**Term Project Paper**

Lisa M. Cook

Department of Research Methods and Information Science, University of Denver

LIS 4220 Data Curation

Dr. Matthew S. Mayernik

March 08, 2021

**Introduction**

Everyday millions of people use social media, such as Twitter, to connect other people with similar health issues. Twitter collects, analyzes, and stores all data used on its platform. They share the collected data with external sources, such as social media researchers, while also offering analytic tools for use on their data. This paper aims to provide information on the curation process of Grover, Kar, and Davies' (2018) data set of tweets that focused on discussions about health and technology on Twitter.

**Data Information**

Grover, Kar, and Davies collected tweets as their data to study users' behavior on discussing health issues with technology (Grover et al., 2018). The data set does not contain any restrictions on the use of it. The Mendeley Data repository displays the Creative Commons Attribution 4.0 International (CC By 4.0), meaning the data set is free to use, copy, share, and adapt in any format for personal, commercial, and research use for the data set. Unfortunately, the data set is incomplete because it lacks the data sample, the study's time frame, the methodology, and the stakeholders' information (Kar et al., 2018).

The next step was to review Gover, Kar, and Davies' (2018) article based on the data set to find the missing information. As luck held, their article contained most of the missing information needed for the data set. For example, the data sample consists of 105,489 tweets from 15,587 unique users, which is defined as people who discussed health issues with technology on Twitter (Grover et al., 2018). Gover, Kar, and Davies employed social media analytics from Twitter, such as the CUP Framework with word analysis, network analysis, and hashtag analysis, as the scientific methods used for the data collection, the data analysis, and the data presentation (2018). Gover, Kar, and Davies used these six hashtags: #healthtech, #digitalhealth, #mhealth, #medtech, #healthit, and #ehealth to identify conversations from Twitter for their research(2018). The actor-network theory is used to view the data because it treats humans and technology as inseparable (Grover et al., 2018). The stakeholders who will benefit from the research are health educators, internet users, marketing firms, health technology companies, health-oriented companies, the medical profession, and the government (Grover et al., 2018). Gover, Kar, and Davies' (2018) article did not contain any information on the timeline when the data was collected and analyzed. The time duration is valuable information needed for the metadata to help promote the use of the data set. An attempt was made to contact Gover and Kar through their emails listed on the article to gain the timeframe information. Regrettably, the email goes unanswered at this time.

**Data Set File Types**

The data set contains two files. The first file is called TotalHealthData.CSV; the file format is a Comma Separated Values used to save tabular data as plain text (Kar et al., 2018). Any spreadsheet program, such as Microsoft Excel, OpenOffice Calc, or Google Docs, can open it. The TotalHealthData.CSV file contains the tweets, organized and stored as tabular data, broken down by ten columns labeled: author.screen\_me, text, retweet\_count, favorite\_count, in\_reply\_to\_screen\_me, favorited, retweeted, lang, author.location, and author.description (Kar et al., 2018). There are 5000 rows used to list the subjects' screen name under the column labeled author.screen\_me (Kar et al., 2018).

The second file is named 8964131.xml; the file format is an Extensible Markup Language (XML) (Kar et al., 2018). It requires an XML file app to open and use because the format is written in markup symbols. The data description is stored in the 8964131.xml file because it organizes, stores, and transmits it over the internet. Search engines use web crawlers to find the 8964131.xml file to make it easier to discover the data set for research.

**Metadata**

The 8964131.xml file, as previously mentioned, is used to store the metadata. Dublin Core is the simple metadata standard used to describe the data set (Kar et al., 2018). The 8964131.xml file contains the minimum number of Dublin Core elements that are required to describe the data (Kar et al., 2018).

For example:

* **Title**: Data for: "Technology enabled Health" – Insights from Twitter Analytics with a Socio-Technical Perspective
* **creators:** Arpan Ka and Purva Gover
* **Identifier type**: DOI:0.17632/rs3c243fnm)
* **Publisher:** Mendeley Figshare web address
* **Description:** This data set consists of all the tweets collected and segregated after extraction on the theme of digital and technology enabled health. This data set has been analyzed with Twitter analytics for establishing the propositions highlighted in our study.
* **Date**: time 2019-07-19 10:19:56 (Uploaded to Figshare)
* **Subject:** Social Media Marketing (Kar et al., 2018).

