DSR methodology is commonly applied for artifact development. The developed frontend should be evaluated following the DSR methodology and its expanded evaluation methodology by Sonnenberg and Brocke [46].

The remainder of the thesis is based on the framework from Gregor & Hevner [23] for publication of DSR. This intends to ensure that both the scientific nature and the relevance for practice are sufficiently recognized and that the knowledge transfer of artifact creation is conveyed in a structured manner. Following the Introduction (1) a chapter on Related Works (2) presents the digital platforms in the context of open data and analyzes existing data platforms. Then, the chapter Methodology (3) introduces the chosen DSR methodology by Pfeffers et al. [41] as well as the chosen evaluation pattern by Sonnenberg and Brocke [46]. The following chapter Results (4) addresses the design and development of the artifact. Design Requirements (DR)s and Design Principles (DP)s are defined and the subsequent instantiation of Design Features (DF)s is described. The conducted evaluation following the evaluation principles by Sonnenberg and Brocke [46] is presented as well. In the following chapter Discussion (5) the results are critically interpreted and compared to the initial goals. Finally, the Conclusion (6) provides a summary of this work and an outlook for the future.

2 Related Works

In this chapter an introduction to the principles of digital platforms in the context of open data is given. This section also presents the currently employed research data repositories of BUA universities, the open data portal Data.gov by the US government and the data science competition platform Kaggle.

2.1 Digital Platforms in the Context of Open Data

Digital Platforms are a transformative concept changing business, the economy and society. Many major companies like AirBnb, Uber and Alibaba are built on the platform concept connecting users with resources [39]. The concept of the digital platform plays a key role in 4 of the worlds largest firms with Microsoft, Apple, Alphabet and Amazon [8]. These digital platforms open the way for radical changes in how we work, socialize and create value in the economy [29]. A digital platforms main defining point is that a digital platform connects individuals for a common purpose or to share a common resource. A common platform bringing together individuals can lead to an immense increase in utility and value [14]. Wikipedia would be an example of a platform that provides a large knowledge base created by the exchange and collaboration of users. In the case of the BOP the platform should bring researchers together enabling them to share and access research data in order to exchange information and enlarge the knowledge base of the platform. Another technique to take advantage of the knowledge base of a digital data platform are cross references to similar datasets using content based recommendation algorithms [40]. In a complex information space, recommendations can guide the user to the desired dataset [38]

A common form of a data platform are governmental open data portals [32]. The Open Data Portal is a widespread concept for institutions sharing and making data publicly available [27]. The path for sharing open data was laid out in 2009 when the Open Government Working Group established eight principles to guide the utilization of public data. They proposed that open data needed to be *complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary and license-free* [16]. As the US government under the Obama administration, laid out the Open Government Initiative defining the three principles *transparency, participation and collaboration* for a more open government [48]. The at the time released open data portal data.gov aiming to provide citizens with public data of the US paved the way for other public bodies to adapt the open government model and develop their own open data portal. Nowadays, many public bodies have their respective open data portal [50]. When analyzing governmental open data portals it can be seen that visualization and analyzation tools are often unsophisticated or non existent [35]. In the reality of complex datasets this can lead to an increasing gap between data producers and users of a data platform. Visualization features could bridge this gap and improve engagement among various stakeholders [7] [6] [31].

2.2 Existing BUA Research Data Repositories

The Berlin based University TU Berlin provides research data on their own platform DepositOnce. DepositOnce is the TU Berlin's institutional repository. It is hosted

by the Service Center Campus Management and operated and continually developed by the University Library [17]. The platform provides research data, publications with academic content, publications that must be made public in connection with examination regulations (i.e. doctoral theses and post-doctoral theses) and theses of TU Berlin students [17]. The platform provides a full text search with filter possibilities on author, date, subject and type of publication. The search page offers information on the publications publication date, publication type, authors, title and abstract. The publication page offers some more detailed information regarding the publication. Digital Object Identifier (DOI) identifier and publication source are provided. Full Metadata information can also be expanded and is therefore available. Attached files are listed and can be downloaded. There is now way of previewing the data. Therefore each file has to be downloaded to be accessed.

The FU Berlin's institutional repository, Refubium provides FU members with a free-of-cost framework to electronically publish documents, doctoral and habilitation theses, as well as research data [42]. Theses by Charité Berlin members are also accepted. The FUs Refubium is also part of the Charité's Open Access Repository [43]. Therefore Refubium also houses research data of the Charité. Refubium's data taxonomy is built on multiple *Communities*. Communities include *Dissertation*, *Research Data*, *Theses*, *Digitised Files*. These Communities allow the user to specify the search. Each Community can have sub-communities and sub-collections to further specify the search. Refubium then offers a full text search. Search results can also be browsed by title, author, department and keyword. Similar to DepositOnce information on the search page includes an abstract, author, title, publication date and keyword. The item page provides full metadata information and lists all attached files. However, there is no preview functionality forcing the user to download files to view their contents. The structure of communities housing sub-communities and collections can feel counterintuitive to the user and might make it harder to find the desired type of publication. Furthermore the lack of an all-encompassing full text search requires the user to know what primary community the desired publication is part of.

The HU Berlin's open access publication server edoc is the HU Berlin's respective research repository. Similar to Refubium the edoc server offers various communities including *research data, historical holdings, digital research journals*. Furthermore the communities offer collections and sub-communities which allow users to specify the search. In contrast to Refubium, edoc provides a full text search including all communities. When browsing a community, all attached sub-communities are listed above the search results. This list can be quite extensive and take up much space, diminishing the UX. Also edoc does not provide any form of data preview feature. Therefore if a datatype can not be opened by the browser it has to be downloaded to access [18].

Concerning the main features of the BUA research data platform all of the universities repositories offer a full text search. The repositories offer additional filters which can include author, title, department or keywords. A missing functionality in all repositories is the data preview. A data preview would allow the user to inspect and view attached files without downloading them. The current repositories can only preview the abstract of a publication. A preview can be especially useful when it comes to large quantities of research data that can be inspected directly on-site.

2.3 Other Data Platforms

Many governments provide public open data portals that are built to share data with users [34]. In the context of providing heterogeneous types of data to users they they are built for use cases similar to the BOP [55]. Therefore a short analysis of the open data portal of the United States of America Data.gov will follow. The Government of the United States' open data portal Data.gov provides access to datasets published by agencies across the federal government. Data.gov is intended to provide access to government open data to the public [15]. The portal is built on CKAN an open source data management system. CKAN provides several features among others a visualization feature for tabular and image data [11]. Tabular data can be rendered in an interactive table. The table can be sorted by the respective columns. Columns can be filtered and search queries can be run on the data. Rich metadata of the datasets provides the user with additional information regarding the dataset. Each column has a short description explaining the content of the column. Data.gov also offers visualization functionality. Columns can be plotted in form of various charts. Namely, bar charts, pie charts, scatter chart and more. Geographical data can also be displayed on a map. Summorized, Data.gov offers rich visualization and analyzation features of tabular data.

Another very popular platform to find and explore datasets is the platform Kaggle [26]. Kaggle is primarily a data science competition platform hosting datasets. Besides hosting data science competitions, the platform also enables users to find and publish datasets and offers extensive data exploration functionalities. Kaggle hosts mainly tabular data but also provides visualization features for heterogeneous data types. Datasets can have several files consisting of various data types. All files of a dataset are listed in the data explorer. The files are selectable via the data explorer. A selected file is opened and can be inspected on-site. On the platform images, audio and video files, tabular data and more can be inspected. As the main focus of the platform lies on tabular data the visualization is especially rich. Tabular data is rendered in an interactive table. At the top of each column additional information on the columns content is given via a histogram or descriptive statistics. Furthermore, more descriptive statistics on column data can be retrieved. Other data types can be inspected on-site but offer no additional descriptive statistics or metadata of the file.

It can be seen that in contrast to the previously examined current BUA research data repositories other data platforms offer rich visualization features leading to an increased discoverability of the data on-site.

3 Methodology

As the main focus of this thesis is the creation of an IT artifact it is conceptualized with the DSR methodology. The DSR methodology is especially suitable for scientific publications developing an IT artifact in regards to an existing problem. This thesis mainly follows the methodology by Pfeffers et al. [41] and evaluation principles outlined by Sonnenberg and Brocke [46].

3.1 Design Science Research

The DSR progress as outlined by Pfeffers et al. [41] consists of 6 main phases. In the first phase *problem identification and motivation* the underlying research problem should be defined and the consequent solution justified. The problem can be split up to further illustrate the solution's complexity. To complete this phase the state of the problem and the importance of the solution have to be understood. This phase has been completed by the BOP project led by the BUA. Co-creation workshops are conducted with stakeholders from research, library and administration. In these workshops problems are identified and passed on to the responsible group at TU Berlin [4]. The second phase *objectives of a solution* derives the solutions objectives from the problem definition. It defines how the artifact can address the research problem. This stage also requires knowledge on already existing solutions and their ramifications. To complete this phase the author has defined requirements to solve the identified problems. This has been done in accordance to Meth et al. [36] who proposes to initially define DR. From the defined DRs DPs are derived that will be instantiated by DFs. The DRs derive from the problems identified by the BUA project, a literature analysis and competing platforms 2. The third phase *design and development* is about creating the artifacts architecture and features and its actual creation. In this phase the author has derived DPs from the previously defined DRs. The derivation of DPs is based on the approach of Gregor & Hevner [22]. After instantiating DPs resulting DFs have been identified by the author. Based on the DFs a working frontend artefact has been developed. The artefact has been populated with exemplary demo data to showcase the developed features. As part of a DF current data taxonomies of BUA repositories have been used to derive a new data taxonomy for the developed artefact. The fourth and fifth phase *demonstration* and *evaluation* have been combined to a single step. The fourth phase includes a demonstration of the capabilities of the artifact. It should be used in a real world scenario to solve a problem. During the fifth phase *evaluation* the artifact's performance during the demonstration is observed and compared to the original objectives of the solution. As the DSR has an iterative character researchers can decide to iterate back to i.e. phase 3 in case the evaluation leaves the researchers wanting to further improve the artifact. Following the Sonnenberg and Brocke [46] evaluation pattern the author has conducted ex-ante and ex-post evaluation. The first evaluation phase has already been conducted through as part of the BUA project [4]. This thesis includes the design phase, evaluation phase 2, the construction phase and evaluation phase 3. The usage phase and evaluation phase 4 will be part of the BUA project in near future. The ex-ante evaluation has been conducted in form of expert interviews with the head of the team at TU Berlin responsible for the development of the BOP [3]. The ex-post evaluation has also been conducted with the same expert. Over

the course of several bi-weekly interviews the developed artifact was evaluated. In the final phase *communication* the problem, the artifact and its effectiveness are communicated to other relevant audiences. The communication will be conducted with this thesis. Further communication will be done by the BUA project. These phases can be followed in sequential order but the DSR process does not have to start with the problem identification and motivation phase. Entry points to the DSR process can be in any of the first 4 phases as described by Pfeffers et al. [41].

The following figure depicts the DSR process by Pfeffers et al. [41] for the artifact developed in the course of this thesis.

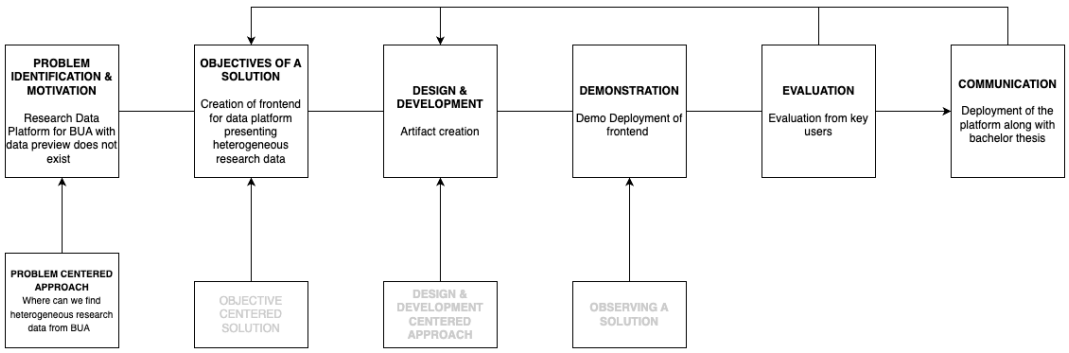

Figure 1: DSRP model according to Pfeffers et al. [41]

3.2 Evaluation Pattern

The evaluation of the developed artifact is based on Sonnenberg and Brocke [46]. They have developed a four step evaluation for DSR processes. They argue that the evaluation and articulation of truth statements about an artifact can not only be done ex-post but also ex-ante. They argue that the evaluation should not be limited to its application to a real world problem but also include its expected impact on the problem it desires to solve. Evaluation addressing the artifacts design differs from evaluation addressing the artifacts use by end users as evaluation addressing the design has to occur prior the its construction thus ex-ante. Thus, also ex-ante evaluation can provide prescriptive knowledge that can be added to the knowledge base. This approach differs from the traditional perspective on the build evaluate pattern proposed by Pfeffers et al. [41] in which the DSR process is split in two phases. The build phase which includes problem identification, definition of a solution and design and development. And the evaluation phase which includes demonstration, evaluation and communication. The Build-Evaluate pattern deduces that only ex-post evaluation is a truth producing evaluation as prescriptive knowledge as the result of the build phase has no truth in itself. Sonnenberg and Brocke [46] present 4 evaluation phases divided into ex-ante and ex-post evaluation. Following the evaluation pattern (Figure 3) by Sonnenberg et al [46], the problem identification and the subsequent evaluation phase 1 has been conducted by the BUA project. This thesis deals with the design phase in which a graphical prototype has been created, the ex-

ante evaluation phase 2, the construction of the artefact in form of the instantiation of the frontend and the ex-post evaluation phase 3.

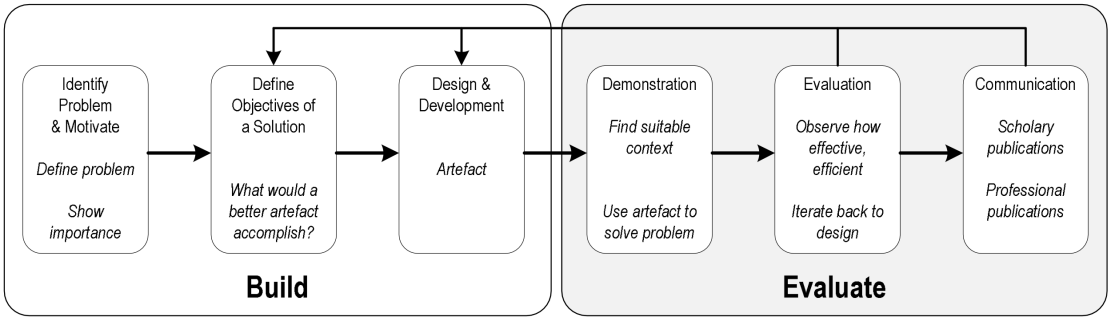

Figure 2: Build-Evaluate pattern proposed by Pfeffers et al [41]

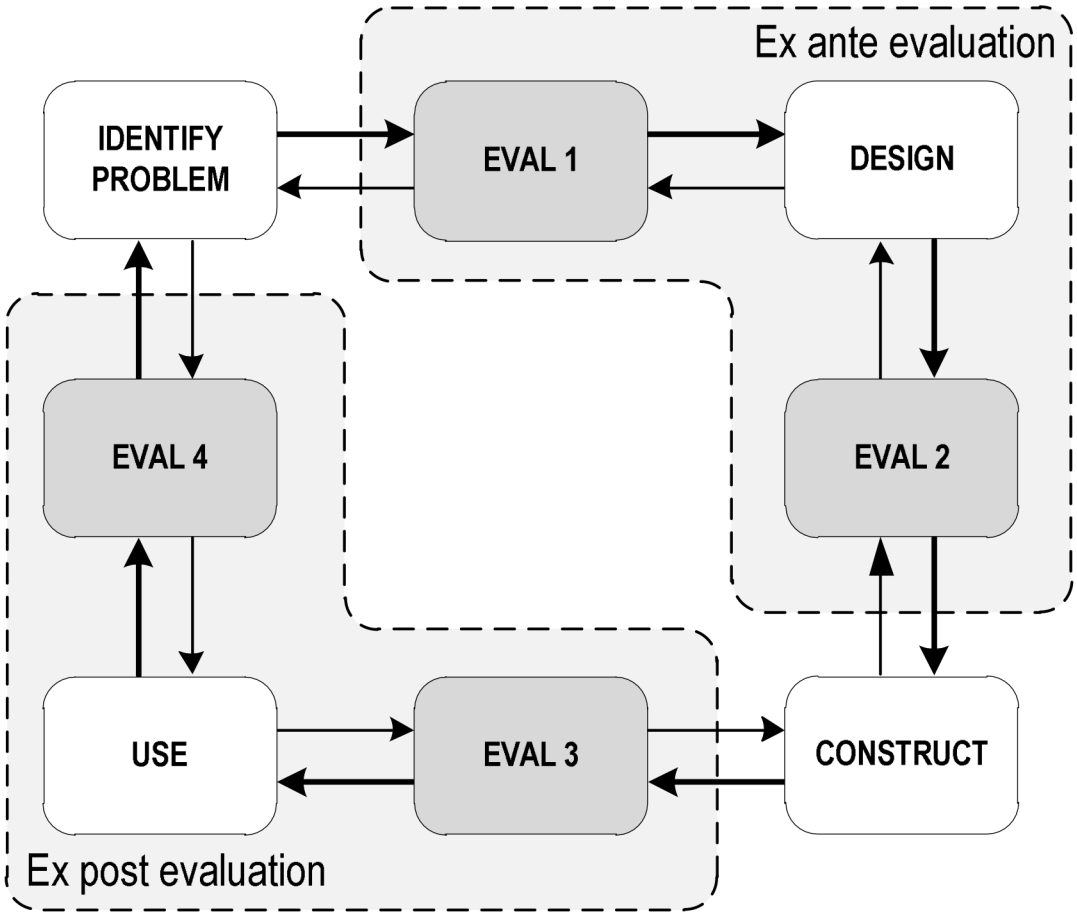

Figure 3: Evaluation activities within a DSR process according to [46]

3.2.1 Ex-Ante Evaluation

The first evaluation phase focuses on the problem identification. It evaluates the problem identification phase to ensure that the DSR problem is correctly identified.

Depending on what started the DSR process (i.e. an observed practical problem, missing research or an existing artifact) this evaluation can have various inputs. The expected output of this stage is a justification of the DSR process. How is the problem statement justified. What is the justification for the design objectives. Also the appropriateness of the design can be justified with already existing solutions [46]. In this thesis this refers to already existing data platforms. This evaluation phase has already been conducted as part of the BUA project *A Digital Research Space for the BUA*. The project conducted meetings with direct and indirect stakeholders to discuss design requirements, examination of prototypes and potential platform designs. Reflecting specific use cases, recommendations have been derived [3].

The second evaluation phase puts its focus on the progress to a solution for the problem to be solved. The design specifications will be compared to the design objectives. It serves as a first demonstration of the artifact to ensure necessary features and functionalities are included. As the artifact is not yet fully finished this is an artificial evaluation. Design specifications should provide value to stakeholders and users. In the design phase a high-fidelity prototype has been created. This phase builds on the created prototype. The design specifications of the prototype have been compared to the design objectives. This evaluation took place in an iterative process with the design phase. In biweekly meetings the author has conducted evaluation sessions with an expert to compare DFs to DRs. After the evaluation sessions the platforms design was iteratively improved.

