



Report on Wikidata in Africa and the Middle East

Houcemeddine Turki, Mohammed Sadat Abdulai, Lane Rasberry, Mohammed Kamal-Deen Fuseini, Mus'ab Banat, Ali Khalil, René Fabrice Bile, Yamen Bousrih, Nassima Chahboun, Nanour Garabedian, Ruby D-Brown

Introduction

Wikidata, a collaborative knowledge base and data repository, has emerged as a crucial resource for the digital representation of diverse cultures and histories. It serves as a central hub for structured data, supporting a wide range of Wikimedia projects and enabling better accessibility and usability of information. In regions like Africa and the Middle East, Wikidata plays an increasingly vital role in documenting and preserving cultural heritage, supporting academic research, and fostering community-driven knowledge sharing. This report explores the development, challenges, and opportunities associated with Wikidata in Africa and the Middle East. We examine the current state of content representation, community involvement, and the use of Wikidata in these regions. Additionally, the report discusses strategies to enhance the visibility and accuracy of information about these culturally rich and diverse areas, ensuring that local knowledge is adequately represented on the global stage. By highlighting successful initiatives and identifying areas for improvement, this report aims to provide a comprehensive overview of the impact and potential of Wikidata in Africa and the Middle East.

The current state of Wikidata in the Middle East and Africa (MEA) reflects a complex interplay of rich cultural diversity and significant challenges. This region, with a population of 1.6 billion across an area of 30.9 million km², encompasses numerous civilizations, local languages, and distinctive features. The three main languages—Arabic, French, and English—highlight the linguistic diversity that characterizes the region. Organized under two Wikimedia regions, MENA and Wiki Indaba, the region faces a substantial gap in readership compared to other regions such as Latin America and Europe. This gap is more pronounced for Wikidata than for the English Wikipedia, pointing to limited engagement and a lack of awareness about the platform's potential. Factors contributing to this include limited local knowledge and a low level of readership, underscoring the need for greater outreach and education to fully leverage Wikidata's capabilities in preserving and promoting the region's cultural and historical data.

Building Wikidata communities in the Region

Building robust Wikidata communities in Africa and the Middle East involves fostering collaboration and leveraging local expertise across different linguistic and technical groups. In the Arabic-speaking world, the Wikidata Arabic Community, chaired by Yamen Bousrih and established during Wikimania 2023, plays a key role in promoting the use of Wikidata. This community focuses on creating and sharing content relevant to Arabic-speaking users, ensuring that the region's data is accurately represented and accessible.

In French-speaking African countries, the Wiki Wake Up Afrique initiative, led by Mermoze Adodo and others experienced Wikimedians, aims to engage and educate users on Wikidata's potential, particularly in preserving and promoting local knowledge. Meanwhile, the AfLIA Wikidata Project, chaired by Alice Kibombo, focuses on English-speaking regions, working to build a community of librarians and information professionals to contribute to and utilize Wikidata.

The Africa Wikimedia Technical Community, led by Joris Darlington, has historically focused on Wikipedia and MediaWiki but is now increasingly interested in Wikidata's development. This community is instrumental in addressing the technical aspects of Wikidata and facilitating training for new contributors. To support these efforts, comprehensive tutorials have been developed in Arabic and French, providing essential guidance for newcomers.

One of the critical goals across these communities is the decolonization of knowledge. This goes in line with previous discussions held in this context such as the Whose Knowledge User Group ones. This involves adding new data relevant to local contexts and updating existing data models and software to better reflect the diverse histories and cultures of the region. For instance, the work of User:Hubaishan in Yemen highlights the importance of including underrepresented perspectives and ensuring that the digital knowledge landscape is inclusive and comprehensive. These initiatives collectively contribute to building a more vibrant and diverse Wikidata community in Africa and the Middle East, empowering local voices and knowledge.