**Metadata Improvements**

The data set cannot stand alone in its current form because it needs Gover, Kar, and Davies' (2018) article as a reference source to fill in the missing information. By adding more information from the article to the metadata, it helps the data set become self-describing. Having more descriptive information is needed to improve the metadata to make the data set more usable by other researchers. One of the best ways to improve the metadata is to add more information, like the sample size, the actor-network theory, the CUP framework approach, and the research questions on the account. It will increase the chances for usability.

When I looked at the metadata, I saw some problems with it. One major problem with the metadata is the missing timeframe used during the research process of the study (Kar et al., 2018). The timeframe is necessary because it is part of the research procedure that permits other researchers to duplicate the same study to verify the results. If people can verify the data set results, it increases the trust in it to use in other studies. Also, the time frame the data was collected helps other researchers decide if this data set meets their criteria in their study. For example, a researcher needs data from the 2020 COVID 19 Pandemic from the Twitter platform to analyze the users' reactions to it; if the potential data set does not contain the research date of 2020, the data set is useless to them.

The second problem is that the metadata only uses one keyword, "Social Media Marketing," limiting its searchability (Kar et al., 2018). The more keywords used to describe the data set increases its findability and potential use. Keywords are the foundation to finding anything online or at libraries because it helps the user narrow down the search topic. Using different vocabulary for keywords helps guide researchers from diverse backgrounds to find the data set faster. Here is a list of keywords to increase the visibility of the data set.

* **Twitter Analytics**
* **Social Media Health Data**
* **Health and Technology Relationships on Twitter**
* **Health Technology**
* **Health Domains**
* **Hashtag Analysis on Health and Technology**
* **Disease Discussion on Twitter**
* **The Actor-Network Theory**
* **Word Analyses**
* **The CUP Framework Approach**

The third issue is the lack of supplement material to explain the value and the purpose of using Twitter Analytics on the data set (Kar et al., 2018). The supplement material helps researchers understand the data set results better, especially if the researcher is unfamiliar with how Twitter Analytics is used on the data set. The data set’s metadata does not include any supplement material on Twitter Analytics (Kar et al., 2018). It would be an excellent idea to include a link in the metadata section in relation to the data set about adding an article that explains why the researchers used Twitter Analysis to evaluate the social media data. I recommend adding this article, *How to Use Twitter Analytics: The Complete Guide for Marketers* by Karin Olafson, to the metadata to help potential users understand Twitter Analytics (Olafson, 2020).

**Metadata Example**

|  |  |  |  |
| --- | --- | --- | --- |
| ID | DataCite-Property | Notes | Allowed values, examples, other constraints |
| 1 | Identifier  identifierType |  | 10.17632/rs3c243fnm.1  DOI |
| 2 | Creator  creatorName  nameType  givenName  familyName  Creator  creatorName  nameType  givenName  familyName |  | Kar, Arpan  Personal  Arpan  Kar  Grover, Purva  Personal  Purva  Grover |
| 3 | Title |  | *Data for: “Technology enabled Health” – Insights from twitter analytics with a socio-technical perspective* |
| 4 | Publisher | The repository that stores the data set. | Mirror of Mendeley Data |
| 5 | PublicationYear | Form yyyy | 2019 |
| 6 | Subject  subjectScheme  schemeURI | Free Text | Social Media Marketing; Twitter Analysis; Social Media Health Data; Health and technology Relationships on Twitter; Health Technology; Health Domains; Hashtag Analysis on Health and Technology; Diseases Discussion on Twitter; The Actor-Network Theory; The CUP Framework Approach; Word Analysis  Free Text  <http://purl.org/dc/dcmitype/Dataset> |
| 7 | Contributor  contributorType  contributorName  familyName  givenName  Contributor  contributorType  contributorName  familyName  givenName |  | DataCollector  Personal  Davies, Gareth  Davies  Gareth  DataCurator  Personal  Cook  Liz |
| 8 | Date  dateType  Date  dateType | Time frame unknown for the research period  (DataCite) | 2019-07-19 10:19:56  Issued  :unav  Created |
| 9 | Language |  | en-US |
| 10 | ResourceType  resourceTypeGeneral |  | Twitter’s hashtag  Dataset |
| 11 | alternateIdentifier  alternateIdentifierType |  | N/A |
| 12 | RelatedIdentifier  relatedIdentifierType  relationType  RelatedIdentifier  relatedIdentifierType  relationType | An article to explain how to use Twitter Analytics  The article that explains the data set. | UR  IsSupplementTo  <https://blog.hootsuite.com/twitter-analytics-guide/>  DOI  IsDocumentedBY  https://doi.org/10.1016/j.ijinfomgt.2018.07.003. |
| 13 | Size | The size of the CSV file that holds the data.  Free Text | 24.7 MB |
| 14 | Format | Use file extension  Free Text  Technical format of the resource. | application/CSV text/tab-separated-values |
| 15 | Version | XML version to record the metadata | 1.0 |
| 16 | Rights  rightsURI  rightsIdentifier  rightsIdentifierScheme  schemeURI | Free Text  URL | https://creativecommons.org/licenses/by/4.0/  https://creativecommons.org/licenses/by/4.0/legalcode  Cc-by-4.0  SPDX  <https://spdx.org/licenses> |
| 17 | Description  descriptionType | Free Text | This data set consists of all the tweets collected and segregated after extraction on the theme of digital and technology enabled health. This data set has been analyzed with Twitter analytics, the CUP framework approach, word analyzed, and the actor-network theory to answer the three research questions. RQ1 What are the popular information technologies in the health domain? RQ2: Which type of diseases, i.e. acute or chronic; communicable or non-communicable diseases being discussed more on social media? RQ3: How technological solutions have been associated with various diseases?  Abstract |