3.2.2 Ex-Post Evaluation

Evaluation phase 3 is the first ex-post evaluation phase which takes advantage of the finished artifact. It is crucial as it reflects on the artifacts design and therefore paves the way for future design iterations. It evaluates the performance of the artifact within a real environment for the first time. This could be a real user evaluating the user experience while solving real tasks. The evaluation can take place in an artificially facilitated user environment with i.e. a guided user experience. The key to this evaluation phase is to gather evidence of the artifacts behaviour in real world scenarios. This phase has been conducted based on the developed artifact in the construction phase. The evaluation took place with an expert. The researcher used the artifact to solve real world problems. The expert had to evaluate the user experience and communicate it to the author. In form of an iterative process of repeating construction and evaluation phases, multiple evaluations have been conducted.

The fourth evaluation phase is the final evaluation putting all playing pieces together. The artifact is evaluated with real tasks by real users on a real system. The artifact is fully functional at this stage. This evaluation phase could be embedded in a case study, survey or field experiment. This evaluation phase should validate the artifacts ability to solve the desired problem. This evaluation phase is not part of this thesis. This evaluation phase will be part of further activities of the BUA project for a digital research space for the BUA.

4 Results

After identifying the problems and defining objectives, this section deals with the design and instantiation of the frontend of the research data platform in form of an it artifact. The methodology used for the instantiation and design of the artifact is the DSR methodology as developed by Pfeffers et al. [41] further developed by Sonnenberg and Brocke [46] as described in the Methodology chapter (3). The author defines DRs deriving from the identified problems. From these DRs DPs are further defined which can be used to specify concrete DFs to be instantiated in the it artifact. The definition of DRs, DPs and DFs has been developed by Meth et al. [36]. According to Rhyn et al . [44] DR can be understood as a meta requirement defining requirements that should be fulfilled by an instantiated artifact. DPs prescribe how the artifact should be built to meet the previously defined DRs. The DFs define a specific way of implementation of DPs in the instantiated artifact. This chapter will describe the process of derivation of DRs from identified problems and the resulting derivation of DPs. It also describes the instantiation of the it artifact on the basis of DFs.

4.1 Defining Design Requirements and Design Principles

As the evaluation phase 1 has already been conducted by the BUA project, the author could derive DRs from the identified problems. The derived DRs have a huge practical relevance as they build on the problems identified in meetings organized by the BUA project that have different stakeholders of the platform attending reflecting on requirements. The DRs also incorporate the knowledge base built by a literature review and an analysis of competing state-of-the-art data platforms and research data repositories. The DRs have been defined in the *Objectives of a solution* phase according to Meth et al. [36]. DR1 *Provide research data from BUA universities* derives from the problem of scattered research data of the BUA universities. This problem has been identified by the BUA project for building a Digital Research Space for the BUA universities. The platform should be a place to for knowledge exchange and research data and accommodate publications from all BUA participating universities. DR2 *Heterogeneous data types should be easily accessible* is a requirement deriving from the project committee. As research data published by BUA universities comes in all forms of data types, the platform should be able to accommodate various data types. The emphasis of this DR lies on the accessibility and discoverability of heterogeneous data types. On current research data repositories of BUA universities it is hard or impossible for the user to identify the content of published data files. The third DR3 *Enable users to conduct data analysis* has also been identified by the project committee. The immaturity of analysis features on competing data platforms has also been identified by Máchová et al. [35] as outlined in the Problem Identification Chapter (1).

DPs have been defined in the beginning of the design phase according to Meth et al. [36]. The DPs are derived based on the previously defined DRs as outlined by Gregor & Hevner [22] who claims DPs should incorporate the intended aim (A), the context (C) and the mechanism (M) to satisfy the DR. They also should be justified in form of a sound rationale (R). The actors involved in a DP are an implementer (I) who instantiates the artifact, a user (U) whose aims are to be achieved by the

artifact and an enactor (E) who executes the mechanism used to achieve the goal [22]. As the context for all DPs is the research data platform it will no longer be mentioned in the following. This also applies to the implementer as it is always the platforms designer and the rationale as it results from the reasoning behind the definition of DRs.

DP1 *Standardization of categories from various BUA repositories (A) to users (U) through the platform owner (E) who makes them available through the platforms data taxonomy (M).* This DP addresses the need for a new categorization system which accounts for the various data taxonomies employed by the BUA repositories. As the BOP should accommodate publications from all repositories and each repository uses a different categorization system it is necessary to design a new data taxonomy deriving from all existing data taxonomies employed currently on the various BUA repositories. Also deriving from DR1 DP2 *Aggregate publications and BUA research data (A) on one platform (M) by the platform owner (E) for users (U)* intends to address the scattering of publications and research data of BUA universities. DP3 *Provide intuitive access to publications and research data (A) to users (U) through the platform owner (E) who makes them available through the platform (M).* The implementation of DR1 intends to address the discoverability of publications from the BUA universities. Next, DP4 *Enable platform users (U) to inspect research data (A) with tools (E) available on site (M).* This DP is derived from DR2 and addresses the availability of heterogeneous data types and their user-friendly presentation. This aims to further improve the data discoverability and accessibility. Then, DP5 *Enable platform users (U) to visualize research data (A) with tools (E) available on site (M).* This DP is derived from DR2 and addresses the availability of heterogeneous data types and their user-friendly presentation. This DP's purpose is to highlight the visualization of tabular data. This aims to further aims to boost the visibility and discoverability of the data. Following, DP6 *Provide rudimentary data analysis (A) tools (E) on the platform (M) to users (U)* is derived from DR3 and aims to provide additional on site functionality to users trying to maximize the value provided to researchers by the platform.

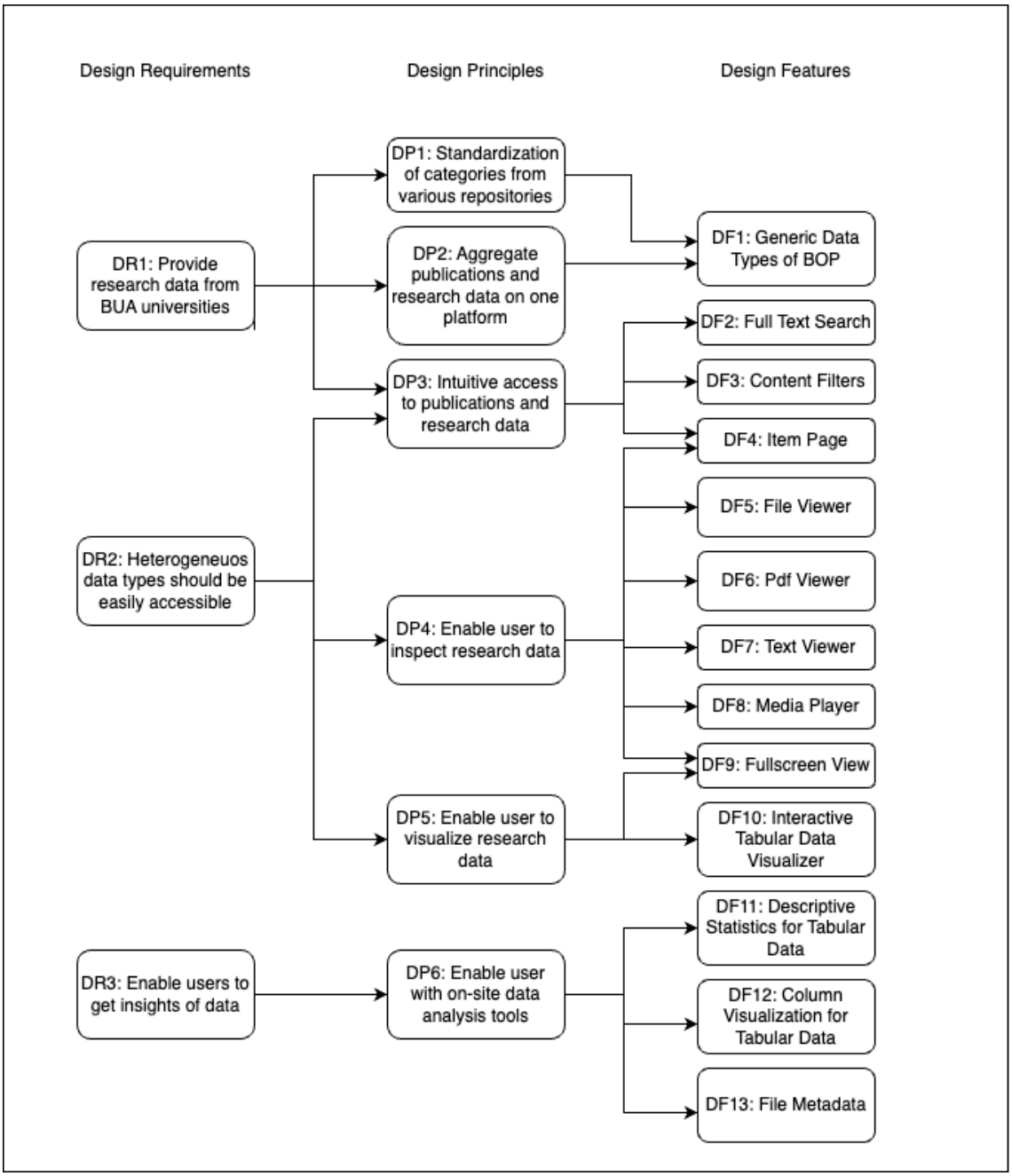

Figure 4: Mapping Design Principles to Design Requirements and Features according to Meth et al. [36]

4.2 Artifact Description through Design Features

Following the design, the subsequent step in the DSR process is the instantiation of the it artifact, the development of the frontend of the BOP. The artifact has been built with Vue a JavaScript framework for building user interface applications. Vue was created by Evan You a former Google employee. It builds on top of standard HTML, CSS, and JavaScript and provides a declarative and component-based programming model [52]. Vue is component based. That means HTML code can be put in single files and thus easily be reused. It also supports declarative rendering meaning that HTML components can be rendered or hidden based on JavaScript states. Another core feature of Vue is its reactivity. Vue tracks the state of JavaScript variables and updates the Document Object Model (DOM) when changes happen. Vue is one of the most popular frontend frameworks as it is a very versatile and light weight JavaScript framework [28]. The frontend is bundled in a Vite project. Vite is a build tool also developed by Evan You. Vite provides a improved developing experience as its development server comes with fast module replacement. This enables faster reloading of the local development server to compile local changes [51]. The frontend uses Bootstrap as its CSS framework. Bootstrap is extremely useful for frontend development as it provides many design CSS design templates [9]. Bootstrap Icons are utilized as the main icon package. The dependencies also include the Node Package Manager (NPM) package *vue-pdf-embed* an open source vue component for embedding pdf documents [19]. The frontend consists of three main pages. The Home Page, the Search Page and the Item Page. The following sections will describe the pages of the research data platform frontend in detail following the derived DFs. The DFs were derived from the previously defined DPs which are shown in Figure 4. These DFs are concrete features to be implemented in the artifact and are therefore suitable for developing the artifact [36].

The first **DF1** *Generic Data Types of BOP* is derived from DP1 and DP2 the standardization of categories from current BUA repositories and the aggregation of publications and research data of BUA universities. The BOP should accommodate publications from all current BUA research data repositories. Since each repository has a different categorization system it is necessary to aggregate all types of categories and define a new categorization system for the BOP deriving from the already existing categories of the respecting repositories. To satisfy this DF the author has conducted an analysis of the categorization methods of the current BUA repositories. The FU Berlin's Refubium structures its categories grouped in 6 communities: *Digital Copies, Dissertations, Research Data, Series and Multi-Volume Works, Student Theses and Other Scientific Publications.* Each of the communities provides the option to filter by type of publication. The community *Research Data* i.e. houses the publication types *Audiovisual, Collection, Dataset, Model, Poster File, Pre-Print PDF, Presentation, Software and Text.* HU Berlin's edoc server organizes categories similarly with collections being grouped in various communities. The community *Research Data* i.e. covers the collections *Other Research Data, Audio, Images, Datasets, Research Data Collections, Models, Software and Video*. TU Berlin's repository DepositOnce on the other hand does not group categories in any way. Thus, the filter bar on the left-hand side of the search results can feel crowded and not well organized. Equivalently to the others repositories research data types DepositOnce provides *Generic Research Data, Audio, Model, Image, Textual Data, Software, Video and Tabular Data.* It can be seen that some categories are equivalent

and will be merged to one category for the BOP. Namely, *Collection* (Refubium), *Generic Research Data* (DepositOnce) and *Other* (Edoc) describe the same type of items and are merged to *Generic Research Data* on the BOP. Similarly categories that house edge cases are merged. Namely, *Poster File, Pre-Print PDF* and *Presentation* from the community *Research Data* of Refubium are merged to the Category *Generic Research Data* on the BOP. Some categories on FU Berlin's Refubium are redundant and can be found in various communities. Most of the categories in the community *Scientific Publications* can be found similarly in the community *Series and Multi-Volume Works*. Categories like *Journal, Book, Dissertation, Habilitation* appear in both communities. Furthermore, the categories *Bachelor Thesis* and *Master Thesis* are part of the community *Student Theses* and the community *Scientific Publications*. Thus, the categories are merged to the three groups *Research Data, Scientific Publication* and *Thesis*. These general categories contain more specific categories similar to the data taxonomy employed by Refubium and edoc. The complete list of aggregated and condensed categories can be seen in Table 1. Tables containing the complete data taxonomies of Refubium, DepositOnce and edoc can be found in the Appendix in Table 2, Table 3, Table 4 and Table 5. The system of nested categories is chosen as the UI can be built more compact and thus provide a better UX since the number of categories to choose from is not so overwhelming to the user. Furthermore the DF also includes the instantiation of the above derived data taxonomy on the BOP. The Generic Data Type is displayed together with the search results above the title in a button like object. Items and Generic Data Type are in a N:N relation. An item can be part of more than one category, i.e. an item of type Research Data can be of type Audio and Image if it contains audio as well as video files. Moreover, an item could be part of more than one general category as it could be of type Scientific Publication and Research Data. Intuitively, a category can contain more than one item. The frontend also provides the option to filter for generic data types. In the filter box the three general Generic Data Types are displayed and can be selected with a click. When clicking on a general Generic Data Type the associated categories appear below and can be selected additionally.

Table 1: Data Taxonomy of BOP

Research Data	**Scientific Publication**	**Thesis**
Audio	Book	Bachelor Thesis
Image	Conference Proceedings	Master Thesis
Video	Book Section	Doctoral Thesis
Tabular	Preprint	Habilitation
Text	Article	Diploma Thesis
Software	Research Paper	
Model	Report	
Generic Research Data	Abstract	
	Periodical Literature	
	Review	
	Teaching Material	
	Letter to the Editor	
	Working Paper	
	Other Publications	

DF2 *Full Text Search* is the second DF. This feature stems from DP3 addressing the need to easily find publications from all BUA universities. The full text search browses all publications of all universities. The full text search is the key feature used for finding publications and research data on the BOP. This is highlighted by the fact that the full text search bar is the main element on the home page centered in the middle being the first element noticed by the user. Contrary to the full text search on i.e. the FU Berlin's research data repository Refubium, the full text search is accessible directly on the home page and on the search page without the need to select a community first. The search can be executed with pressing the enter key when the search bar is focused. The search query browses the metadata of publications like authors, title, abstract, etc. The search is based on string comparison. On a more technical level, the JavaScript function *String.includes()* is used to determine if the search query string occurs in the searched publication. The search query is transmitted via the Uniform Resource Locator (URL) under the parameter query. Thus, the user could theoretically manipulate the search query in the URL. This also enables sharing a search with other users in the form of a link. Search results are displayed vertically. Each result includes the title, author, abstract and further information. The title serves as a link to the item page. Above the title, information regarding the provider, which is the university from which the publication originates, is displayed. Also the generic data Type is specified. The generic data Type is the above explained data taxonomy. Furthermore the department of the university of which the item originates from is displayed. The right side of the search result additional information is displayed concerning the specific file types of files attached to the item and tags of the respective item.

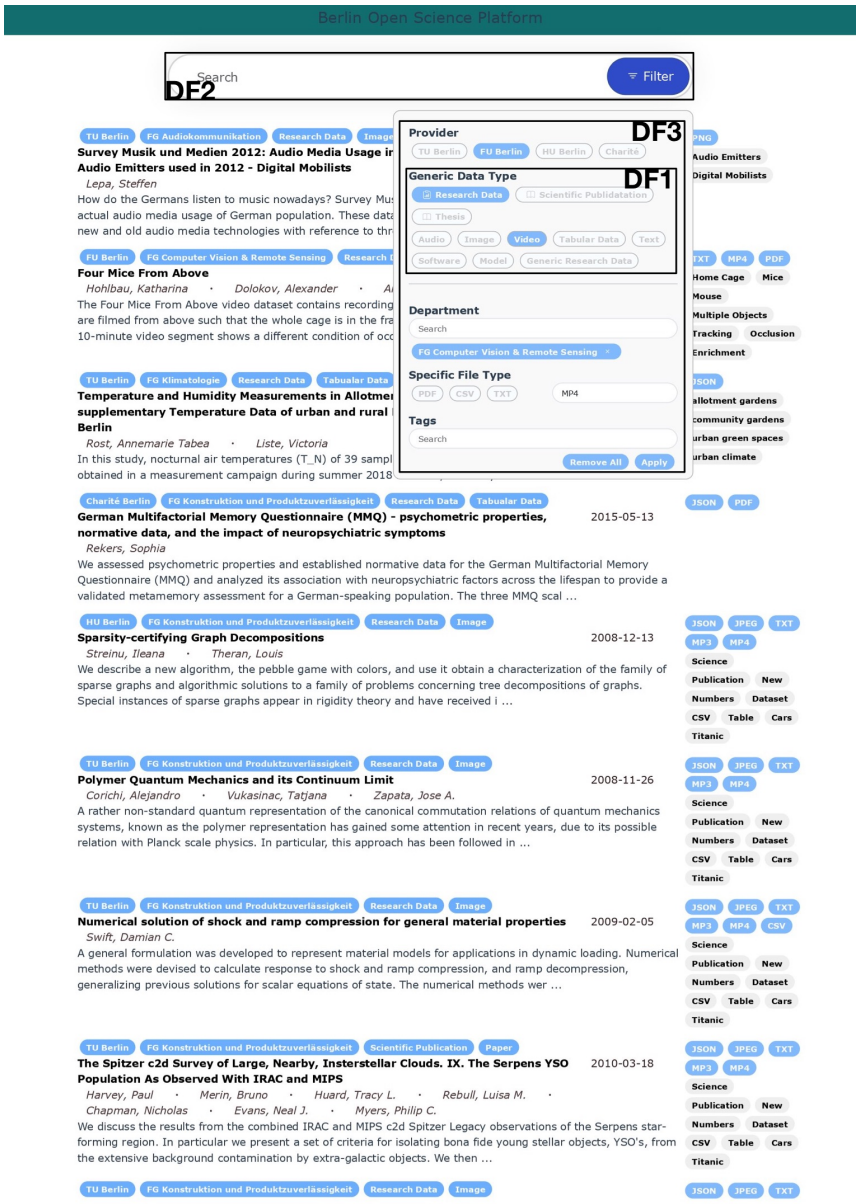

Figure 5: Search Page with annotated DFs

DF3 *Content Filters* is the third DF. Similar to the full text search, the filters are used to satisfy DP3 to provide intuitive access to BUA publications. The filters can be found in a box which can be expanded with a click on the filter button on the right-hand side inside the search bar. The filters provide additional options to specify the search. The filters can also be selected without a full text search query. The filters include the provider, the generic data type, the department, specific file types and tags. To comply with the visual identity of the platform the filters are designed in the same button like objects which can be found all over the platform. The button like objects come in a blue and a grey version. By default, a grey color indicates a not selected filter. The filters are additive. In each filter group multiple filters can be selected. A user could i.e. select TU Berlin and FU Berlin as providers. If a filter is selected it is highlighted as the color of the button like object switches to blue. If no providers are selected all available providers are displayed. The filters for provider and generic data type provide all selectable types directly as a button like object. The filter for department provides a search bar as there

exist hundreds of departments in the 4 BUA universities. A department can be selected with pressing the enter key when typing the department search bar. The selected department is displayed beneath the search bar as a button like object. The filter can be removed with clicking on the x of the button like object. Multiple departments can be selected. The filter for specific file type provides 3 preselected file types which are displayed as a button like object and can be selected with a simple click. Additionally for other file types there is a search bar with the same functionality as previously outlined for the department filters. The filter for tags works similar to the department filter. Selected filters can be applied with a click on the apply button in the bottom right corner of the filter box. Left to the apply button a remove all button enables the user to remove all selected filters with one click. All filters can be individually unselected by either clicking the blue button like object replacing it with its grey counterpart or clicking on the x.