Wikimedia Deutschland to the Rescue

Wikimedia Deutschland has been instrumental in supporting the growth of the Wikidata community in Africa and the Middle East by providing resources, organizing events, and developing tools. They offer valuable resources, such as the Wiki Mentor Africa initiative, which connects experienced Wikidata contributors with newcomers in Africa, providing mentorship and guidance. Additionally, the Hardware Donation Program supplies essential hardware to community members, facilitating access to technology. WMDE collaborates with smaller groups to develop software, such as the new Special:Lexeme page with the community in Indonesia. It also provides learning resources, such as the MOOC, for those who want a more structured way of learning about Wikidata and becoming part of the community.

Wikimedia Deutschland organizes various events to foster community engagement and learning. These include the XXX Days, which are thematic workshops focusing on specific aspects of Wikidata, and the Wikimania Tandem Scholarships, which support community

members from underrepresented regions to attend the annual Wikimania conference. Another significant event is WikidataCon, a global gathering of Wikidata enthusiasts and experts, where participants can share knowledge, experiences, and ideas.

In terms of tools, Wikimedia Deutschland has developed several key tools to enhance the functionality and accessibility of Wikidata. The Query Builder tool simplifies the process of constructing SPARQL queries, making it easier for users to extract specific data from Wikidata. The REST API provides developers with a straightforward way to interact with Wikidata, enabling the creation of custom applications and services that utilize Wikidata's vast dataset. These tools and initiatives collectively contribute to building a more robust and inclusive Wikidata community in Africa and the Middle East.

Ways to contribute

Wikidata Research and Development

While Wikidata does provide extensive user documentation, the availability of its user community and their willingness to answer questions is a major attractive feature. Any researcher who is either new to Wikidata, or who is considering how they can collaborate with Wikidata, should post to a Wikidata discussion forum and start conversation with experienced Wikidata editors.

While Wikidata's design allows researchers to host many sorts of projects, some sorts of projects have a precedent of success, others have a history of being more challenging, and researchers have hardly attempted others.

The most common sort of Wikidata research activity is sharing research metadata, which has benefits including bringing visibility to an existing published research project. Another common type of project is sharing enough data to support a human in telling a story.

Projects which researchers consider, but which are not usually a good fit for Wikidata, include using Wikidata as a data repository or publishing data without plans for a use case such as storytelling.

The Wikidata community may give researchers advice which is impractical. Common Wikidata editor requests include that researchers join the Wikidata community, engage in Wikimedia projects beyond Wikidata, or that they start their Wikidata project with plans to curate Wikidata content outside the scope of the main research. While these activities are helpful to Wikidata and facilitate the research relationship, setting expectations greatly helps collaborations.

Wikidata for Education

The Covid-19 pandemic disrupted education for 1.6 billion students, highlighting the need for accessible educational data. Wikidata for Education (WD4E) aims to enhance global access to school curricula data and align open educational resources (OER) with national curricula.

WD4E tackles challenges like missing Wikidata items related to curricula data, related properties, a data model or structure, and diverse document formats by structuring curriculum data, improving OER quality, supporting Wikimedia volunteers, creating a curriculum-aligned education portal, enabling global curriculum analysis, and facilitating machine-readable data for innovative solutions. Using Ghana's curriculum as a key example, WD4E addresses knowledge gaps on Wikipedia in multiple languages. The integration of Ghanaian and Uruguayan curricula into Wikidata highlights the project's impact. Currently, curriculum data for multiple subjects and levels in Ghana are uploaded, with a final report pending. Future plans include expanding to more countries, adding more curricula to Wikidata, developing user-friendly applications, educating stakeholders on the interface, and continually improving the curriculum model. WD4E aims to democratize educational resources, making them more accessible and adaptable worldwide.