**Publication**

I realized I was going to have problems finding any publications based on the data set that uses only one keyword and has no time frame when the researchers collected the data. This investigation was not going to be easy. The search for the data set cited by other articles was going to be challenging at the least. What is the first step in finding statistical information on a data set cited by other publications? Let's start at the beginning by looking at the metadata record on Figshare.com (Kar et al., 2018). The data set's landing page keeps track of the statistics of the data set. It was not surprising to see the sad results on how little anyone viewed the data set and use (Kar et al., 2018). The facts do not lie that this data set needs a lot of help to get noticed and used. The data set had 0 citations, 17 downloads, and 130 views (Kar et al., 2018).

The next step was looking at the article's the data set was used to create. The statistics were still horrible for the data set. The article was credited with 37 citations with no citations for the data set (Grover et al., 2018). Finally, it was time to use Google Search to use the data set, which fails me. Google Search results led me to the Mendeley data page, which displayed 1146 views and 51 downloads with 0 citations (Kar, 2018). Another Google search led to the websites dataone.org (2021) and [researchdata.ands.org.au](https://researchdata.ands.org.au/) (2021) with zero results for both of them. There is no evidence that the data set is used in any other publication besides the researcher's article.

**Metadata Solutions**

The data set and the metadata both need a lot of work to create a self-describing data set. The one major issue stopping the data set from being used by other studies is the missing research time frame. The best solution to fix the missing time frame was by reaching out to the researcher through email to provide it. The researchers have not yet responded. All the other issues are fixable with a lot of work and research from Grover, Kar, and Davies’*(2018)* article*, “Technology enabled Health’-Insights from twitter analytics with a social-technical perspective.*

**The Requirements for a Repository**

When the time came to search for a repository to store my data set, I was totally lost. I reread the articles assigned in class, but I was still unsure which direction was the correct path for the data set. I was not sure where to start. How do I go about finding a repository to store social media data? Let's say my brain answered the question with silence. The idea on how to find a home for the data set materialized by accident because I stumbled across a video on YouTube called *The Power in Effective Data Storytelling* by Malavica Sridhar (TEDxTalks, 2018).

The analogy of finding a good story to read is useful in finding a repository consists of the same concepts. I was armed with a set of questions to help define the type of repository for the data set by thinking of the type of story I want to read (TEDxTalks, 2018). First, I decide on the type of world that I want to immerse myself in for the next few days (TEDxTalks, 2018). Secondly, I look at the different authors who create the world I want to explore (TEDxTalks, 2018). Thirdly, I explore the different stories written by the chosen author (TEDxTalks, 2018). When I am looking for an excellent story to read, I want a vivid, compelling story to keep me interested in the plot (TEDxTalks, 2018). One of the best parts of a good story is well-written characters who add enrichment to the story. When I find the story I want to read, I study the colophon, the table of contents, the prologue, and the front endpaper to help make the final decision to read the book.