DF4 *Item Page* is derived from DP3 and DP4 as the item page is the location for getting access to the publication and research data (DP3) as well as the location where functionality for visualization of research data can be found (DP4). The item page displays the full item information. The items title is displayed as the main heading. Below, the items authors names are displayed. Further below the page is vertically split with a bootstrap grid. The left side consists of the about and metadata section. This section can display two types of information. The about or abstract section and information on metadata. By default it displays the abstract of the paper or a short text with information on the item. With a click on the metadata navigation pill the metadata can be displayed. The about section will be hidden and replaced with the metadata. All metadata is provided except the abstract as it is already displayed in the about section. The navigation pills are identifiable to the user as such as the pill turns to a blue font color when hovered. Furthermore the currently selected navigation pill is underlined. On the right side information on the provider, tags and department is displayed in the typical button like objects.

The following **DF5** *File Viewer* is also instantiated on the item page and derives from DP4 to enable the user to visualize research data and publications. The file viewer can be found on the item page below the about and metadata section. It is rendered as a box containing all files attached to the respective item. All attached data can be downloaded as an archive file with a click on the download button displayed above the top right corner of the file box. The download button has a download icon to make it identifiable as such. The item box displays the files name in the top left corner. An icon also indicates the file type. The file box is split to the file previewer and the file explorer. The file explorer on the right-hand side lists all attached files in alphabetical order. The file explorer can be hidden with a click on the hide icon above the file explorer to maximize the size of the file previewer. If the file explorer has been hidden and can be revealed with a click on the reveal icon which appears in the top right corner when hiding the file explorer. The files can be selected with a click on the file name in the file explorer. When selected, the filename in the file explorer is underlined and highlighted with a grey background color to emphasize the selected file. The file previewer on the left-hand side opens the file and shows its content.

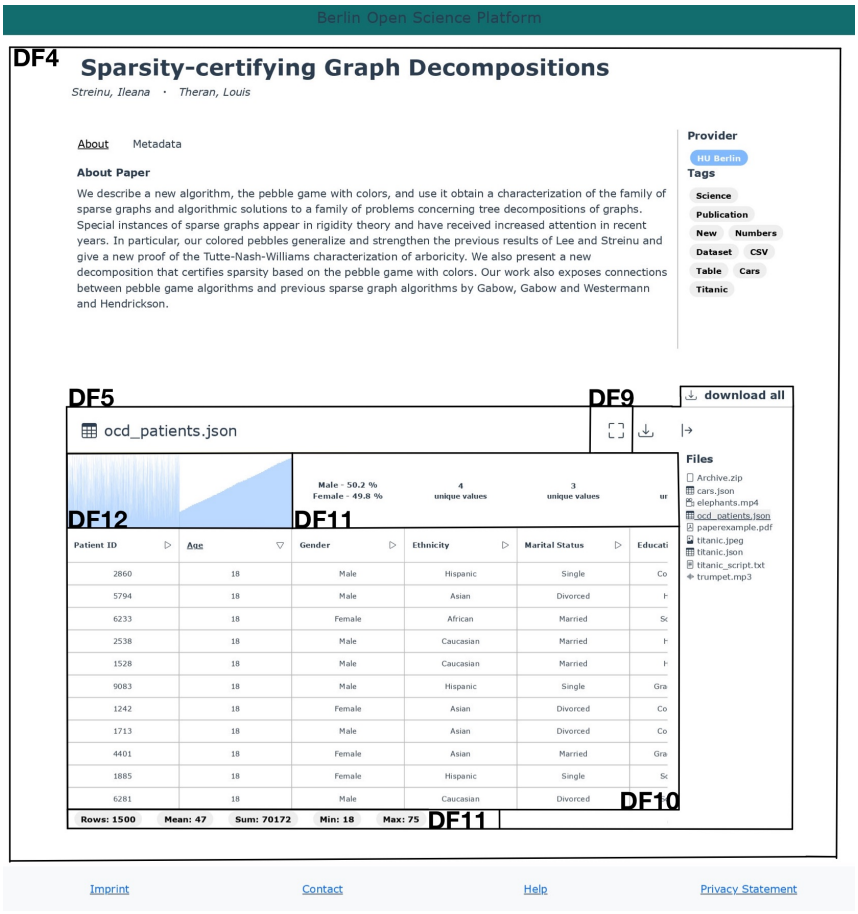

Figure 6: Item Page displaying Tabular Data with annotated DFs

DFs 6-8 refer to DP4 and aim to enable the user to inspect research data on-site. The file viewer is populated with the selected file and currently supports a PDF viewer powered by the JavaScript package vue-pdf-embed [19]. There is also a text viewer rendering text files. Additionally, a media player powered by the standard HTML audio player enables the playback of audio files. Furthermore, a video player enables the user to play video on-site.

DF9 *Fullscreen View* is derived from DP4 and DP5 and aims to provide a further improved UX. The above explained file viewer can be expanded into full screen mode utilizing the full browser window. This fullscreen view can be especially useful when working with the tabular data visualizer. In the standard view, larger tables require the user to scroll in order to see all columns and rows. The fullscreen view reduces this need of scrolling as more columns and rows are visible. The size of the view is only limited by the browser windows size.

DF10 *Interactive Tabular Data Visualizer* is derived from DP5 and enables the user to visualize research data. The file viewer provides additional visualization functionality for tabular data. The data is rendered as an interactive table. The table is rendered utilizing the standard HTML table combined with Vue's data binding capabilities. Additionally, the table is sortable by column values. When clicking on the triangle next to the column's head the table cycles between sorting the data ascending, descending and the default sorting.

Figure 7: Item Page displaying a Text in Fullscreen with annotated DFs

DF11 *Descriptive Statistics for Tabular Data* is a DF derived from DP6, the integration of rudimentary data analysis. This DF concerns itself with tabular data visualization. The above described rendered table offers additional analytical functionality. When clicking on a columns head the mean, min, max value and the sum of all values of the selected column is displayed. The selected column is highlighted as it's head is underlined. Additionally, the number of rows of the table is always displayed. All of this information can be found below the table in form of a sticky table footer. The selected column is highlighted as the columns head is underlined. This functionality is only given with columns whose content is of the JavaScript type number.

Another type of descriptive statistics is the value distribution in a row. The value distribution is displayed in the same row as the bar charts above each column. The distribution of values in the respective column is only displayed if the column contains binary data, thus only two different values. Column data can be in form of

JavaScript types String, Number or Boolean. The distribution is displayed in form of a string describing the distribution in percent (i.e. Value A - 20 % : Value B - 80 %). Additionally, the number of unique values of a column can be displayed. The number of unique values is additional analytical data provided to the user above the respective column of a table. This information is displayed when the column data is of JavaScript type String and contains more than two individual values. The number of individual values is displayed in form of a string. This DF concludes the DFs occupied with offering descriptive statistics as analyzation functionality to the user when displaying tabular data. Thus, each case of type of data in a column is covered. Each column provides additional descriptive statistics be it the distribution of values or the number of unique values.

Figure 8: Item Page displaying a Video with annotated DFs

DF12 *Column Visualization for Tabular Data* is another DF based on DP6 and aims to provide further tabular data analysis functionality to the user. When tabular data is displayed, above the table each column has additional analytical information displayed. The type of information depends on the respective column. If a column is of type number and does not contain binary data thus more than two different values, the data is plotted in form of a bar chart. The bar chart gives the user an overview on the content and the distribution of data in the respective column. The bar chart is created with the JavaScript package ChartJS [24] with the Vue wrapper vue-chartjs [25]. ChartJS is a lightweight charting JavaScript open source library.

The individual bars can be hovered to display the exact values. Since the space for displaying the chart is very limited the chart has been stripped of all non essential parts leaving only the bars. Information on x and y data can be retrieved with hovering over the bars. The charts rely on Vue's data binding and display data live. Thus, the chart is also sorted when sorting the respective column. Furthermore, all other charts will adjust it's order based on the newly sorted table.

DF13 *File Metadata* is the last DF derived from DP6 and represents the alternative to DF11 and DF12 for other data types than tabular data. This DF also provides the user with additional analytical data on the file currently selected. The file metadata is again displayed above the file contents corresponding to the existing visual identity of the frontend similar to DF11 and DF12. The metadata displayed depends on the selected file type. When selecting an image the top row displays information regarding the image's size in pixels, the image's size in kilobytes and the file type. If a text file is selected the number of words, the number of characters, the file size in kilobytes and the file type is displayed. When displaying an audio file the additional data provided consists of the file size in kilobytes, the audio's length in seconds and the audio file's bitrate. Previewing a video file the top row gives information on the videos resolution in pixels, the file size in megabytes and the video length in seconds. This feature does not support pdf files. This is done in order to maximize the space occupied by the pdf. Additionally the metadata of pdf files is usually of little value to users on the BOP.

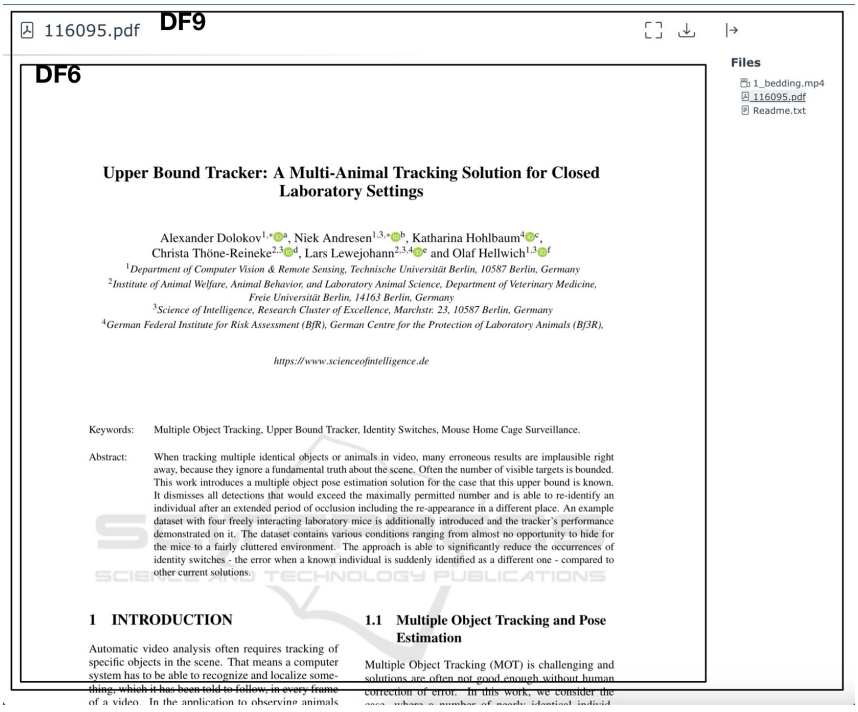

Figure 9: Item Page displaying a PDF in Fullscreen with annotated DFs

4.3 Ex-Ante Evaluation

Following the evaluation pattern proposed by Sonnenberg and Brocke [46] the acDSR process incorporates ex-ante evaluation phases. The first evaluation phase in which the problem the identified problem has to be justified and evaluated has already

been conducted as part of the BOP project in workshops discussing and evaluating requirements for the research data platform. Evaluation phase two, which succeeds the design phase in a DSR process after Sonnenberg and Brocke [46], has been conducted in form of an interview with an expert. In the design phase of the DSR process a graphical prototype has been created with the interface design tool Figma. The prototype has been created to showcase all features of the platform and evaluate them before developing the artifact. The evaluation of the prototype took place in form of biweekly meetings with the head of the team responsible for the development of the BOP [3]. As proposed by Sonnenberg and Brocke, the input for the ex-ante evaluation includes the Design Specifications, the Design Objectives and the Design Tool. These could be found as part of the graphical prototype created with Figma. The design was evaluated by various criteria. The criteria included feasibility of the design and its instantiation, accessibility and understandability to the user, the consistency of the visual identity of the platform and the UI, the operationability and completeness of the platforms ability to merge publications from all BUA participating universities. The process of evaluation as proposed by Sonnenberg and Brocke is of iterative character. Thus, the process of ex-ante evaluation and design phase took place over several weeks. Following feedback from the evaluation sessions, the prototypes design could iteratively be improved. Through the ex-ante evaluation phase various design decisions could be justified. Namely, the newly developed category system employed by the BOP to aggregate and merge all existing categories of BUA repositories. Furthermore, the design decision of the file previewer in combination with the file explorer used to equip users with the ability to preview data on-site, was positively evaluated. Other design decisions regarding data analysis functionality of DF11 and DF13 were improved iteratively over the course of multiple evaluation sessions.

4.4 Ex-Post Evaluation

After concluding the Design phase of the Design and Development phase of the DSR process, the next phase as described by Sonnenberg and Brocke is the construction phase succeeded by the ex-post evaluation phase 3. Evaluation phase 3 as described by Sonnenberg and Brocke [46] was conducted as part of this thesis. This ex-post evaluation is based on the developed artifact and aims to validate the artifacts instance in an artificial setting. The evaluation is closely intertwined with the construction phase. Similar to the ax-ante evaluation described above, the ex-post evaluation is of an iterative character and is conducted over the course of several weeks. The evaluation phase alternates with the construction phase allowing for iterative improvement of the artifact. The ex-post evaluation of the artifact took advantage of the developed artifact which was developed employing the JavaScript framework Vue. In the form of biweekly meetings, interviews with an expert were conducted to justify the design tools employed. Criteria by which the artifact was evaluated included the feasibility of features, their ease of use, the effectiveness and efficiency of features, the fidelity of the frontend with real world phenomena, the robustness of the platform as well as the suitability of features in light of respective use cases. The evaluation phase was used to investigate the technical feasibility of DF10, the plotting of data in form of multiple small plots above the respective columns. The DF had to be evaluated in light of performance issues of the frontend

leading to the testing and finally the usage of various JavaScript charting libraries. Since, multiple plots are instantiated and bound to live data from the displayed table the computing capacities needed for the plots quickly rose to considerable size. Furthermore, since VueJs supports conditional rendering the charts are newly rendered when displaying a new table. Thus, the performance of the employed plotting library was of great importance. Finally, ChartJS could be chosen as the most light weight and best performing library. As a result of this evaluation phase the instantiated artifact could be validated in an artificial setting.

Following the evaluation pattern described by Sonnenberg and Brocke the succeeding phases in a DSR process are the Use phase and the evaluation phase 4. This thesis is limited to evaluation phase 3 as the last evaluation phase 4 as described by Sonnenberg and Brocke [46] will be conducted later in the process of the BOP project. Evaluation phase 4 incorporates a fully finished artifact which is validated in a naturalistic setting. This evaluation phase is not feasible in the scope of this thesis as this thesis' aim is to develop a frontend filled with demo data to validate a proof-of-concept. Evaluation phase 4 as described by Sonnenberg and Brocke [46] would require a working prototype of the platform with a connected back end to conduct the evaluation with a fully working artifact.

5 Discussion

This thesis is part of the BUA project aiming to build the joint research data platform BOP for the BUA universities. BUA research data is currently scattered on the respective research data repositories of the universities. Furthermore, the repositories lack visualization and data analyzation functionalities. Thus, the need for a new joint research data platform has been identified by the BUA. The goal of this thesis is the design and instantiation of a frontend for the BOP focusing on the presentation of heterogeneous research data. The central contribution of this thesis is the definition of DRs from which DPs have been derived that have been used to further derive DFs for the modern research data repository of the BUA. The DPs can be used as a guideline and be applied in the development process of other research data repositories to outline the main requirements that have to be fulfilled by a modern research data repository. Thus, this thesis contributes to the problem of research data repositories and their challenge to uniformly present heterogeneous research data.

Another contribution of this thesis is the instantiation of the platform in form of an it artifact. A frontend has been developed and populated with demo data. With the instantiation of the frontend the raised research question *what is a suitable way of presenting heterogeneous research data* has been answered in a practical context. A file viewer has been built in which heterogeneous data types can be inspected and visualized. Additionally, the file viewer provides descriptive statistics, information on metadata and visualization features for tabular data. In contrast to other data platforms which struggle with important visualization or analyzation features [35] [53] [49] [20] this challenge has been addressed with the developed file viewer. The concept of the file viewer employing modular previewers for various data types has been identified as an appropriate way to present heterogeneous research data and improve discoverability and visibility of research data.

To satisfy the requirement of a joint research data platform for all BUA universities it was necessary to create a new data taxonomy. The taxonomy accounts for all currently employed data taxonomies. A full list of the currently employed data taxonomies and the newly created data taxonomy can be found in the appendix 6.

As the DFs have been evaluated ex-ante and ex-post after their instantiation they contribute to the existing body of knowledge of prescriptive and descriptive knowledge concerning data platforms. Research concerning open data platforms can build on the created and evaluated DPs and DFs to further investigate visualization and analyzation techniques in the context of the presentation of heterogeneous data on data platforms.

Limitations of this thesis include the comparatively small amount of demo data. Only a few data samples from the current research data repositories have been picked to populate the artifact. Real data is potentially formatted in other ways and includes more data types than what can be covered by the demo data. The full scope of variance in data could not be covered as part of this thesis due to time limitation. Also large files would need to be accounted for when looking at the performance of the artifact. Further development would need to be done to ensure a stable performance of the platform when visualizing large tabular data as increased latency negatively impacts the efficiency of data exploration sessions [45] [54]. In light of the fact that visualization of data is especially useful when dealing with large

datasets [21] a stable performance would increase the value of the visualization functionality provided on the research data platform. To take full advantage of the knowledge base of the research data platform a content based recommendation system would help users finding related publications in a complex information space [38]. A recommendation algorithm was not utilized as part of this thesis due to time limitation. Additionally, the evaluation could have been expanded as only one expert was consulted for the evaluation sessions over the course of several interviews. The utilized evaluation pattern recommends additional evaluation phases [46]. This evaluation could have been extended to a broader scope with i.e. multiple focus groups to include experts from various user groups as stakeholders of the platform.