Wikidata in specific countries

Arab region

Country	Status
Tunisia	<p>First contributions from Tunisia to Wikidata date back to 2015, when Helmi Hamdi (Helmoony), a veteran Wikipedian, became a Wikidata ambassador in the Middle East and North Africa. He worked to establish a community of Arabic Wikipedia users who are familiar with using Wikidata. Additionally, his efforts aimed to address issues in the Arabic Wikipedia community, such as the representation and support of Arabic dialects, which often lead to controversies due to political considerations.</p> <p>In 2019, the first Wikidata Workshop in Sfax, Tunisia, was organized by Houcemeddine Turki (Csisc), Mohamed Ali Hadj Taieb, and Mohamed Ben Aouicha to establish a research community focused on Wikidata. This initiative led to the creation of a Wikimedia-related research unit called Data Engineering and Semantics. The community primarily worked on developing medical and GLAM (Galleries, Libraries, Archives, and Museums) applications using Wikidata and applied machine learning and semantic technologies to enhance and curate medical knowledge on the platform. In 2022, this effort earned the community funding from the Wikimedia Research Fund, making it the first Arab and African research organization to receive this support. Beyond this, the community worked with Wiki Wake Up Afrique (French-speaking) and Wikidata Arabic Community (Arabic-speaking) to establish learning resources for the Middle East and Africa region to bridge the gap between this region and other developed ones.</p> <p>In 2020, the Wiki World Heritage User Group was founded by Yamen Bousrih and other Wikimedia enthusiasts with the aim of enhancing the coverage of cultural heritage globally, with a particular focus on the Middle East and Africa. The group sought to promote the use of Wikidata in documenting and preserving cultural heritage. To achieve this, they</p>

	<p>organized a series of sessions in English to introduce users to tools and techniques for contributing to and utilizing Wikidata. These sessions covered the basics of the Wikidata Query Service and provided training on using web programming languages such as HTML, CSS, and JavaScript to develop tools that support cultural heritage initiatives. The group developed several tools that leverage the capabilities of the Wikidata Query Service, enabling more effective data retrieval and visualization related to cultural heritage sites and objects. This initiative has helped to increase the visibility and accessibility of cultural heritage information on Wikidata, fostering greater global awareness and appreciation.</p>
Iraq	<p>In Iraq, Wiki projects have not yet reached their full potential. However, in the past two years, efforts have been made to rejuvenate these projects, particularly in Baghdad. In 2022, the Iraqi Users Group held the country's first Wikidata workshop, which saw the participation of 20 individuals. The workshop revealed a significant demand for learning Wikidata tools and their applications. In response, Wikimedia Iraq is now developing comprehensive training materials to be made available to interested individuals by the end of 2024.</p>
Other countries	<p>Libya: As a merged community the activities are focused for now on Wikipedia and a little more on Commons. We have beginners workshops on our current annual program for Wikidata which will be addressed to data science students aiming to raise awareness about Wikidata and gain more contributors on Arabic content.</p> <p>Gender gap on Wikidata: As many other projects, there is a huge gap on Wikidata female contributors in our community. There are individual efforts but not enough. We hope to plan a clear road map regarding this gap in the Arabic Wikidata Community.</p> <p>Jordan: Wikidata awareness in Jordan is still forming. The use of Wikidata was mainly in linking articles with other languages (interwiki) or adding descriptions to Wikidata items. In 2024, Jordanian Digital Arabic Content Association, an allied-organization to the movement, received a grant and scheduled a Wikidata workshop in September which will be first Wikidata in person Workshop in Jordan.</p> <p>Other countries:</p> <ul style="list-style-type: none"> ● User:Mr. Ibrahim from Yemen has made over 1,639,972 contributions to Wikidata. His bot did 40,441,348 contributions. ● User:FShbib from Saudi Arabia has made over 1,285,864 contributions to Wikidata. ● User: Hubaishan from Yemen has become the first interface administrator of Wikidata in the Arabic-speaking community and is contributing to the inclusion of local knowledge through technical contributions like adding support for the Hijri calendar. ● User:Mohammad Abul-Futooh from Egypt made 9,325,830 contributions to Wikidata. ● User:Meno25 from Egypt has made 668,596 contributions to Wikidata. ● User:Sky xe, from Palestine, made 3,025,356 contributions to Wikidata. ● Birzeit University from Palestine has a specific course and research

	community on Semantic Web. Professor Mustafa Jarrar (Mjarrar) and Dr. Qais Shreda (Qais.Shreda) made significant contributions to Wikidata in this context, particularly through collaboration with Wiki Palestine Community.
--	---

