Luming Zhang

Professor Matthew S. Mayernik

LIS 545 B Winter 2024: Data Curation

Feb 20th, 2024

**Term Project- Repository Profile**

For this term project, I have chosen the GlobeData repository[[1]](#footnote-1), a decision driven primarily by my research interest in area studies. GlobeData, hosted by the Leipzig Research Centre Global Dynamics (ReCentGlobe), presents an intriguing opportunity to delve into a rich array of resources and materials relevant to my field. Furthermore, GlobeData’s openness to researchers worldwide is significant. This global inclusivity indicates a wealth of varied perspectives and methodologies, enriching the repository with a diverse range of data. Additionally, GlobeData’s systematic approach to data collection assures a level of comprehensiveness and reliability in the data available, which is crucial for thorough and robust research.

As I navigated through the intricacies of selecting an appropriate repository for my data curation project, I found myself delving into the details of GlobeData’s policies, procedures, and data access mechanism. My exploration began at GlobeData’s official website (<https://globedata.uni-leipzig.de/>), which served as a primary source for the information I gathered.

**Openness of Data Submission:**

GlobeData, a repository under the Leipzig Research Centre Global Dynamics (ReCentGlobe), is open to data submissions from researchers around the world. This aspect resonated with my aim to contribute to and benefit from a global pool of interdisciplinary research, particularly in area studies.

**Criteria for Data Acceptance:**

While reviewing GlobeData’s content types and their open approach to data submission, I was impressed by the multidisciplinary scope of the repository. However, I noted a lack of detailed information on the exact requirements for the Submission Information Package (SIP). This observation led me to ponder whether the vagueness is intentional, aiming to cast a wide net by not deterring potential contributors with stringent requirements. Such an approach could indeed be strategic, designed to encourage a broad spectrum of submissions by maintaining flexibility in submission criteria. This consideration reflects GlobeData's commitment to fostering a diverse and inclusive research environment, though it also highlights the importance of clear communication in repository guidelines to ensure potential contributors understand the submission process.

**Guidance for Submitters:**

Although GlobeData’s website provided an overview of the types of data accepted and the general ethos of FAIR principles, I noticed a lack of detailed information on the exact requirements for the Submission Information Package (SIP). This led me to consider reaching out to the provided contacts (fdm\_recentglobe@uni-leipzig.de and eva.ommert@uni-leipzig.de) for more specific guidance.

**Assistance and Metadata Standards:**

One of the aspects I appreciated about GlobeData is its provision of human support for data submitters. This support is pivotal for addressing specific queries or concerns regarding submission. Additionally, the repository’s adherence to established metadata standards like DataCite Metadata Schema and Dublin Core ensures that submitted data is well-documented and easily discoverable, a factor that adds to the reliability and accessibility of the repository.

**Login and Account Management:**

The user guide of GlobeData is extensive, covering aspects like account information, login options, account creation and editing, password reset, remote authentication, API token management, and more. This guide serves as a valuable resource for new users to navigate the repository’s interface effectively. However, when I accessed the login page to download data, I noticed a limitation: the absence of an institutional login option. This was contrary to what I expected, given the repository’s extensive user guide. As a result, I had to opt for an individual account setup. While this was not a significant hindrance, it did differ from my initial expectation of a more streamlined institutional access. The individual account setup process was straightforward, but I couldn’t help but think that an institutional login could have expedited the process, especially for users affiliated with academic institutions.

**Data Access Mechanisms:**

The repository provides several ways to access data. Users can perform basic and advanced searches, browse the dataverse installation, and even utilize APIs for more advanced data interactions. The availability of API tokens and guidance on their creation and use, particularly the SWORD and REST APIs, is a notable feature, allowing for automated and programmatically controlled data access. This multifaceted approach to data access caters to both casual browsers and more technically inclined users.

**Metadata Display Standards:**

GlobeData adheres to several recognized metadata standards, ensuring that data is accurately and comprehensively described. These standards include the DataCite Metadata Schema, OAI-ORE, Dublin Core, and DDI (Data Documentation Initiative). The inclusion of DDI, in particular, is noteworthy. The Data Documentation Initiative standard is crucial for the documentation of social, behavioral, economic, and health data, among others. By providing metadata exports in formats aligned with these standards, GlobeData enhances the discoverability and usability of datasets, facilitating easier access and understanding for researchers.

**Dissemination Information Package (DIP):**

While examining a specific dataset titled “Index of Key Documents of the Organisation of African Unity (OAU) and the African Union (AU), 1963–2023,”[[2]](#footnote-2) I gained insight into the elements of GlobeData’s DIP. This dataset, for instance, included detailed metadata like the persistent identifier, publication date, author information, and subject categorization. Additionally, there were clear terms of use and access restrictions, with the option to contact the author for access to restricted files. This level of detail in the DIP ensures that users are well-informed about the data they are accessing, including its usage rights and limitations.

**Conclusion and Personal Reflection**:

In reflecting on my exploration of the GlobeData repository, the realization that it utilizes the Dataverse software for its repository infrastructure adds an important dimension to my assessment. Dataverse is a widely recognized software suite for data repositories, known for its comprehensive feature set that supports data curation, sharing, and preservation. This understanding reinforces my view of GlobeData’s robustness and the reliability of its data access mechanisms. The repository’s use of Dataverse suggests a solid foundation for managing and disseminating research data, further assuring users of the quality and accessibility of its datasets. Recognizing the significance of Dataverse in the context of GlobeData has deepened my appreciation for the complexities of data management and the critical role that advanced repository platforms play in the broader landscape of research and information management. (1022 words in total)

References

*Account creation + management*. Dataverse.org. (n.d.). <https://guides.dataverse.org/en/5.12.1/user/account.html>

Engel, U. (2023, June 6). *Index of key documents of the Organisation of African Unity (OAU) and the African Union (AU), 1963–2023*. GlobeData. <https://globedata.uni-leipzig.de/dataset.xhtml?persistentId=doi%3A10.48736%2FGD1PZPUVT>

*GlobeData*. Brand. (n.d.). <https://www.re3data.org/repository/r3d100014174>

GlobeData. (n.d.). <https://globedata.uni-leipzig.de/>

1. The link to GlobeData repository’s introduction page on re3data is <https://www.re3data.org/repository/r3d100014174>. The GlobeData’s own link is: <https://globedata.uni-leipzig.de> [↑](#footnote-ref-1)
2. The link to this dataset in GlobeData is <https://globedata.unileipzig.de/dataset.xhtml?persistentId=doi:10.48736/GD1PZPUVT> [↑](#footnote-ref-2)