6 Conclusion and Outlook

The project *A Digital Research Space for the BUA* was launched to build the joint research data platform BOP for the BUA. The need for a new research data platform was identified as BUA research data is currently scattered on the respective research data repositories of the universities. After analyzing the currently employed research data repositories of BUA universities it became clear that the repositories lack features concerning data discoverability and on-site data analyzation. In the course of this thesis a frontend for the BOP has been designed and instantiated. This thesis describes the DSR process after Meth et al. [36] which has been used to design and instantiate the frontend for the BOP. To answer the research question of how heterogeneous research data can be presented in a uniform and clear manner, DRs have been defined from which DPs could be derived. The DRs and DPs have been defined after an analyzation of other data platforms in regard to the discoverability of data and their visualization and analyzation features. The subsequently derived DFs have been used to design, instantiate and evaluate the frontend. The construction and evaluation pattern used were proposed by Sonnenberg and Brocke [46]. Furthermore a new data taxonomy which is derived from the current BUA research data repositories is presented.

Future research on the presentation of heterogeneous data on a digital platform can profit from this thesis as the DPs and DFs can be applied to other domains. This thesis also contributes to the existing body of knowledge as the DFs have been evaluated in form of expert interviews on the instantiated artifact.

Forthcoming research could investigate more sophisticated visualization tools. The literature research revealed that visualization of data increases the human ability to better understand data [53]. Thus, the impact of extended visualization and analyzation tools could be investigated. Possible would be i.e. a dedicated visualization dashboard with extended visualization and data analyzation tools. The analyzation tools could also be extended to more data types than tabular data to provide broader on-site analyzation features.

A recommendation functionality would also be conceivable. Recommending related publications would further utilize the knowledge base provided by the platform and increase the findability and exchange of research data on the platform.

To improve handling of larger files, future research could investigate other visualization tools with respect to their performance. Namely, visualization for large tabular data or video data can easily impact the platforms performance. Thus, the respective tools would need to be developed in a way that maintains fast previewability also for large data files to the users.

References

[1] Berlin University Alliance. *Berlin University Alliance*. 2023. URL: https://www.berlin-university-alliance.de/. (accessed 01.12.2023).

[2] Berlin University Alliance. *BUA Commitment*. 2023. URL: https://www.berlin-university-alliance.de/en/commitments/index.html. (accessed 01.12.2023).

[3] Berlin University Alliance. *BUA Digital Research Space*. 2023. URL: https://www.berlin-university-alliance.de/en/commitments/sharing-resources/digital-research-space/index.html. (accessed 01.12.2023).

[4] Berlin University Alliance. *Co Creation Workshops*. 2023. URL: https://www.berlin-university-alliance.de/commitments/research-quality/forschung/ongoing/co-creation/index.html. (accessed 01.12.2023).

[5] Berlin University Alliance. *Research Data Management and Open Access*. 2023. URL: https://www.berlin-university-alliance.de/en/commitments/sharing-resources/_inhaltselemente/overview-fdm-oa/index.html. (accessed 01.12.2023).

[6] Bahareh Ansari, Mehdi Barati, and Erika G Martin. "Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research". In: *Government Information Quarterly* 39.1 (2022), p. 101657.

[7] Raissa Barcellos et al. "Transparency in practice: using visualization to enhance the interpretability of open data". In: *Proceedings of the 18th Annual International Conference on Digital Government Research*. 2017, pp. 139–148.

[8] Carla Bonina et al. "Digital platforms for development: Foundations and research agenda". In: *Information Systems Journal* 31.6 (2021), pp. 869–902.

[9] Bootstrap. *Bootstrap*. 2023. URL: https://getbootstrap.com/. (accessed 01.12.2023).

[10] Katrin Braunschweig et al. "The state of open data". In: *Limits of current open data platforms* 1 (2012), pp. 72–72.

[11] CKAN. *CKAN Visualization*. 2023. URL: https://ckan.org/features/visualisation/. (accessed 01.12.2023).

[12] Felix Creutzig. "An integrated data platform to leverage the benefits of smart mobility". In: (2020).

[13] Theresa M Culley. *The frontier of data discoverability: Why we need to share our data 1*. 2017.

[14] Michael A Cusumano, Annabelle Gawer, and David B Yoffie. *The business of platforms: Strategy in the age of digital competition, innovation, and power*. Vol. 320. Harper Business New York, 2019.

[15] Data.gov. *US Open Data Portal*. 2023. URL: https://data.gov/. (accessed 01.12.2023).

[16] Sharon S Dawes. "Stewardship and usefulness: Policy principles for information-based transparency". In: *Government information quarterly* 27.4 (2010), pp. 377–383.

[17] DepositOnce. *DepositOnce Policy*. 2023. URL: `https://depositonce.tu-berlin.de/info/depositonce-policy`. (accessed 01.12.2023).

[18] Edoc. *Edoc Server HU Berlin*. 2023. URL: `https://edoc.hu-berlin.de/`. (accessed 01.12.2023).

[19] Vue Pdf Embed. *Vue Pdf Embed*. 2023. URL: `https://www.npmjs.com/package/vue-pdf-embed`. (accessed 01.12.2023).

[20] Stephen Few. *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc., 2006.

[21] Parke Godfrey, Jarek Gryz, and Piotr Lasek. "Interactive visualization of large data sets". In: *IEEE transactions on knowledge and data engineering* 28.8 (2016), pp. 2142–2157.

[22] Shirley Gregor, Leona Chandra Kruse, Stefan Seidel, et al. "Research perspectives: the anatomy of a design principle". In: Association for Information Systems. 2020.

[23] Shirley Gregor and Alan R Hevner. "Positioning and presenting design science research for maximum impact". In: *MIS quarterly* (2013), pp. 337–355.

[24] Chart JS. *Chart JS*. 2023. URL: `https://www.chartjs.org/`. (accessed 01.12.2023).

[25] Vue Chart JS. *Vue Chart JS*. 2023. URL: `https://github.com/aperture less/vue-chartjs`. (accessed 01.12.2023).

[26] Kaggle. *Kaggle Datasets*. 2023. URL: `https://www.kaggle.com/datasets`. (accessed 01.12.2023).

[27] Evangelos Kalampokis, Efthimios Tambouris, and Konstantinos Tarabanis. "Open government data: A stage model". In: *Electronic Government: 10th IFIP WG 8.5 International Conference, EGOV 2011, Delft, The Netherlands, August 28–September 2, 2011. Proceedings 10*. Springer. 2011, pp. 235–246.

[28] Marin Kaluža, Krešimir Troskot, and Bernard Vukelić. "Comparison of front-end frameworks for web applications development". In: *Zbornik Veleučilišta u Rijeci* 6.1 (2018), pp. 261–282.

[29] Martin Kenney and John Zysman. "The rise of the platform economy". In: *Issues in science and technology* 32.3 (2016), p. 61.

[30] Sungchul Lee, Ju-Yeon Jo, and Yoohwan Kim. "Performance testing of web-based data visualization". In: *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2014, pp. 1648–1653.

[31] Taejun David Lee, Seulki Lee-Geiller, and Byung-Kwan Lee. "Are pictures worth a thousand words? The effect of information presentation type on citizen perceptions of government websites". In: *Government Information Quarterly* 37.3 (2020), p. 101482.

[32] Martin Lnenicka. "An in-depth analysis of open data portals as an emerging public e-service". In: *International Journal of Humanities and Social Sciences* 9.2 (2015), pp. 589–599.

[33] Martin Lnenicka and Anastasija Nikiforova. "Transparency-by-design: What is the role of open data portals?" In: *Telematics and Informatics* 61 (2021), p. 101605.

[34] Renata Máchová and Martin Lněnička. "Evaluating the quality of open data portals on the national level". In: *Journal of theoretical and applied electronic commerce research* 12.1 (2017), pp. 21–41.

[35] Renáta Máchová, Miloslav Hub, and Martin Lnenicka. "Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective". In: *Aslib Journal of Information Management* 70.3 (2018), pp. 252–268.

[36] Hendrik Meth, Benjamin Mueller, and Alexander Maedche. "Designing a requirement mining system". In: *Journal of the Association for Information Systems* 16.9 (2015), p. 2.

[37] Engineering National Academies of Sciences, Medicine, et al. "Open science by design: Realizing a vision for 21st century research". In: (2018).

[38] Tommaso Di Noia et al. "Sprank: Semantic path-based ranking for top-n recommendations using linked open data". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.1 (2016), pp. 1–34.

[39] Geoffrey G Parker, Marshall W Van Alstyne, and Sangeet Paul Choudary. *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company, 2016.

[40] Michael J Pazzani and Daniel Billsus. "Content-based recommendation systems". In: *The adaptive web: methods and strategies of web personalization*. Springer, 2007, pp. 325–341.

[41] K Pfeffers et al. "The design science research process: A model for producing and presenting information systems research". In: *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006), Claremont, CA, USA*. 2006, pp. 83–106.

[42] Refubium. *Refubium*. 2023. URL: `https://refubium.fu-berlin.de/`. (accessed 01.12.2023).

[43] Charité Open Access Repository. *Charité Open Access Repository*. 2023. URL: `https://bibliothek.charite.de/en/publishing/how_to_publish_a_dissertation_or_habilitation_thesis/dissertations/electronic_publishing/`. (accessed 01.12.2023).

[44] Marcel Rhyn et al. "Opening the Black Box: How to Design Intelligent Decision Support Systems for Crowdsourcing." In: *Wirtschaftsinformatik (Zentrale Tracks)*. 2020, pp. 50–65.

[45] Ben Shneiderman. "Response time and display rate in human performance with computers". In: *ACM Computing Surveys (CSUR)* 16.3 (1984), pp. 265–285.

[46] Christian Sonnenberg and Jan Vom Brocke. "Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research". In: *Design Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings 7*. Springer. 2012, pp. 381–397.

[47] Open Distributed Systems. *Berlin Open Science Platform*. 2023. URL: `https://www.tu.berlin/ods/forschung/projekte`. (accessed 01.12.2023).

[48] Jeffrey Thorsby et al. "Understanding the content and features of open data portals in American cities". In: *Government information quarterly* 34.1 (2017), pp. 53–61.

[49] John W Tukey and Martin B Wilk. "Data analysis and statistics: an expository overview". In: *Proceedings of the November 7-10, 1966, fall joint computer conference*. 1966, pp. 695–709.

[50] Jürgen Umbrich, Sebastian Neumaier, and Axel Polleres. "Quality assessment and evolution of open data portals". In: *2015 3rd international conference on future internet of things and cloud*. IEEE. 2015, pp. 404–411.

[51] Vite. *Vite*. 2023. URL: `https://vitejs.dev/guide/`. (accessed 01.12.2023).

[52] Vue. *Vue*. 2023. URL: `https://vuejs.org/guide/introduction.html`. (accessed 01.12.2023).

[53] TM Willigen. "Measuring the user experience of data visualization". MA thesis. University of Twente, 2019.

[54] Eugene Wu et al. "Combining Design and Performance in a Data Visualization Management System." In.

[55] Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. "Innovation with open data: Essential elements of open data ecosystems". In: *Information polity* 19.1-2 (2014), pp. 17–33.

A Data Taxonomies

Category
Doctoral Thesis - C3
Article - B5
Research Paper - B6
Conference Object - B2
Book - B1
Book Part - B3
Report - B7
Periodical Part - B9
Tabular Data - A4
Generic Research Data - A8
Master Thesis - C2
Preprint - B4
Conference Proceedings - B2
Image - A2
Textual Data - A5
Software - A6
Video - A3
Habilitation - C4
Audio - A1
Model - A7

Table 2: Data Taxonomy of DepositOnce

Research Data	**Qualification Papers**
Collection - A8	Final Theses - C1 / C2
Dataset - A4	Dissertations - C3
Audio - A1	Habilitations - C4
Model - A7	
Image -A2	
Software - A6	
Video - A3	
Other - A8	

Table 3: Excerpt of Data Taxonomy of Edoc

Research Data	**Scientific Publication**	**Student Thesis**
Audiovisual - A1 / A2 / A3	Abstract - B8	Bachelor Thesis - C1
Collection - A8	Bachelor Thesis - C1	Master Thesis - C2
Dataset - A4	Book - B1	
Model -A7	Chapter of Book - B3	
Text - A5	Dissertation - C3	
Poster File - A8	Habilitation - C4	
Pre-Print PDF - A8	Conference Proceedings - B2	
Presentation - A8	Master Thesis - C2	
Software - A6	Review - B10	
	Movie - A3	
	Working Paper - B13	
	Journal - B9	
	Images - A2	
	Diploma Thesis - C5	
	Teaching Material - B11	
	Letter to the Editor - B12	
	Poster - A8	
	Preprint - B4	
	Report - B7	

Table 4: Data Taxonomy of Refubium Part 1

Digital Copies	**Dissertations**	**Series and Multi-Volume Works**
Book - B1	Dissertation - C3	Working Paper - B13
Part of a Periodical - B9	Habilitation - C4	Journal - B9
		Chapter of Book - B3
		Dissertation - C3
		Habilitation - C4
		Conference Proceedings - B2
		Poster - A8
		Preprint - B4
		Other Publications - B14
		Part of a Periodical - B9
		Article - B5
		Teaching Material - B11

Table 5: Data Taxonomy of Refubium Part 2

Table 6: Data Taxonomy of BOP

Research Data	**Scientific Publication**	**Thesis**
Audio - A1	Book - B1	Bachelor Thesis - C1
Image - A2	Conference Object - B2	Master Thesis - C2
Video - A3	Book Section - B3	Doctoral Thesis - C3
Tabular - A4	Preprint - B4	Habilitation - C4
Text - A5	Article - B5	Diploma Thesis - C5
Software - A6	Research Paper - B6	
Model - A7	Report - B7	
Generic Research Data - A8	Abstract - B8	
	Periodical Literature - B9	
	Review - B10	
	Teaching Material - B11	
	Letter to the Editor - B12	
	Working Paper - B13	
	Other Publications - B14	