Sub-Saharan Africa

Country	Status
Western Africa	<p>In Igbo Wikimedia community, there is a program Wiki Mentor Africa which focuses on Tech skills. They have online Wikidata's sessions; they bring together French speaking people to allow them to receive the same level of knowledge. In 2024, they thought about physical approach; some countries like Congo benefited from this experience. The program is led by Udeh and there are some wikidata experienced individuals who joined to share their knowledge and skills.</p> <p>In Cameroon, Wikidata activities are mainly conducted through campaigns and occasionally community projects. There is a user group and unaffiliated organizations (workgroups) working on these initiatives, including AGRIPO and WAFTAI. WAFTAI, in particular, focuses on helping young girls and women master digital techniques and technologies. Throughout the year, they organize activities related to Wikidata and other projects. WAFTAI initiated "Wednesday with Wiki" in Cameroon, providing training to contributors in both Cameroon and Madagascar. Following this activity, they were able to train two assistant trainers who later assisted in the Wiki For Human Rights 2024 campaign in Cameroon.</p> <p>The Wiki Wake Up Afrique initiative aimed to establish a strong and active French-speaking Wikidata community in Western Africa. This initiative focused on engaging and empowering French-speaking individuals and groups across the region to participate in Wikidata projects. The goal was to increase the representation of African knowledge, culture, and heritage in French on the Wikidata platform, thereby addressing the underrepresentation of African content and perspectives in global knowledge bases. Through a series of workshops, training sessions, and community-building activities, Wiki Wake Up Afrique sought to equip participants with the necessary skills to contribute to Wikidata. This included teaching them how to create and edit entries, manage data, and understand the intricacies of the platform.</p> <p>In Ghana, the Dagbani Wikimedians User Group (DWUG) has been a great resource organizing Wikidata training for their community members and other affiliates and organized communities in Ghana. The DWUG has also been adding Dagbani lexemes to Wikidata as one of the flagship languages for the Wikidata Lexicographical data. Gurene Wikimedia Community started the Wikidata loves Gurene names where they documented the indigenous names in the Gurene language on Wikidata which has now been adapted by other communities. The Global Open Initiative has also started a project on democratizing Access to Supreme Court Cases through Wikidata where</p>

	cases are documented on Wikidata to allow for easy access.
Other countries	The African Library and Information Associations and Institutions (AfLIA) made an effort to establish a vibrant Wikidata community across Eastern and Southern Africa through the launch of its Wikidata Online Course, with the first cohort starting in 2023. This initiative was particularly aimed at engaging professionals from the Library and Information Science (LIS) communities in these regions. The course sought to introduce participants to the fundamentals of Wikidata, enhance their skills in data management and curation, and encourage the use of Wikidata as a tool for improving access to information. By focusing on the LIS sector, AfLIA aimed to leverage the expertise and reach of librarians and information professionals to enrich Wikidata with locally relevant data, thus promoting a more diverse and representative global knowledge base. The initiative also sought to foster a network of trained individuals who could continue to contribute to and expand the Wikidata community in their respective countries.

Challenges

Language Support

There is a need for Wikidata content to be available in the diverse languages and dialects of these regions to ensure accessibility for all communities. Although Wikidata is multilingual, it faces several challenges in supporting multiple languages effectively. The newly created “mul” labels only support Latin script, which can cause issues for languages using other scripts, particularly Arabic, leading to the sharing of labels among several items. There is also a challenge in translating and maintaining accurate and comprehensive data due to limited digital resources in many local languages. Additionally, issues with the accuracy and appropriateness of automatic translations, managing variations in dialect within the content, and the task of documenting and preserving endangered languages present further complications. Technical challenges in supporting diverse languages, especially those with complex scripts, add to the difficulty. Moreover, ensuring that content is appropriate and respectful of local cultures and languages is essential for cultural and linguistic sensitivity.