These steps for finding a good story are the same steps to find a suitable repository for storing the data set. First, I need to find a repository that stores social media data, since the data set contains Twitter's conversations. Secondly, I go to re3data.org to look at all the different repositories on the website (Re3data.org, n.d.) . The desired repository must have compelling storage attributes, such as a permanent identifier, open access, accept data sets, and is Interoperable (Re3data.org, n.d.). I typed "data set" in the search box on the re3data web page and marked off the defined characteristics, like Digital Object Identifier (DOI), open access, and Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Re3data.org, n.d.). The search engine returned these six repositories: Research Data Unipd, CaltechDATA, KNB Data Repository, NAKALA, TROLLing, and ioChem-BD (Re3data.org. n.d.). I reviewed the policies and procedures of all six repositories. I picked NAKAKA as the desired repository because it fits all the requirements for the data set.

**New Website**

I ran into a new problem as I worked on improving the second paper about Nakala. In the span of two weeks, Nakala's website changed its web designs from the last time I visited it (Nakala, n.d.). The information in the first paper is obsolete, so I am starting over from scratch. The first change I noticed on their website was Nakala only capitalized its first letter instead of the whole word (Nakala, n.d.). I laughed when I saw on the left top side of the screen a pink one eye monster that looks like it escaped from the movie *Monster Inc*. (Nakala, n.d.). This adorable picture is their new icon that lets users navigate anywhere on the website back to Nakala's homepage (Nakala, n.d.). The new web page offers a button on the right top side of the screen to translate the French text into Spanish or English, which opens the repository to potentially more users (Nakala, n.d.). On the center of the homepage is a large blue button that says upload my data in Nakala (Nakala, n.d.). Below the button, are these words listed “search, cite and reuse scientific data” describing what information the large search engine provides to users (Nakala, n.d.). There is a new web page dedicated to searching, uploading, and deleting data (API, n.d.).The Nakala provides information, explaining how to upload data to the repository, but there is still no human interaction (API, n. d.). The repository asks users to follow all the laws, especially the Data Protection Act and the Formalities required by the Commission Nationale de L'informatique et des Libertés (CNIL) for any data submission (Mentions légales, n. d.) . The database uses SPARQL endpoint of Isidore as the query language to discover structured RDF (ISIDORE, n.d.). A new web page called Open Archival Information (OAI) provides information for identification, list metadata formats, list identifiers, and list records. The website no longer requires an user account to explore and discover data information (*OAI 2.0 Nakala Request Results, n. d.*). Nakala’s website has become more user friendly compared to the frustration I encountered the first time I used the website.

**Repository’s Policies & Procedures**

Lets dive into answering those burning questions about why Nakala was picked as the new home for the social media data. The repository accepts all types of data that relates to humanities and social sciences, including social media data sets, from research teams (Re3data.org n.d.). The procedure for data set approval is made easier with the new web design. It is a simple two-step process for asking permission from Nakala. The first step is creating a HumanID that consists of user id, name, and password (HumanID, n. d.). The second step is asking for permission to access Nakala that consists of affiliation, project title, and project description (HumanID, n. d.). The repository uses the International Image Interoperability Framework (IIIF), Resource Description Framework (RDF), Web Ontology Language (OWl), Handler System, and Digital Object Identifier, (DOI), ensuring data are easy to find and reference online (Re3data.org n.d.). NAKALA uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Re3data.org n.d.). The Repository has a long list of controlled vocabulary accepted, so I am going to list three of them: Library of Congress Subject Headings (LCSH), RAMEAU, and OpenEdition index (Isidore.science/vocabularies, n.d.). The Repository uses Extensible Markup Language (XML) or JavaScript Object Notation (JSON) as schemas to store and distribute metadata (Isidore.science/api, n.d.). Dublin Core is recommended as the main metadata to use for the repository (Huma-Num, 2018), but it accepts all types of metadata to encourage the exchange and dissemination of data (API, n.d.). The repository provides the International Standard Serial Number (ISSN 2495-8972) as their persistent identification number (Re3data.org n.d.). The new website does not provide any information on the size limit for uploading data because the webpage of the constraints of data uploading is not created.