B Figma Prototype

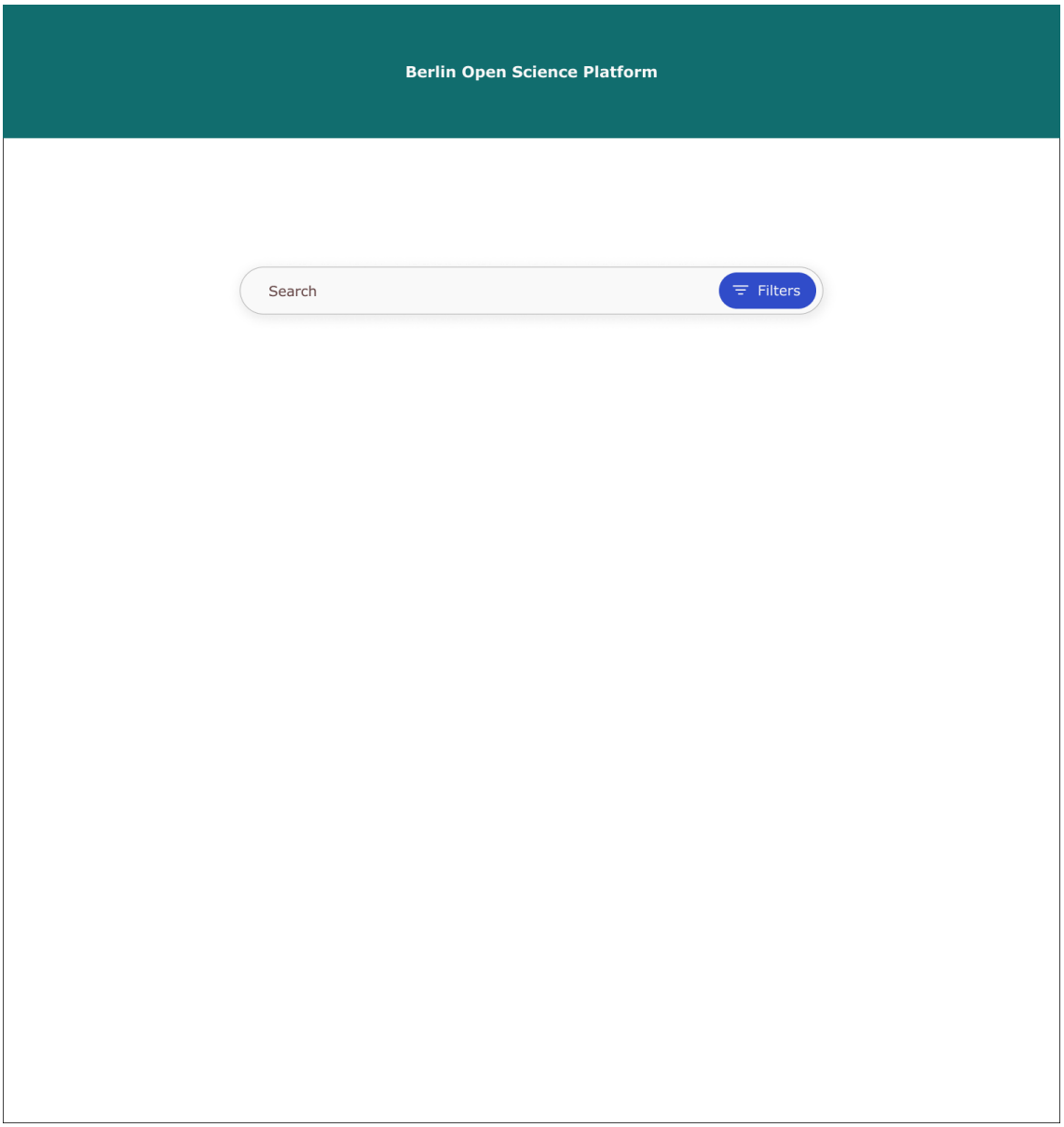

Figure 10: Home Page

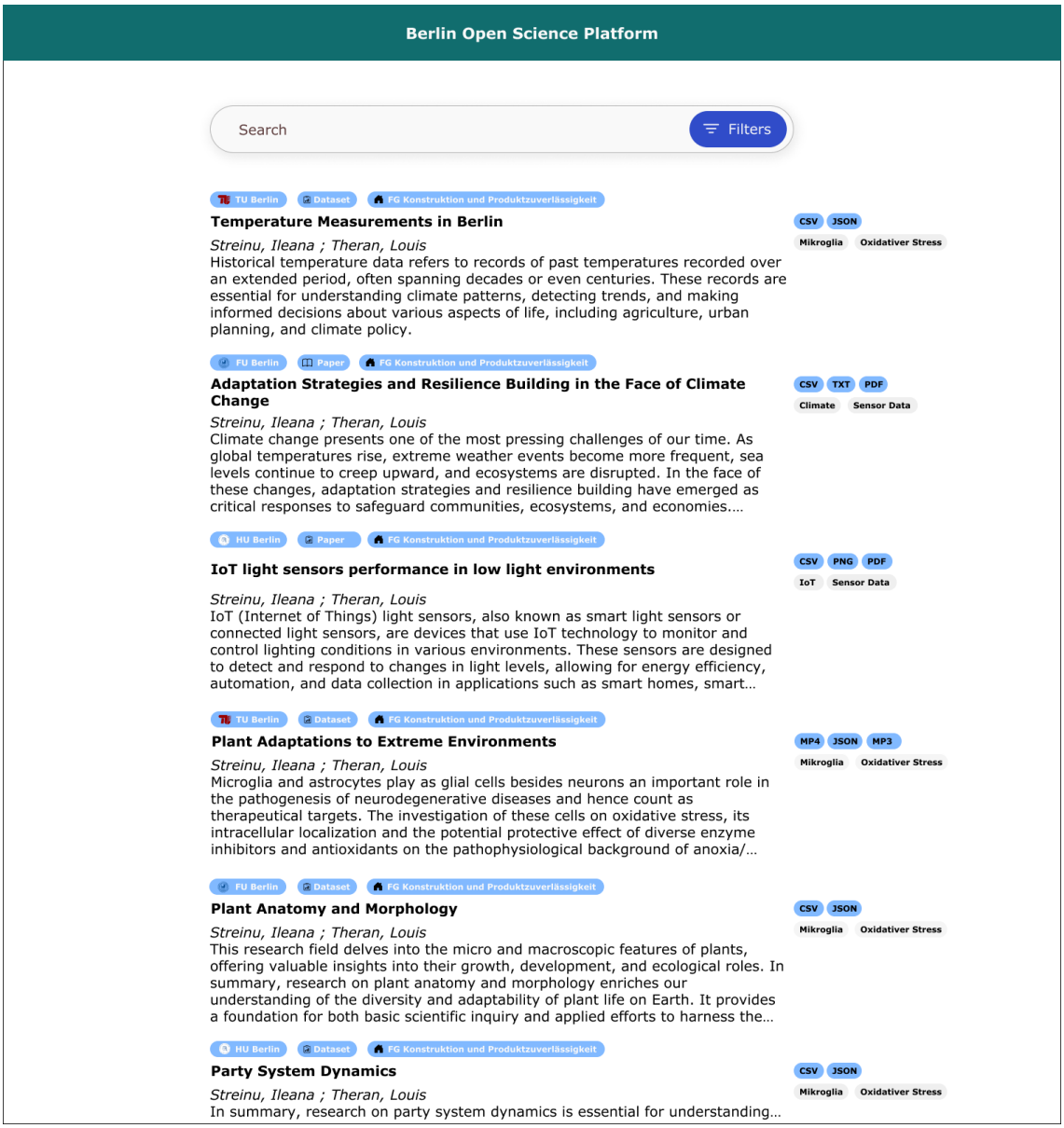

Figure 11: Search Results

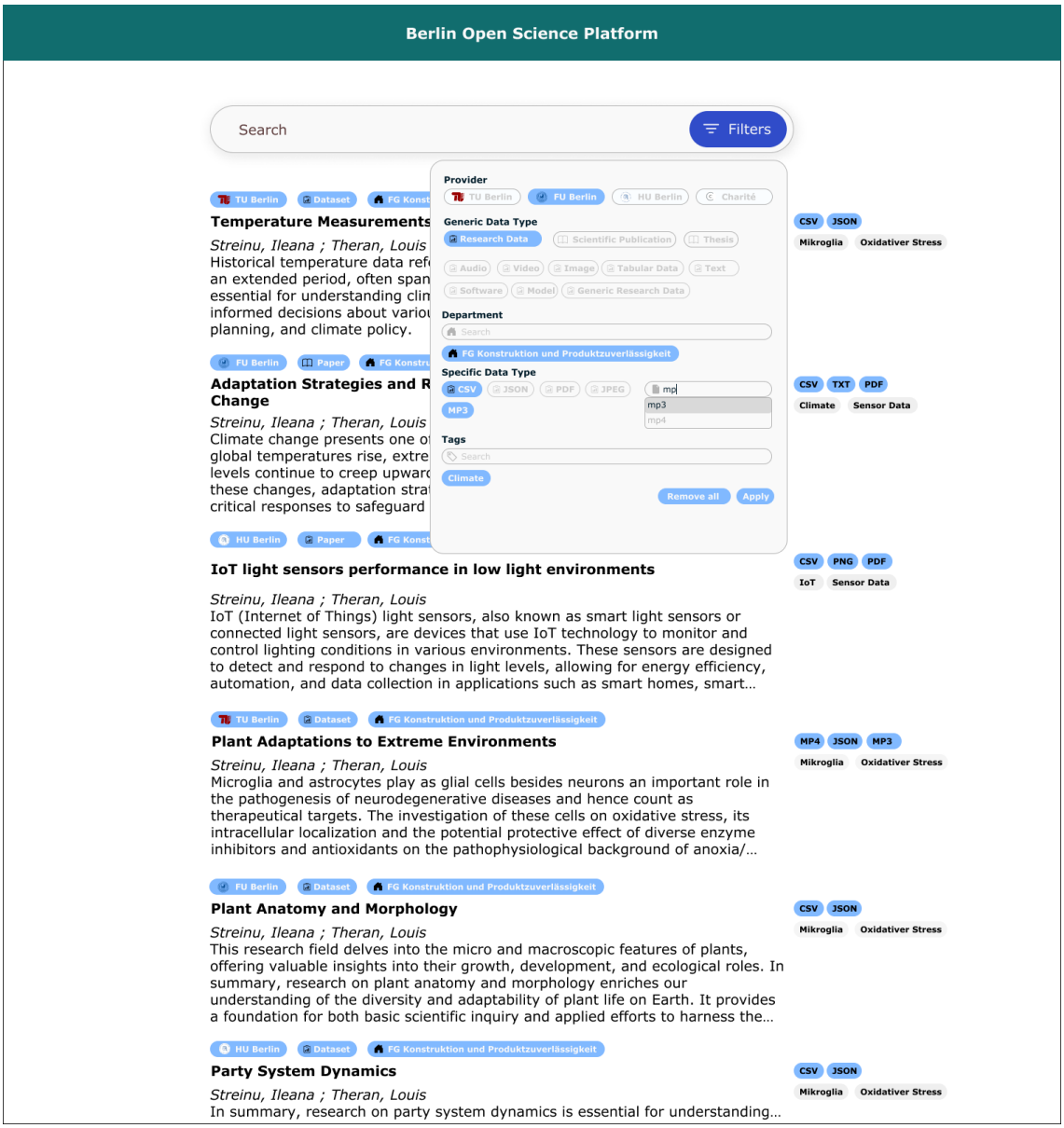

Figure 12: Search Results - Filters Extended

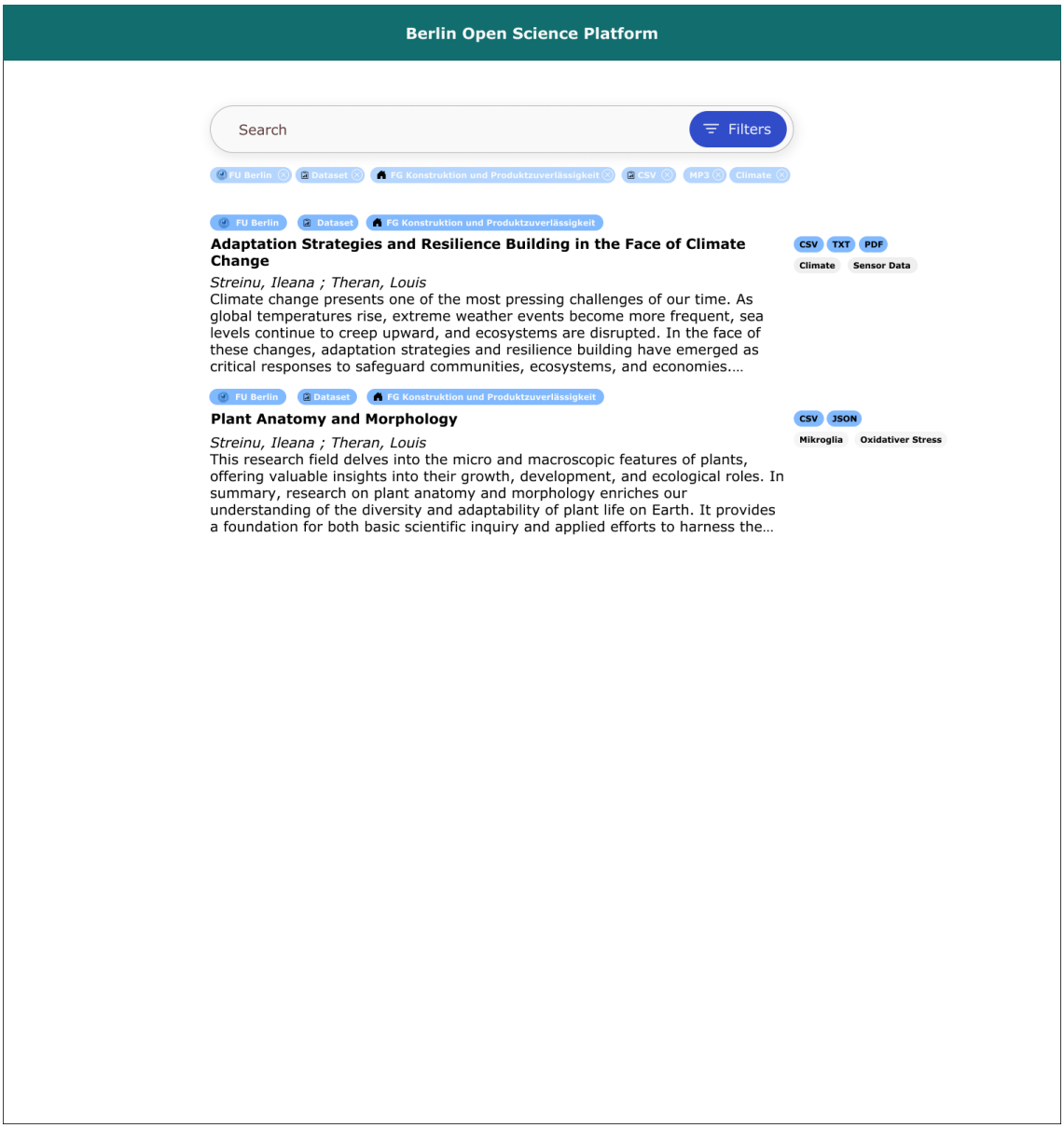

Figure 13: Search Results - Filters Applied

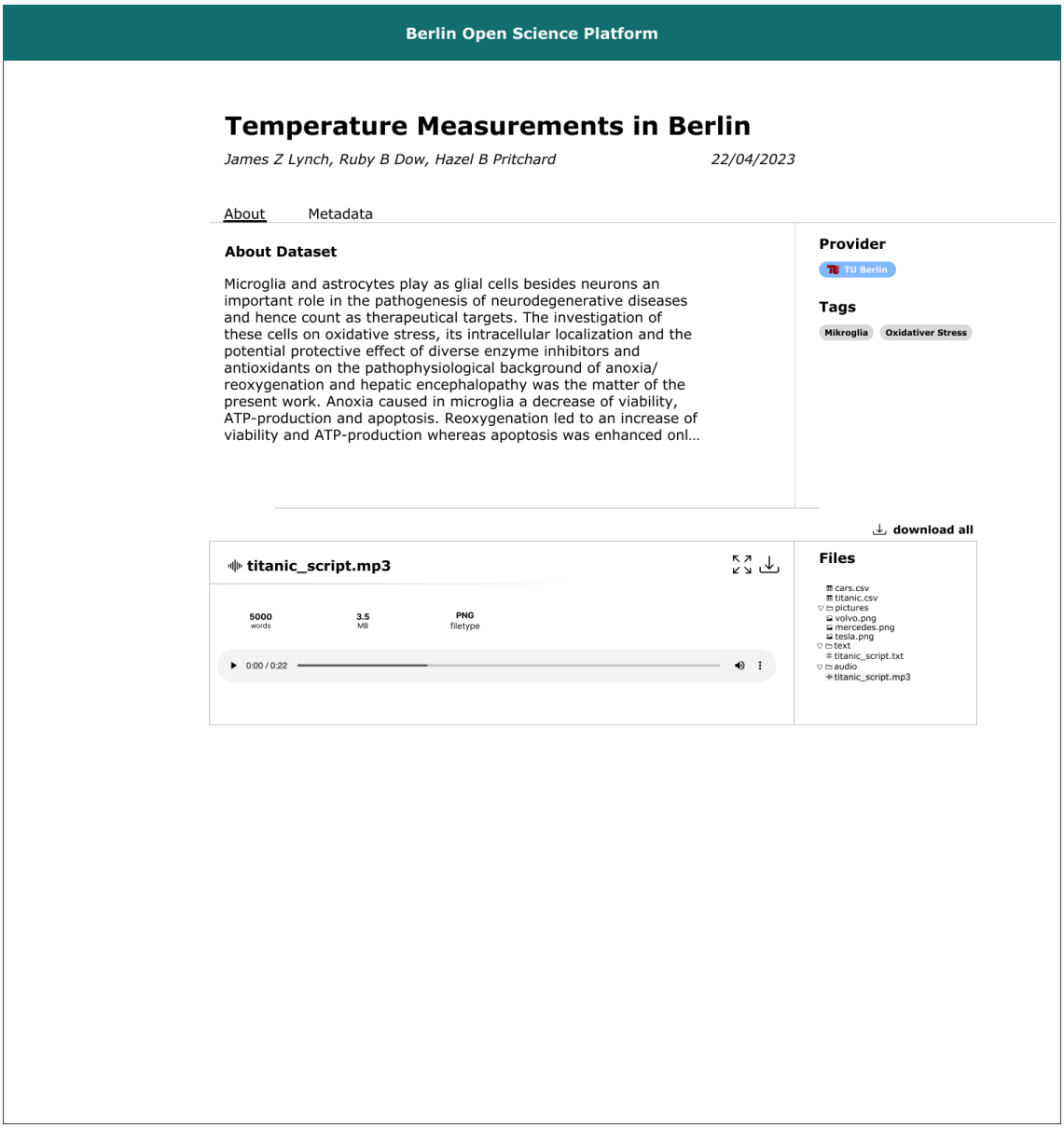

Figure 14: Dataset View - Audio

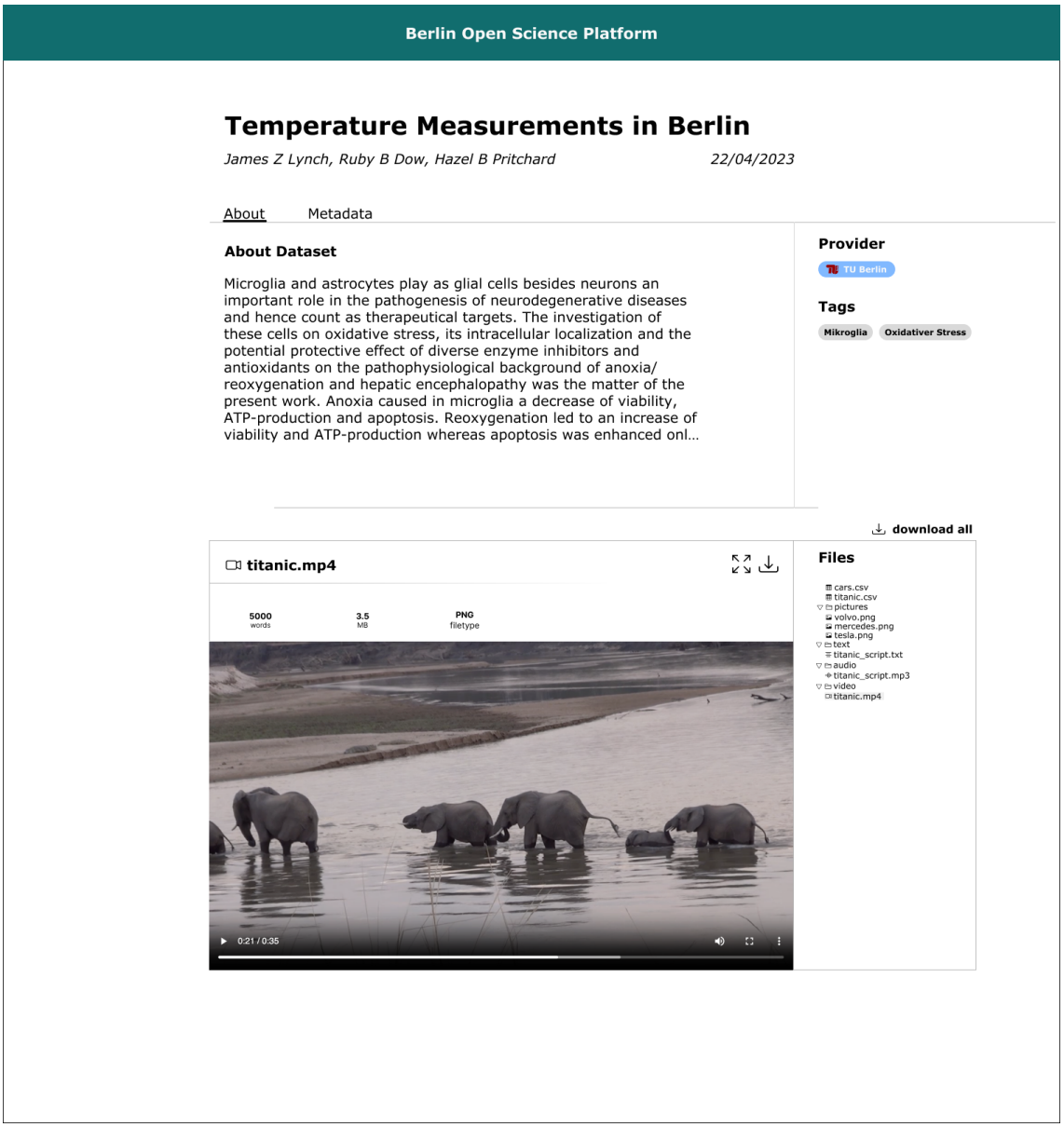

Figure 15: Dataset View - Video

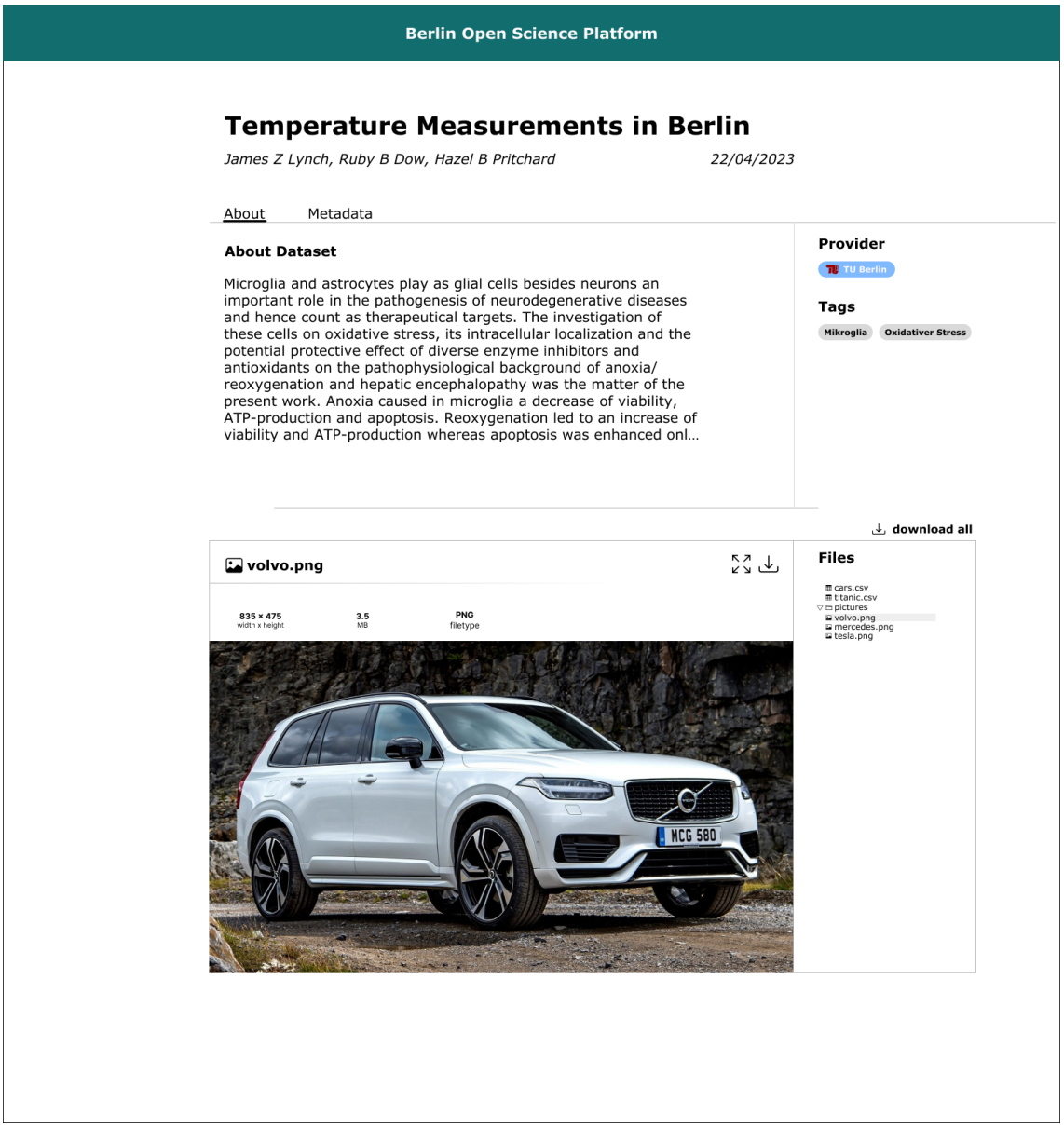

Figure 16: Dataset View - Image

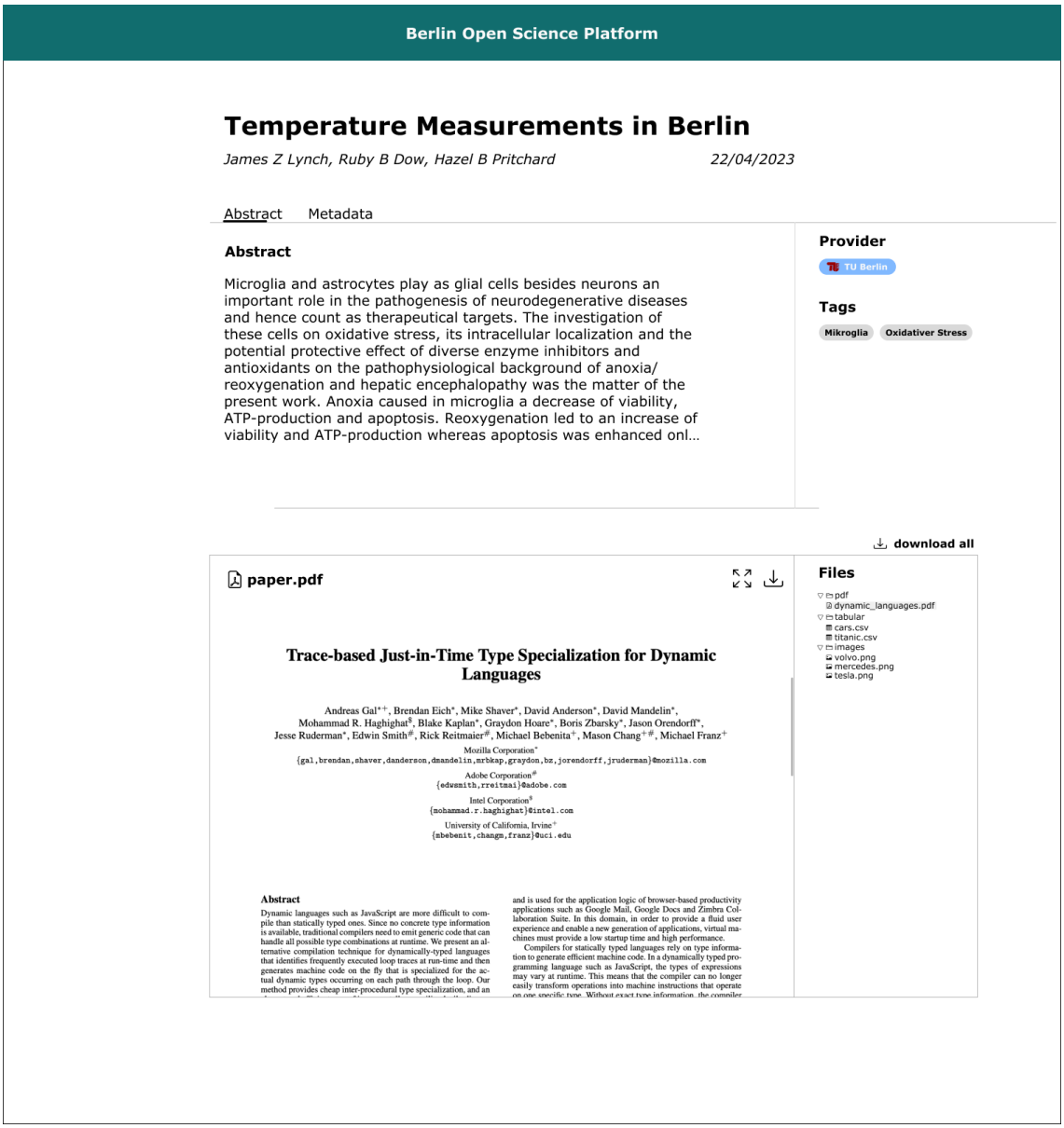

Figure 17: Dataset View - PDF

Figure 18: Dataset View - PDF Fullscreen Fileexplorer Extended

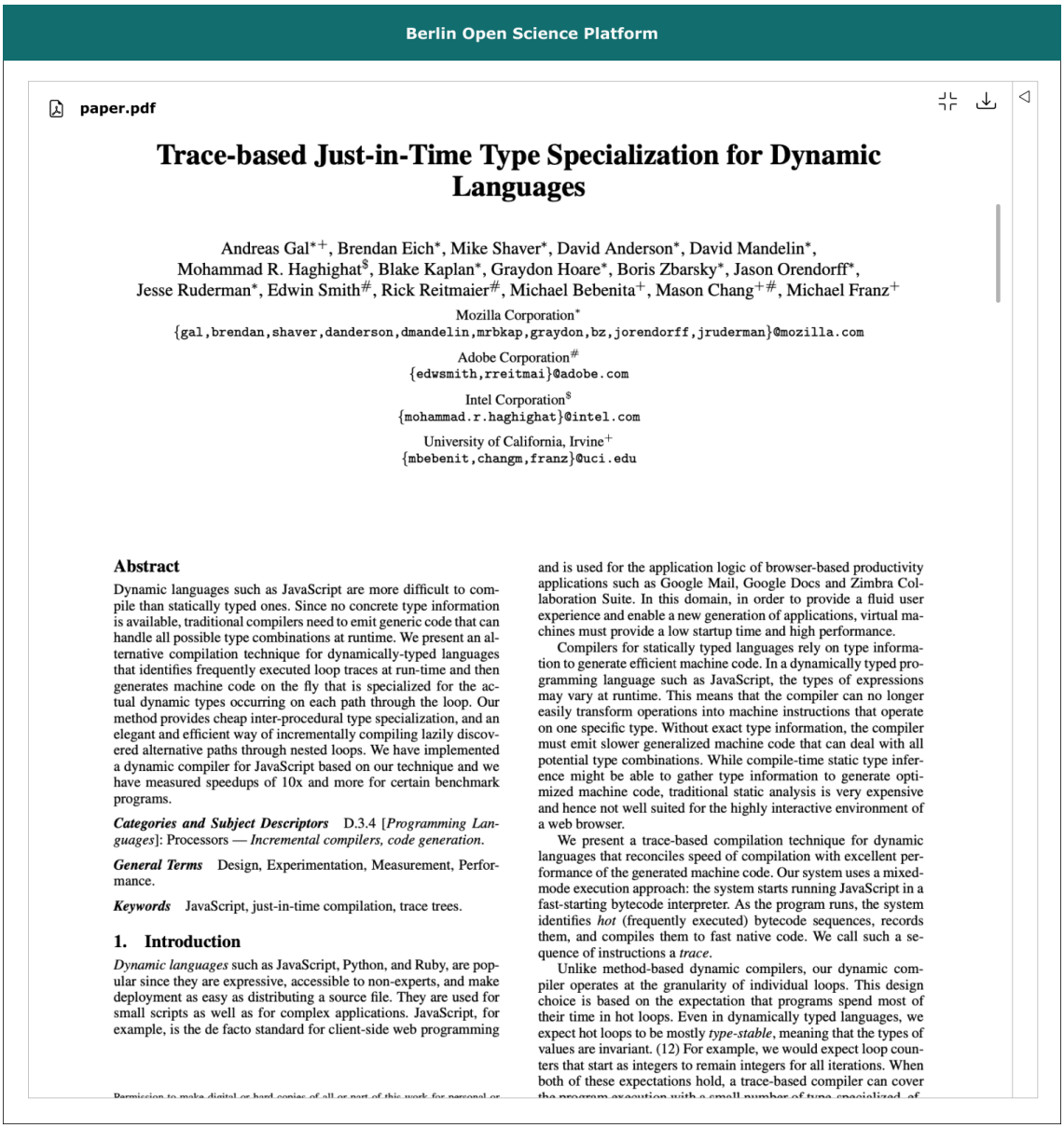

Figure 19: Dataset View - PDF Fullscreen Fileexplorer Hidden

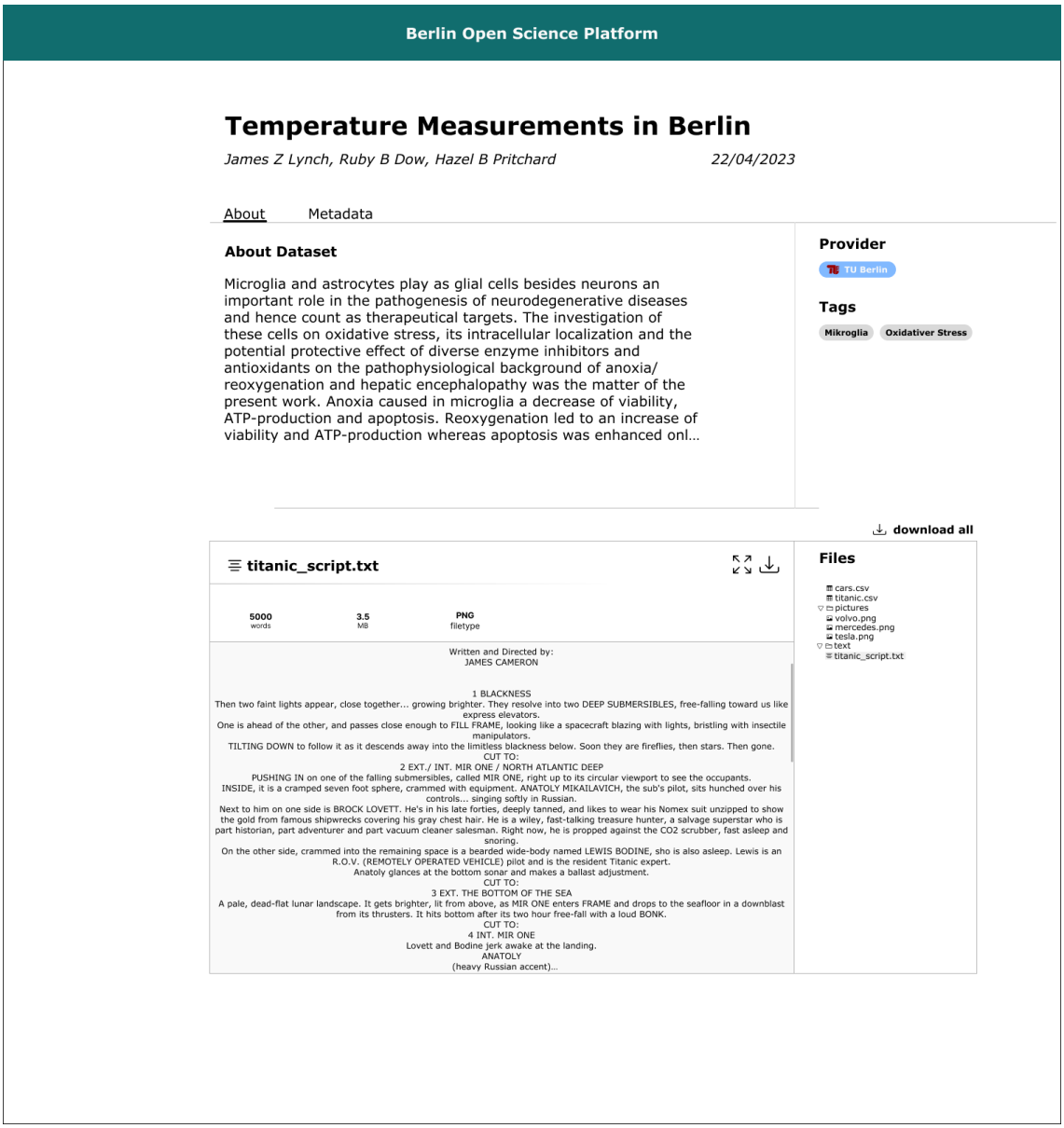

Figure 20: Dataset View - Text

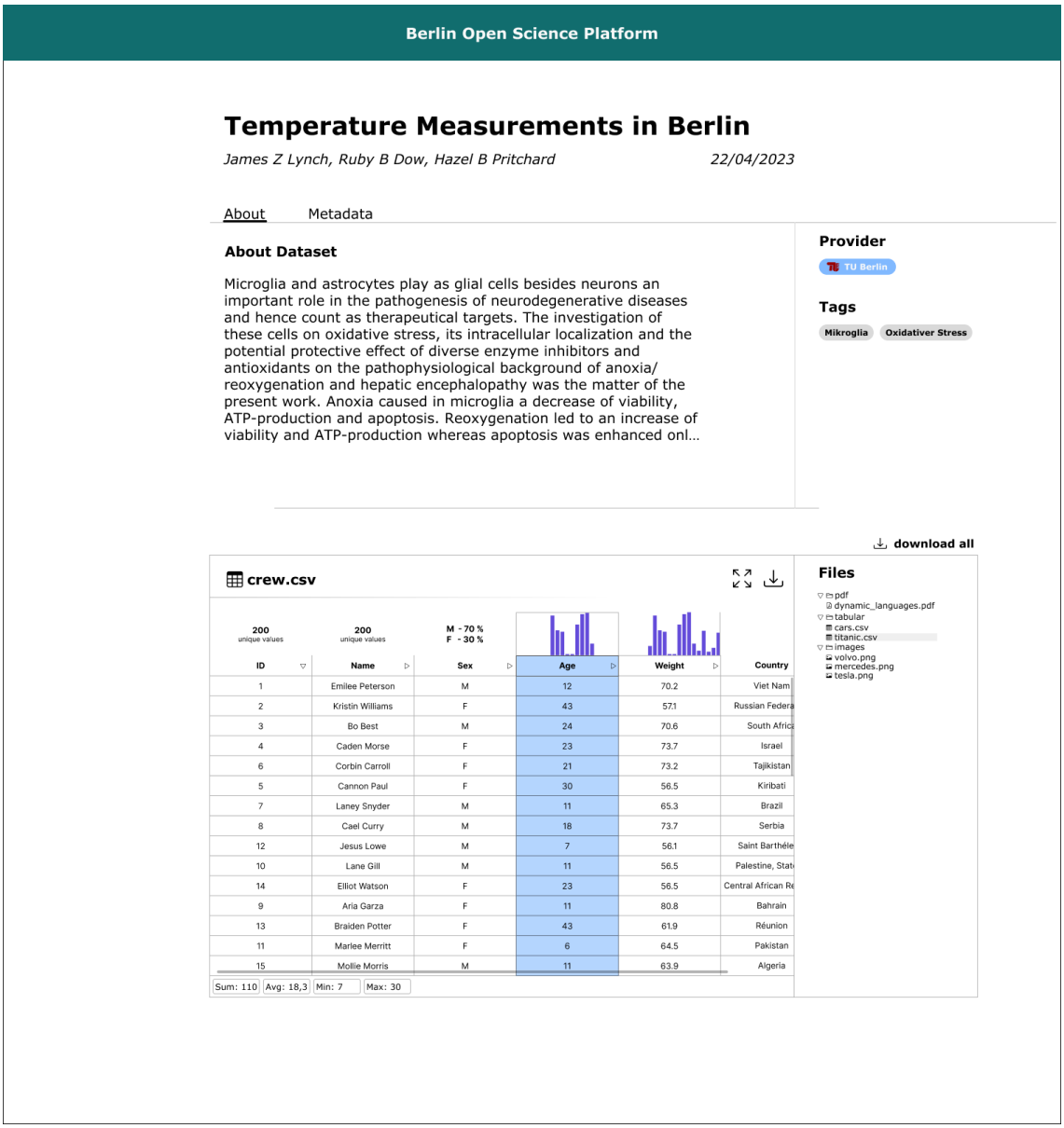

Figure 21: Dataset View - Tabular - About

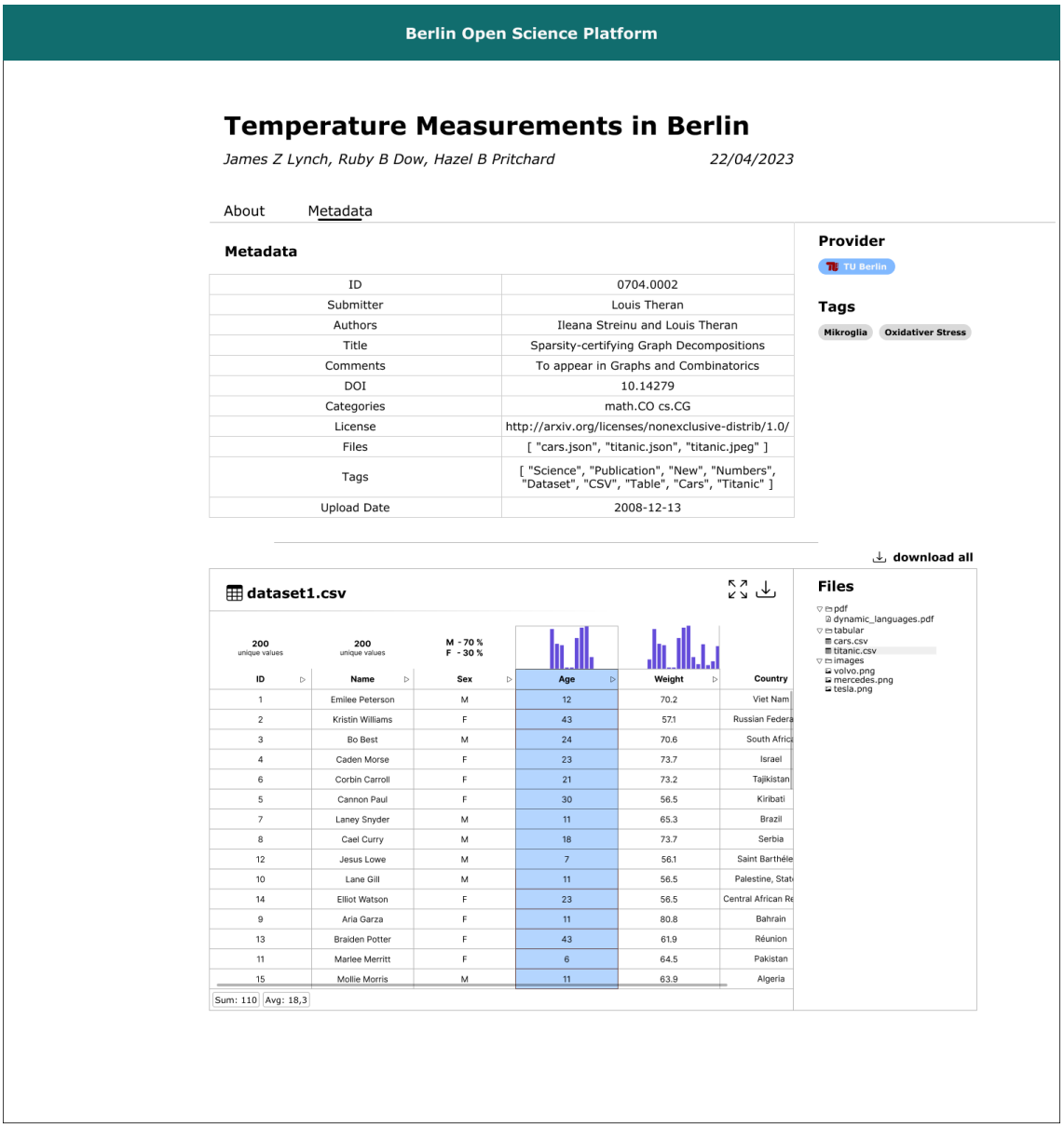

Figure 22: Dataset View - Tabular - Metadata

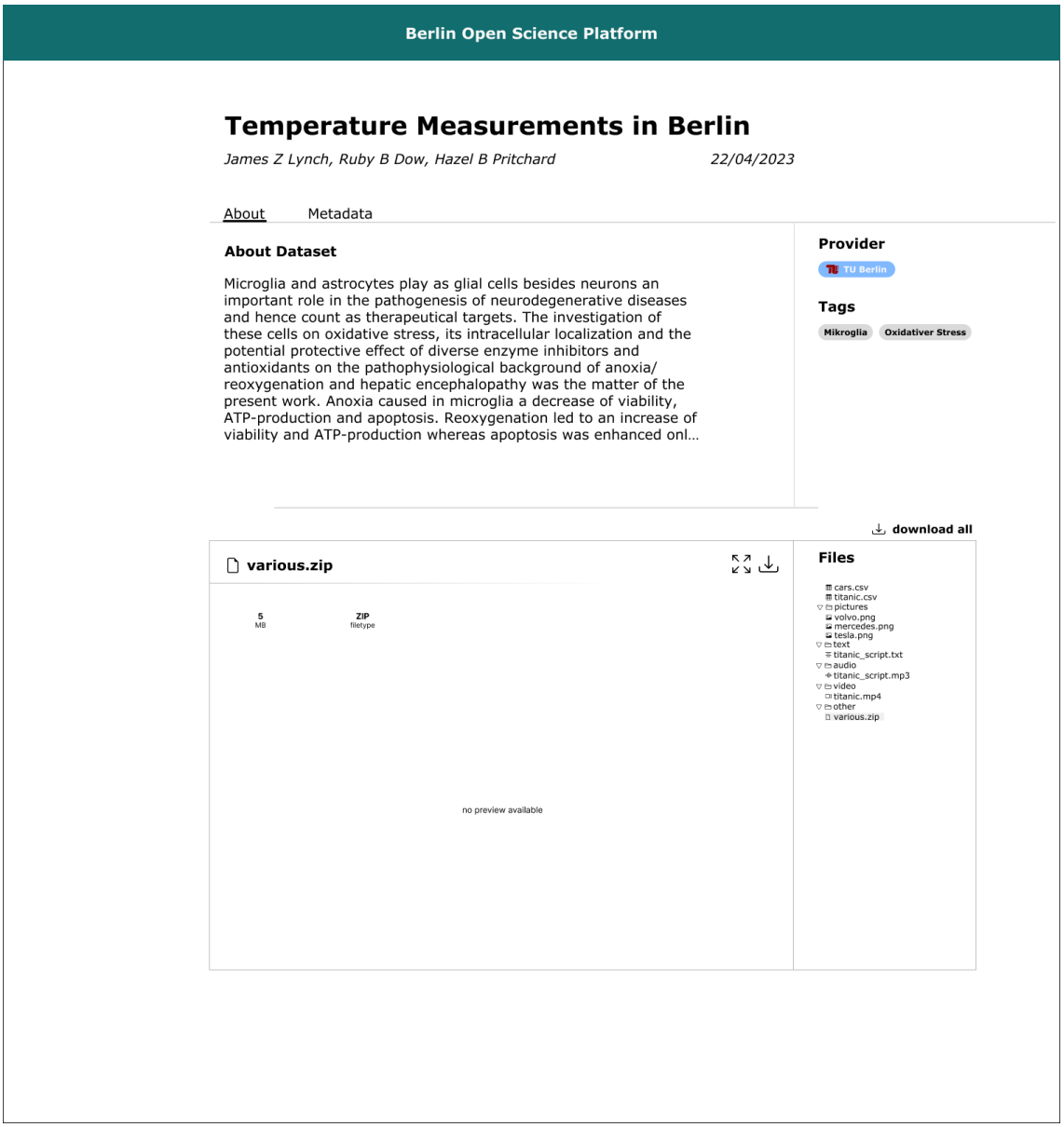

Figure 23: Dataset View - no preview available

C Artifact Images

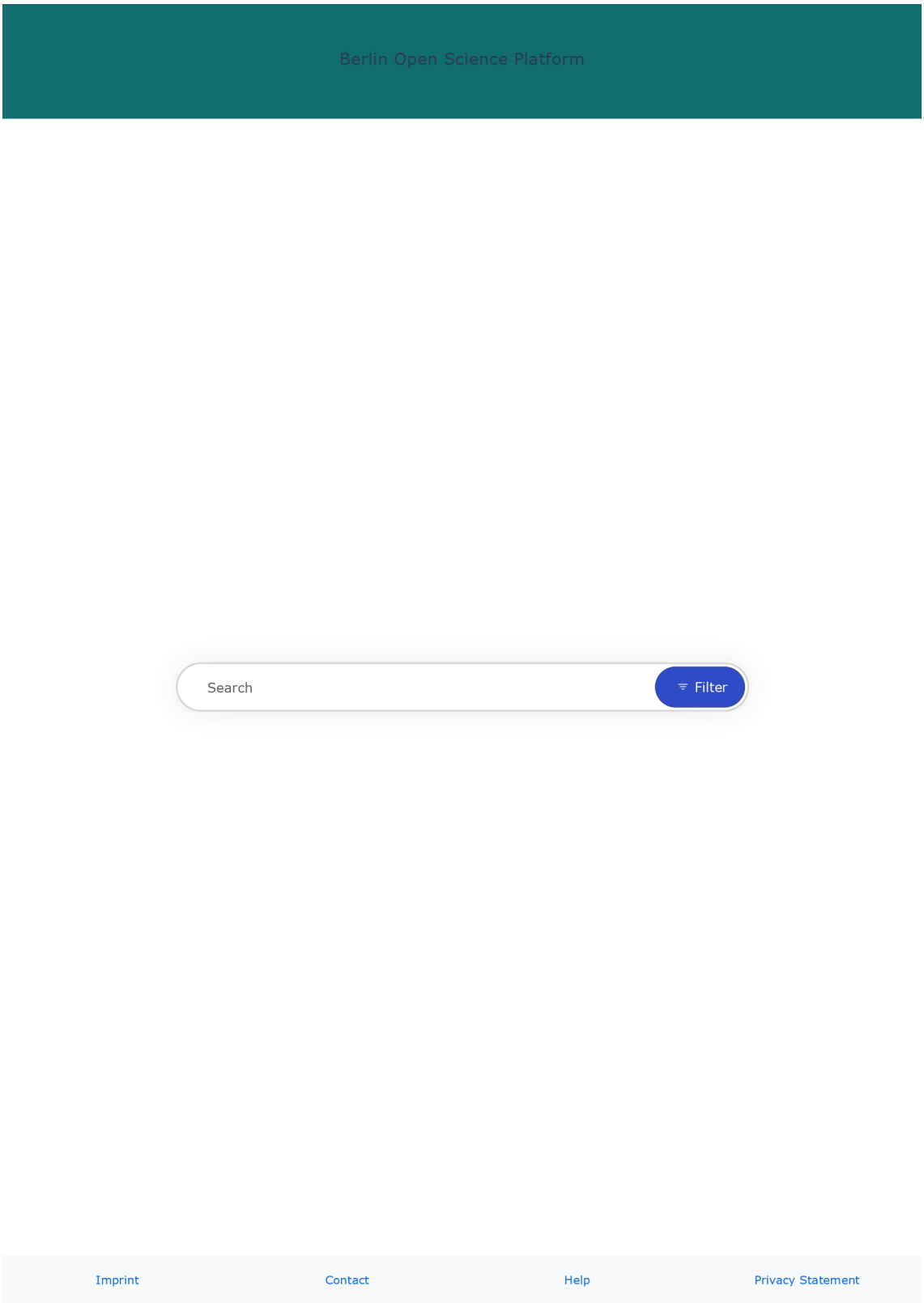

Figure 24: Home Page

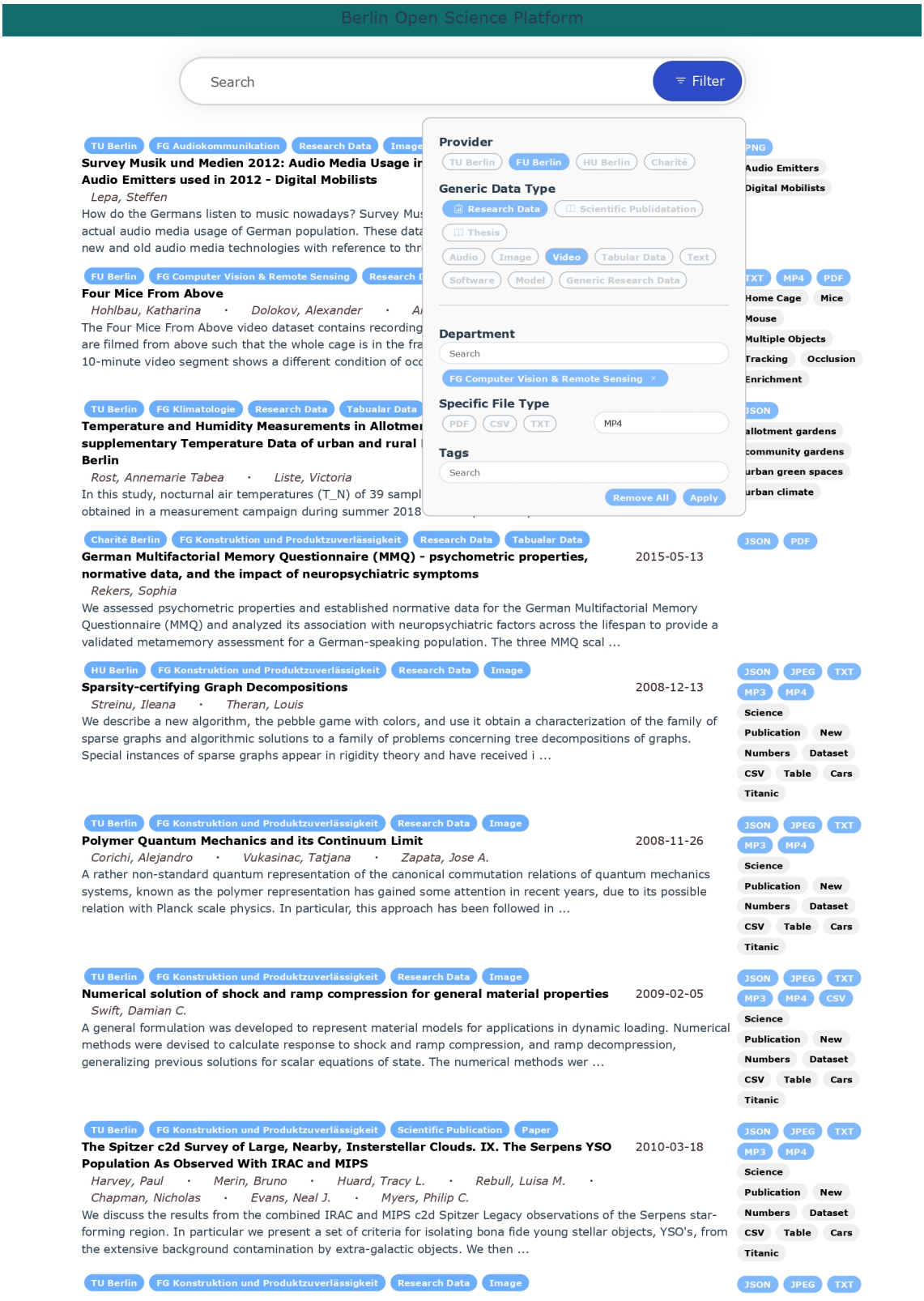

Figure 25: Search Page with Filter Box

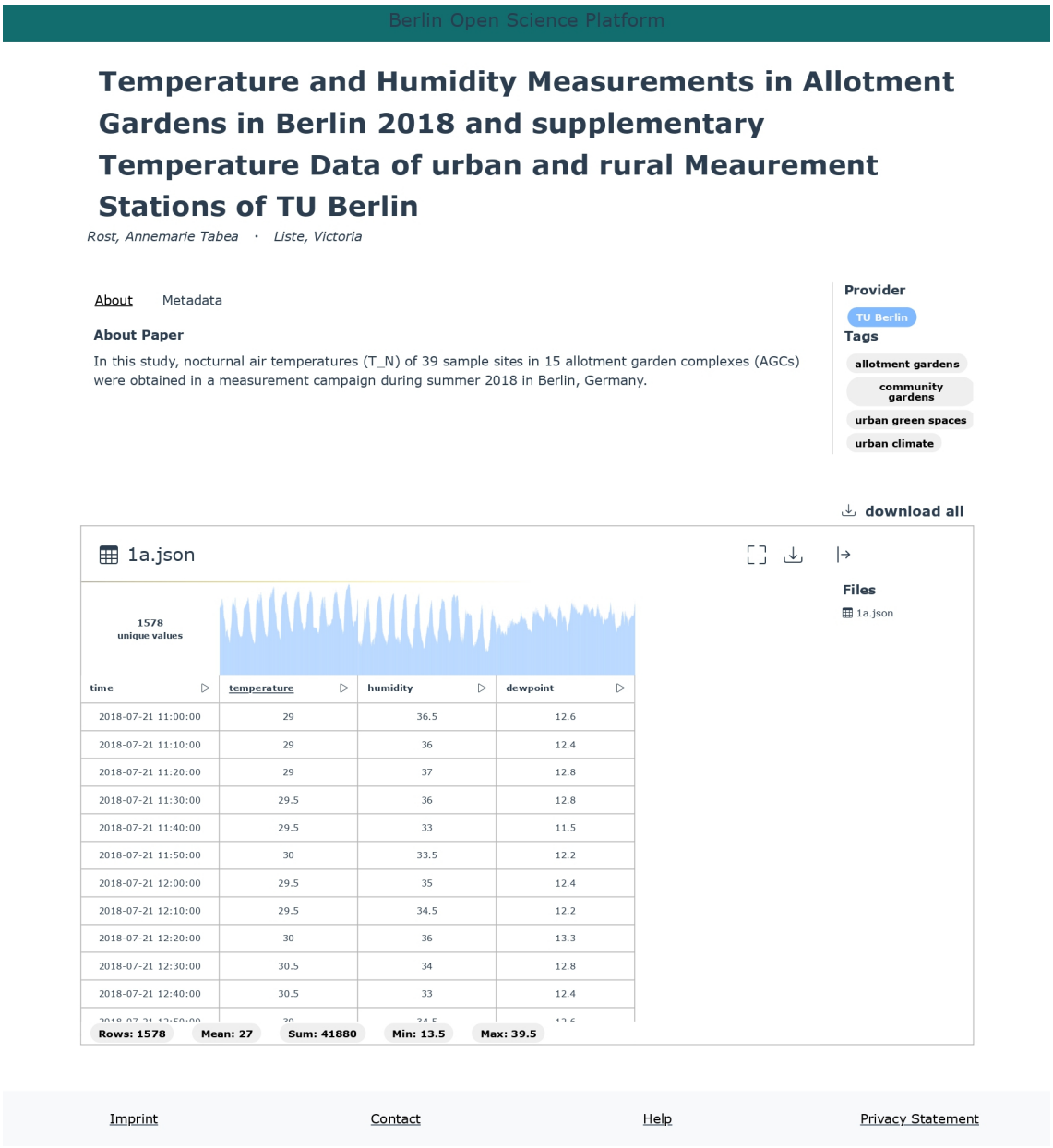

Figure 26: Item Page Tabular

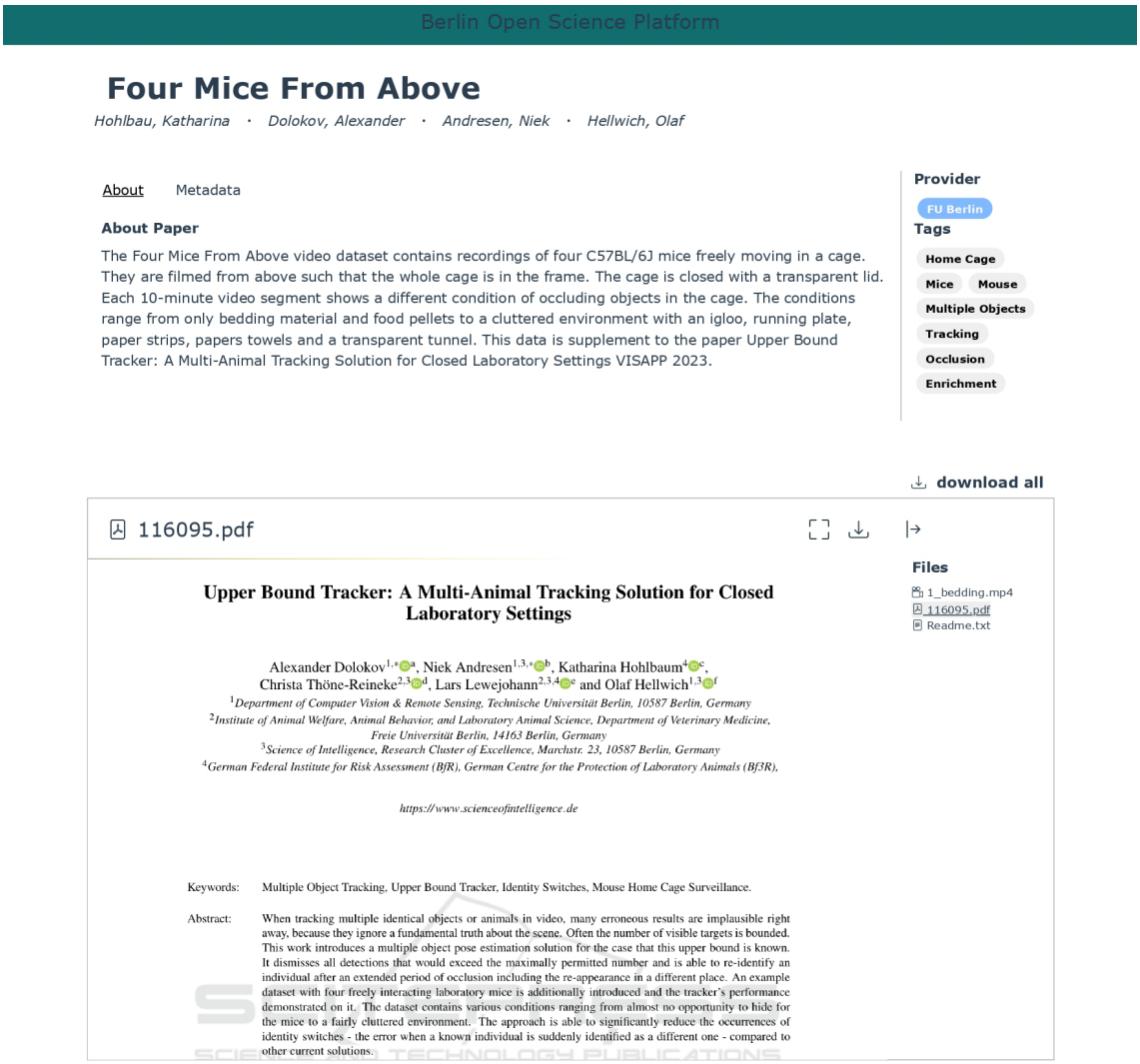

Figure 27: Item Page PDF