Geopolitical Factors

Limited internet penetration due to infrastructure issues, economic barriers, and political restrictions limits access, creating a digital divide with disparities in digital literacy and access to technology that hinder engagement with platforms like Wikidata. Censorship and governmental control, including regulations or restrictions on internet usage and content, affect accessibility and the spread of information. Economic challenges, such as poverty and economic disparities, restrict access to resources like computers and the internet, with basic needs taking precedence over participation in digital platforms. Lower literacy rates and a lack of digital literacy further hinder engagement and contribution. Restrictions on freedom of expression impact contributions to Wikidata, while ensuring accessibility for people with disabilities remains a challenge. Volunteers in conflict or unstable areas face physical and

mental health risks, and geopolitical tensions, along with cross-border tensions, affect collaboration and information flow. Conflict zones present challenges in accessing reliable information and contributing, while economic sanctions impact the availability of technology and resources for participation. Navigating bureaucracy and engaging with local governments add complexities, especially in rural areas with limited internet speeds and power sources. Economic crises affect community participation, and political repression poses challenges in politically repressive regimes. Lastly, digital initiatives often compete with other community needs, leading to competing priorities.

Community Factors

Protecting the anonymity and safety of contributors is crucial, along with ensuring inclusive and representative local Wikimedia community leadership. Addressing apathy or disinterest in contributing to global platforms is necessary to increase participation. Bridging the technical skills gap in data science and programming is essential, while volunteer fatigue and burnout need to be managed to maintain a motivated contributor base. Recognizing and rewarding volunteer efforts is important for sustaining engagement, but high turnover rates can lead to gaps in knowledge continuity and the need for continuous training. Limited awareness about Wikidata and its benefits in the region necessitates promotion and education efforts. Building a local community of contributors and editors is vital for maintaining data accuracy and relevance, though partnerships and collaborations with local organizations and institutions can be challenging. Ensuring the credibility and accuracy of contributors is crucial, as is securing sustainable long-term funding for Wikidata projects. Handling sensitive or controversial topics with cultural sensitivity is necessary to avoid misunderstandings or conflicts. The scarcity of information about local events, figures, and cultures affects Wikidata's usefulness, and ensuring accurate information in less-documented areas is challenging.

Resource constraints, including funding and support limitations for technology infrastructure and training, are significant. Local infrastructure limitations, such as unreliable electricity and internet, need to be addressed, and Wikidata must adapt to specific technological and cultural contexts. Ensuring data privacy and protection in compliance with local laws and respecting users' privacy is essential. There is an imbalance in the representation of indigenous and local knowledge systems, and concerns about misrepresentation or cultural appropriation need attention. Skepticism towards digital platforms due to past experiences with misinformation or lack of privacy protections poses challenges, as does ensuring the sustainability of local initiatives. Engaging local experts and institutions is necessary for content accuracy and relevance, but the lack of training resources in local languages or tailored to regional needs hinders progress. Developing technical skills among potential contributors, navigating copyright restrictions, and legal barriers to data sharing are also significant challenges.

Addressing the underrepresentation of marginalized groups and gender disparities in content creation and representation is critical. Implementing appropriate protocols for local content, building trust with NGOs, and scaling up training and outreach programs with limited resources are necessary steps. Tailoring training materials to different linguistic and cultural contexts and balancing openness with responsibility to prevent misuse or misinterpretation of open data are important. Effective allocation of funds, complexities in cross-sector collaboration, and ensuring long-term partnership sustainability are ongoing concerns. Additionally, complying with international laws and understanding local legal frameworks

impacting information sharing are essential for successful operations. Incorporating Wikidata into educational curricula poses challenges, as does the need for training and resources to help educators use Wikidata effectively.

Technical Challenges

Using Wikidata in the Middle East and Africa presents several technical challenges. Establishing consistent data standards across diverse languages and regions is crucial to ensure uniformity and reliability. Server access and maintenance can be problematic in areas with unstable internet connections, impacting the platform's functionality. Scalability issues arise as the platform must manage increasing data and traffic from a growing number of diverse contributions. Localizing user interfaces to accommodate different languages and cultural contexts is necessary for effective user engagement. Mobile accessibility is vital, as many users in these regions rely on mobile devices for internet access. Aligning data semantics across various languages and cultural contexts presents additional complexity. The platform must also keep up with rapid technological advances and adapt to mobile-first environments, ensuring that it remains user-friendly on mobile devices. Managing large volumes of data while maintaining quality and relevance is an ongoing challenge. Additionally, ensuring data privacy and handling sensitive information ethically are crucial for maintaining user trust and compliance with local standards.