Huma-Num created NAKALA, a non-profit disciplinary institution, to ensure a large digital infrastructure exists for research teams to have interoperable access, and storage to data. Huma-Num is a large research infrastructure that is composed of different research institutions, such as The French National Centre for Scientific Research, Aix-Marseille University, and Campus Condorcet (À propos, n.d.). Huma-Num's mission is to secure data: sound recordings, videos, maps, photographs, observation data, information systems, surveys, databases, and corpus for preservation, enhancement, and uses in the Humanities and Social Science research programs (Huma-Num : l'infrastructure des humanités numériques, n.d.). NAKALA charges a required membership fee displayed on the re3data page; however, this piece of information is missing from the repository’s website (Re3data.org n.d.). An interesting note about Nakala’s terms of access to the date, it specifies only the users from Universities and research teams are the only ones who can have access to the data. But this is not true, I was able to search and download data from the website before I created a user’s account. I was able to download a file in the Portable Document Format (PDF), and a data set file in Comma Separated Values (CSV).

**Life Cycle**

NAKALA offers services at different stages of the data life cycle, such as Archiver, Signaler, Exposer, Diffuser, Traiter, and the Stocker. The Archier is designed for the long-term storage of the Humanities and Social Sciences digital data (Services et outils, n.d.). The National Computer Center for Higher Education (CINES) has certified Nakala for use because of its infrastructure (Services et outils, n.d.). SIGNALER uses ISIDORE, a French-speaking harvester, that collects metadata and records then translates it into three languages ( English, French, and Spanish) to help the data discoverability (Services et outils, n.d.). Exposer is the interface used for depositing and managing data (Services et outils, n.d.). The Stocker provides a secure professional space to store and organize the digital data, which is operated by Huma-Num's file server (NAS) and the CC-IN2P3/CNRS for the iRods service (Services et outils, n.d.). The SHAREDOCS package provides access to files through the web and WebDAV clients but requires the researcher to have advanced computer skills to use this service (Services et outils, n.d.). The Huma-Num Box package provides access through all types of computers, and only requires average computer skills to use this service (Services et outils, n.d.). The Diffuser is used to disseminate the data on the internet by promoting free access to the data and the metadata (Services et outils, n.d.). The framework controls the reuse of data and metadata by insisting on the use of Creative Commons and follows the policy set by Etalab, a French government agency for open and shared public data) for all items stored in the repository (Mentions légales, *n. d)*. The Traiter can be used at different stages of a digital project to provide analytical information (Services et outils, n.d.).

**Data Citation**

The current citation needs to stay the same, including the same DOI. I do recommend the data set not be used or cited until the privacy issue of the subjects are protected. I am providing an updated citation for the data set because of the work I created for the new metadata.

Kar, Arpan; Grover, Purva (2018): Data for: “technology enabled health” – insights from Twitter Analytics with a socio-technical perspective. Improvements to the Mirror of Mendeley Data. Dataset. https://doi.org/10.17632/rs3c243fnm.1

**Copyright/licensing statement**

The data set has a Creative Commons Attribution 4.0 International (CC By 4.0) attached with it.

**Human Subject Consideration**

I have issues with how little care was taken into consideration when publishing the data set. The data set contains the Twitter users’ names that lead to profiles. I understand some of the profiles are made up, but most of them reveal people’s identity. I am not happy the date set provides information that can be used to track down individuals who use Twitter; it is dangerous. There are no steps taken to keep anyone’s identity anonymous or private. This is another major issue with the data set because of the lack of ethical concerns.

**Preservation**

I do not think the data set’s format of CSV is going to be obsolete any time soon because of how many spreadsheets can open this file type. The problem for long-term storage is why save this data if no one can use it because it does not have a time frame on when the data was collected. I am looking at the space the data set is taking up at the repository that could be used for another data set that is usable to other researchers. The data set provides interesting information on social media, but it cannot be used in other studies or replicated. The other major issue is not protecting anyone’s identity in the data set.

**Reference**

*À propos*. (n.d.). Nakala. https://www.nakala.fr/about.

*API*. (n.d.). Nakala. https://api.nakala.fr/doc.