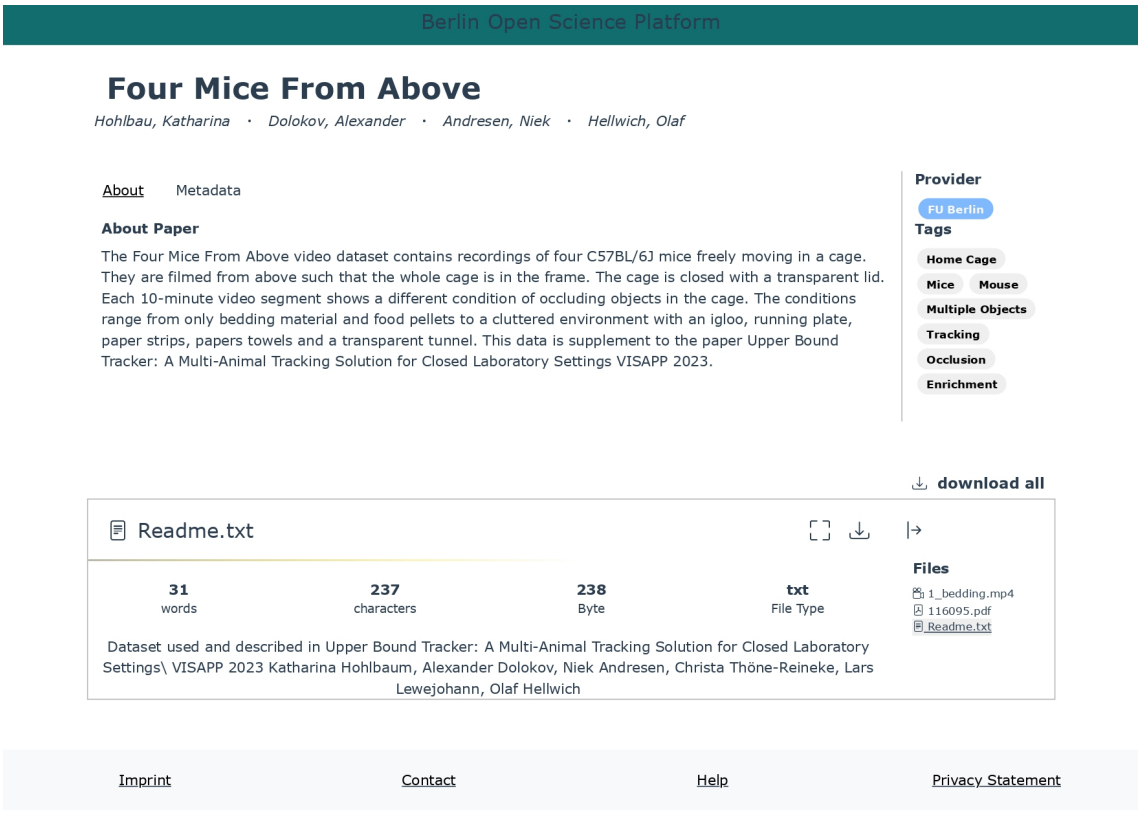

Figure 28: Item Page Text

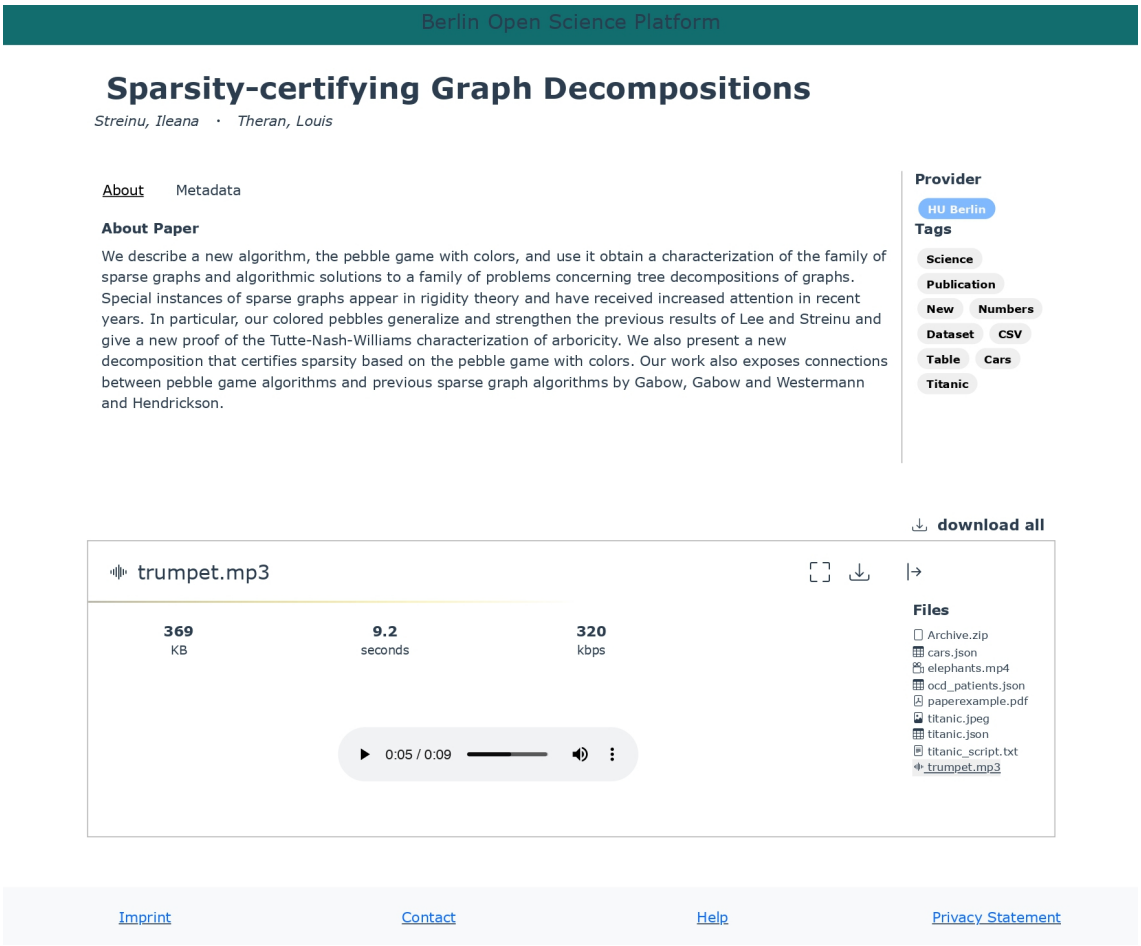

Figure 29: Item Page Audio

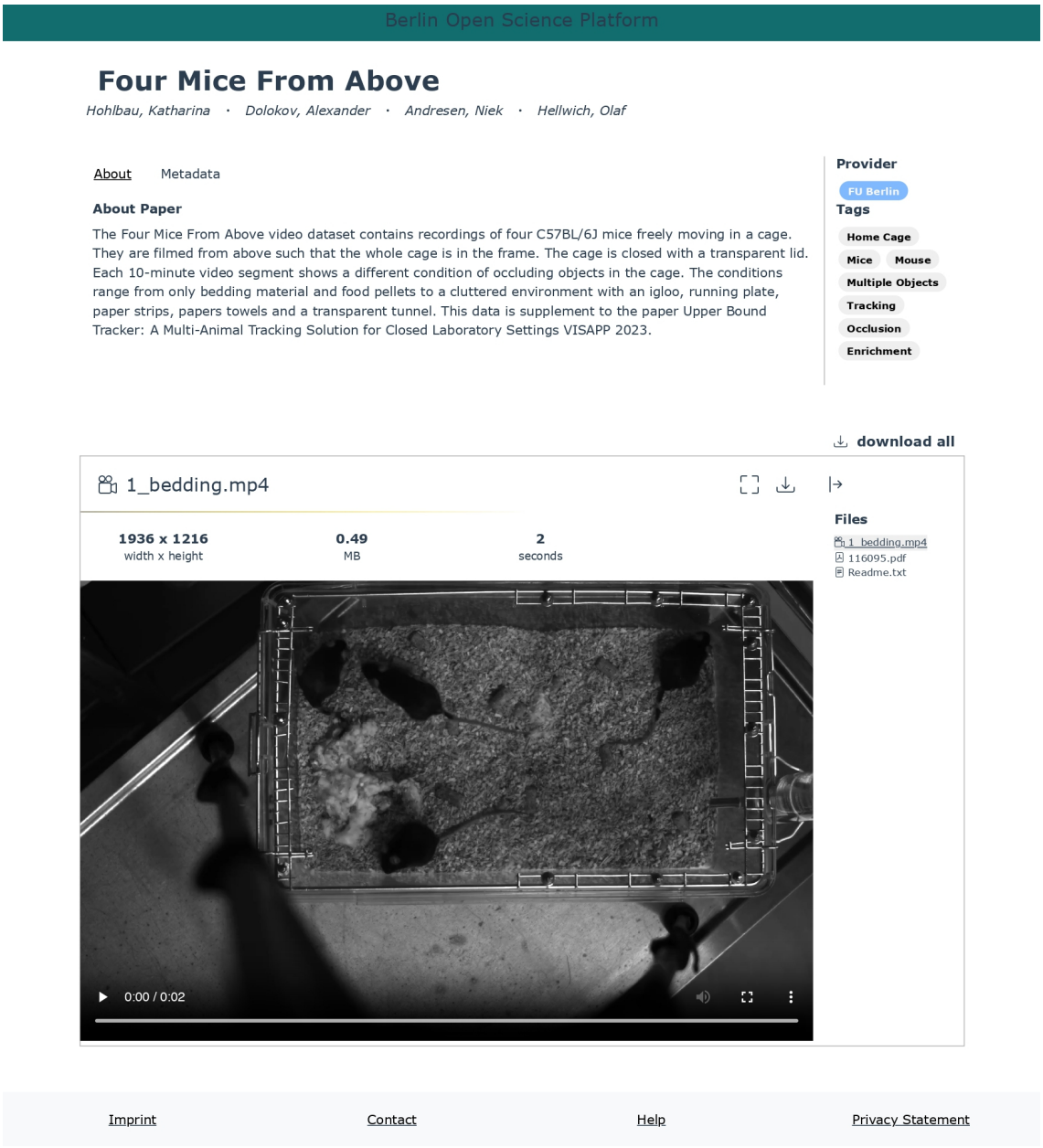

Figure 30: Item Page Video

116095.pdf

Files
1_bedding.mp4
116095.pdf
Readme.txt

Upper Bound Tracker: A Multi-Animal Tracking Solution for Closed Laboratory Settings

Alexander Dolokov[1,*][a], Niek Andresen[1,3,*][b], Katharina Hohlbaum[4][c], Christa Thöne-Reineke[2,3][d], Lars Lewejohann[2,3,4][e] and Olaf Hellwich[1,3][f]

[1]*Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany*
[2]*Institute of Animal Welfare, Animal Behavior, and Laboratory Animal Science, Department of Veterinary Medicine, Freie Universität Berlin, 14163 Berlin, Germany*
[3]*Science of Intelligence, Research Cluster of Excellence, Marchstr. 23, 10587 Berlin, Germany*
[4]*German Federal Institute for Risk Assessment (BfR), German Centre for the Protection of Laboratory Animals (Bf3R),*

https://www.scienceofintelligence.de

Keywords: Multiple Object Tracking, Upper Bound Tracker, Identity Switches, Mouse Home Cage Surveillance.

Abstract: When tracking multiple identical objects or animals in video, many erroneous results are implausible right away, because they ignore a fundamental truth about the scene. Often the number of visible targets is bounded. This work introduces a multiple object pose estimation solution for the case that