Cultural and Data-Related Challenges

Using Wikidata involves addressing several complex challenges related to data ownership and sovereignty, particularly concerning the local knowledge shared on a global platform. Ensuring that translated content remains culturally relevant and accurate is essential, as is managing vandalism and misinformation to maintain high content quality. Balancing neutrality and bias, especially in sensitive areas, is crucial to ensure unbiased representation. There are significant data gaps and coverage issues that need to be addressed, as well as a need to balance global and local content representation. Verifying and sourcing accurate local information can be challenging, and customization of Wikidata tools and interfaces for specific community needs is necessary. Maintaining data integrity with multiple sources and contributors, while navigating cultural differences, requires sensitivity and respect for diverse cultural contexts. Concerns about digital colonialism and equitable data ownership highlight the need to address dominant cultural narratives and ensure fairness. Revisionist histories and colonial legacies can impact historical representations, and internal community conflicts can affect collaboration. Epistemological differences and challenges in documenting informal knowledge systems must be managed carefully. It is important to balance global standards with local practices to avoid marginalizing local knowledge. For example, the focus on Gregorian and Julian calendars, as opposed to indigenous calendars like the Hijri Calendar, illustrates how marginalized data models can be overlooked. Despite efforts such as User:Hubaishan's source code to support the Hijri Calendar, this task has faced delays, highlighting the need for a more inclusive approach.

Further Reading

- Wikimedia Regions: https://meta.wikimedia.org/wiki/Wikimedia_regions
- Wikidata Arabic Community: https://www.wikidata.org/wiki/Wikidata:Wikidata_Arabic_Community
- Wiki Wake Up Afrique: https://meta.wikimedia.org/wiki/Wiki_Wake_Up_Afrique
- AfLIA Wikidata Project: https://meta.wikimedia.org/wiki/Aflia_Wikidata_Project
- WAFTAI page: https://meta.wikimedia.org/wiki/Women_Association_For_Technology_And_Innovation
- Africa Wikimedia Technical Community: https://www.mediawiki.org/wiki/Africa_Wikimedia_Technical_Community
- Arabization of Wikidata Leveling Up Days 2024 and other important videos: <https://w.wiki/AX3J>
- French Tutorial of Wikidata by Wiki Wake Up Afrique: <https://w.wiki/5pjc>
- WikiProject Africa: <https://w.wiki/Ahfh>
- Updating datetime to cover other calendars including Hijri: <https://phabricator.wikimedia.org/T252627>
- Wiki Mentor Africa: <https://w.wiki/Ahvk>
- Hardware Donation Program: <https://w.wiki/6i3C>
- Software Collaboration for Wikidata: <https://w.wiki/Ajfi>
- Wikidata MOOC: <https://w.wiki/9kfs>
- XXX Days: <https://w.wiki/Ajg2>
- Wikimania Tandem Scholarships: <https://w.wiki/Ajg9>
- WikidataCon: <https://w.wiki/AjgN>
- Query Builder: <https://w.wiki/6cnU>
- REST API: <https://w.wiki/7sfj>
- Data Engineering and Semantics Research Unit: <https://deslab.org>
- Wikimedia Research Fund Project: https://meta.wikimedia.org/wiki/Research:Adapting_Wikidata_to_support_clinical_practice_using_Data_Science,_Semantic_Web_and_Machine_Learning
- Sawtpedia: <https://github.com/csisc/Sawtpedia>
- MuslimSites: <https://github.com/csisc/MuslimSites>
- Wiki World Heritage User Group Capacity Building: https://meta.wikimedia.org/wiki/Wiki_World_Heritage_User_Group/Capacity_building
- Decolonizing the Internet discussion: <https://whoseknowledge.org/initiatives/decolonizing-the-internet/>
- Default values for labels and aliases: https://m.wikidata.org/wiki/Help:Default_values_for_labels_and_aliases