Bilkent University Institutional Repository | re3data.org. (n.d.). Re3data.Org. Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100013352

CaltechDATA | re3data.org. (n.d.). Re3data.Org. Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100012384

Danish National Birth Cohort | re3data.org. (n.d.). Re3data.Org. Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100011824

*Data Observation Network for Earth*. DataONE. (2021, February 18). https://www.dataone.org/.

Grover, P., Kar, A. K., & Davies, G. (2018, July 25). *"Technology enabled Health" – Insights from twitter analytics with a socio-technical perspective*. International Journal of Information Management. https://www.sciencedirect.com/science/article/pii/S0268401218304997.

HumanID. (n.d.). https://humanid.huma-num.fr/oauth2/authorize?state=f5764ed97f548ca3a551f2797f9cd4f5&scope=openid+profile+email+orcid&response\_type=code&approval\_prompt=auto&redirect\_uri=https%3A%2F%2Fwww.nakala.fr%2Fhumanum%2Flemon%2Ftoken&client\_id=nakala\_huma-num.

Huma-Num. (2018, October 9). *La base de connaissance des services d'Huma-Num - How does one push one's data in ISIDORE with metadata and the OAI-PMH protocol?* Num. https://documentation.huma-num.fr/content/12/196/en/how-does-one-push-one%E2%80%99s-data-in-isidore-with-metadata-and-the-oai\_pmh-protocol.html?highlight=Dublin+Core.

Huma-Num TGIR. (2021, January 12). *Exposer ses données avec Nakala* | Huma-Num : l’infrastructure des humanités numériques. Https://Www.Huma-Num.Fr/. https://www.huma-num.fr/services-et-outils/exposer

ISIDORE. (n.d.). https://isidore.science/.

ioChem-BD | re3data.org. (n.d.). *Re3data.Org.* Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100012553

Kar, Arpan; Grover, Purva (2018). *Data for: “technology enabled health” – Insights from Twitter Analytics with a socio-technical perspective.* Mirror of Mendeley Data. Dataset. https://doi.org/10.17632/rs3c243fnm.1

KNB Data Repository | re3data.org. (n.d.). *Re3data.Org.* Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100010092

*Mentions légales*. (n.d.). Nakala. https://www.nakala.fr/credit.

NAKALA | re3data.org. (n.d.). *Re3data.Org.* Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100012102

*OAI 2.0 Nakala Request Results*. (n.d.). https://api.nakala.fr/oai2?verb=Identify. https://api.nakala.fr/oai2?verb=Identify.

Olafson, K. (2020, September 30). *How to Use Twitter Analytics: The Complete Guide for Marketers*. Social Media Marketing & Management Dashboard. https://blog.hootsuite.com/twitter-analytics-guide/.

Re3data.org. (n.d.). *Re3data*. Retrieved February 12, 2021, from https://www.re3data.org/

Research Data Australia. (n.d.) https://researchdata.edu.au/.

Research Data Unipd | re3data.org. (n.d.). *Re3data.Org.* Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100012955

Ressources | Huma-Num : *l’infrastructure des humanités numériques.* (2015). TGIR Huma-Num. https://www.huma-num.fr/ressources

*Services et outils*. (n.d.). Accueil. https://www.huma-num.fr/services-et-outils.

TEDxTalks. (2018, August 30). *The power in effective data storytelling | Malavica Sridhar | TEDxUIUC*. YouTube. https://www.youtube.com/watch?v=0e52QfQngrM.

TGIR Huma-Num. (2015). *Services et outils | Huma-Num : l’infrastructure des humanités numériques*. Huma-Num. https://www.huma-num.fr/services-et-outils

TROLLing | re3data.org. (n.d.). *Re3data.Org.* Retrieved February 12, 2021, from https://www.re3data.org/repository/r3d100011623

World Data | Predictability Models Powered by. (n.d.). *WorldData.AI.* Retrieved February 12, 2021, from https://worlddata.ai/

WorldData.AI. (2020). *How has WorldData.AI revolutionized open data? - WorldData.AI. Medium.* https://medium.com/@worlddata/how-has-worlddata-ai-revolutionized-open-data-cca2c9ad6f53

WorldData.AI | re3data.org. (n.d.). *Re3data.Org.* https://www.re3data.org/repository/r3d100013323