this upper bound is known. It dismisses all detections that would exceed the maximally permitted number and is able to re-identify an individual after an extended period of occlusion including the re-appearance in a different place. An example dataset with four freely interacting laboratory mice is additionally introduced and the tracker's performance demonstrated on it. The dataset contains various conditions ranging from almost no opportunity to hide for the mice to a fairly cluttered environment. The approach is able to significantly reduce the occurrences of identity switches - the error when a known individual is suddenly identified as a different one - compared to other current solutions.

1 INTRODUCTION

Automatic video analysis often requires tracking of specific objects in the scene. That means a computer system has to be able to recognize and localize something, which it has been told to follow, in every frame of a video. In the application to observing animals there can be the additional requirement to track not only one individual and its body parts, but multiple simultaneously. To the human observer individuals can appear identical, while - through the utilization of visual appearance and the time component - the system has to be able to distinguish and identify them.

1.1 Multiple Object Tracking and Pose Estimation

Multiple Object Tracking (MOT) is challenging and solutions are often not good enough without human correction of error. In this work, we consider the case, where a number of nearly identical individuals and their pre-defined (body) parts should be tracked across all frames (Multi-Object Pose Estimation). Our contribution is not limited to the task of pose estimation, but can also be used for situations where no keypoints play a role. Since it is most useful in laboratory animal settings, in which pose is often necessary, we present it in the Multi-Object Pose Estimation context.

1.2 Typical Frameworks

A typical Multi-Object Pose Estimation framework performs three steps (top-down, Figure 1 (a)): 1) Object Detection, 2) Body Part Detection and 3) Track-

[a] https://orcid.org/0000-0003-0207-4372
[b] https://orcid.org/0000-0002-3596-0795
[c] https://orcid.org/0000-0001-6681-9367
[d] https://orcid.org/0000-0003-0782-2755
[e] https://orcid.org/0000-0002-0202-4351
[f] https://orcid.org/0000-0002-2871-9266
*These authors contributed equally to this work

Dolokov, A., Andresen, N., Hohlbaum, K., Thöne-Reineke, C., Lewejohann, L. and Hellwich, O.
Upper Bound Tracker: A Multi-Animal Tracking Solution for Closed Laboratory Settings.
DOI: 10.5220/0011609500003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 5: VISAPP*, pages 945-952
ISBN: 978-989-758-634-7; ISSN: 2184-4321

Figure 31: Item Page Full Screen PDF

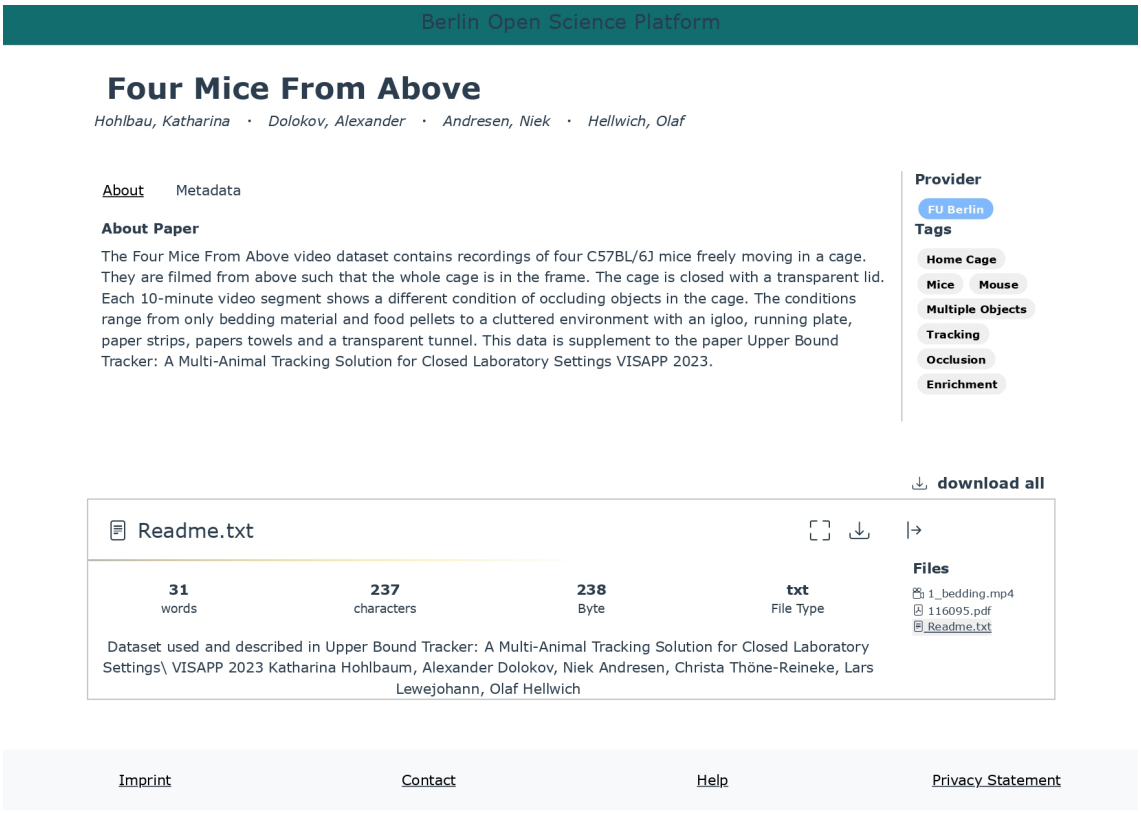

Figure 32: Item Page Text