

Wikimedia Movement Data Hub

Wikimedia metrics to understand and improve Diversity, Inclusion & Equity in the movement

Setting

Wikimedia Movement Ecosystem

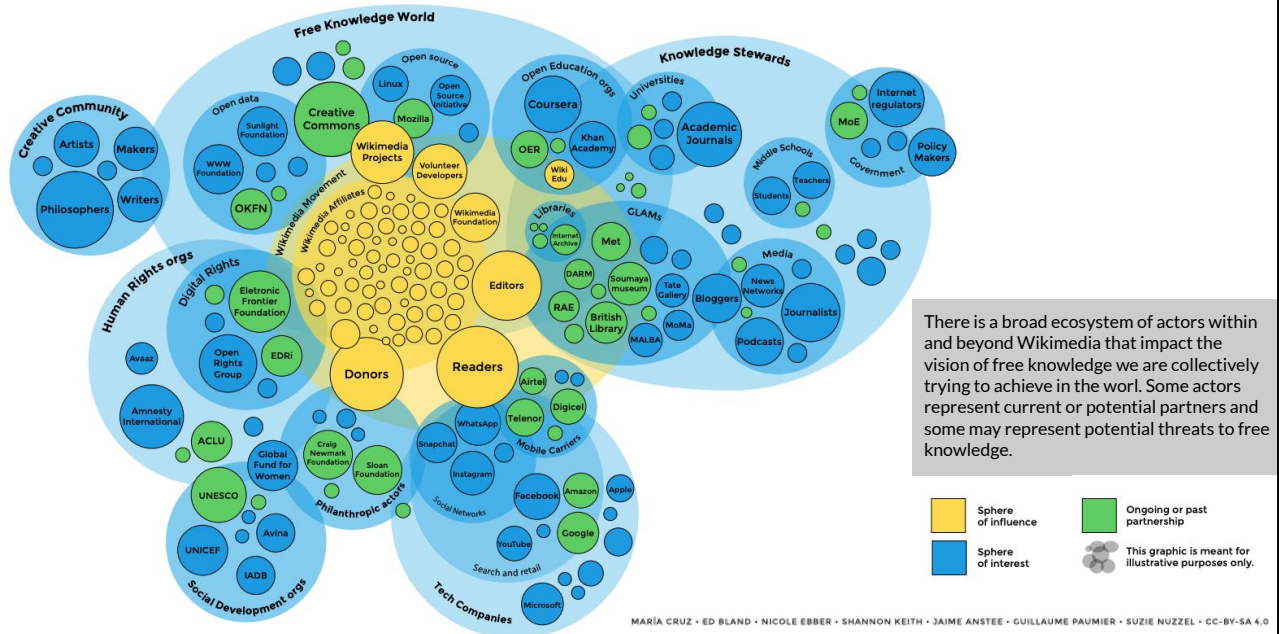
A diverse landscape of Wikimedia projects and movement actors

In 2017 the “Wikimedia Movement” got together and produced a shared strategic direction:

By 2030, Wikimedia will become the essential infrastructure of the ecosystem of free knowledge, and anyone who shares our vision will be able to join us.

Slide graciously borrowed from “What is Everything” by Kelsi Stine-Rowe

Wikimedia Movement Ecosystem and Actors



The [Movement](#) is a global, volunteer-driven network of individuals, chapters, affiliates and partners who work to advance the agenda of Open Knowledge. At the core of this agenda is the position that universal access to reliable and verifiable knowledge is a basic human right. And, that a consensus-based dialogue on what constitutes verifiable knowledge is the best method we currently have for collectively generating the knowledge all of us can depend upon. A central modality of the Movement is that dialogue is valued, fluid and ongoing across languages, cultures, geographies and interest groups. We seek consensus but nothing is ever settled.

Who is part of our movement?

How are people getting involved?

Where are our
fastest growing communities?

**Where are we showing promise in
diversity and inclusion; where are we
most challenged, and how?**

What supports movement collaboration;
how and where can we improve?

**Are we equitably supporting
emerging communities?**

Where are they?

Where are Wikimedia projects and
communities keeping up or falling behind
global infrastructure gains?

**What voices are missing from our
projects and communities?**

Are our growth targets equitable?

Many of our questions about our communities ask where; historically we have only been able to examine data by Wikimedia project or special data collection event. To understand the context of movement organizing, we must understand the where geographically as well.

Of course, thematic and language areas intersect these geospaces where our communities exist and grow - for this reason we are working to map languages and thematic groups to geo spaces this year also as we work to develop our initial dashboard.

Situation

Measuring our Movement Ecosystem

A need for a more diverse datascape

Our MTP and movement strategy bring a need to define what equity, diversity, and inclusion mean in data.

Diversity, inclusion, and equity are now central to the Foundation's MTP and movement strategy aims; in order for us to understand the political, social, and economic barriers contributors may face in their movement engagement we need to measure and track changes.

However, measuring inequality and diversity is not straightforward, it requires decisions to be made on the metrics and distributional characteristics of interest. These decisions influence the conclusions and must be closely linked to specific use cases for movement decision-making.

We have a lot of data, the world has a lot of data, let's compile what we have to begin answering basic questions.

Following the lead of the National Equity Atlas (2019) and Chicago Beyond (2019), we have begun to map our existing wealth of landscape data to develop an equity landscape data reference and index to track the Wikimedia movement's progress in breaking down the social, political, and technical barriers to full participation in free knowledge, based on what is already being captured.

To begin to do this, we have assembled our internal engagement and social research metrics alongside aligned global data indicators to enable easy examination of the Foundation's progress within national or regional contexts using available data.

We intend to view our data through the lens of social inequities already mapped in available global data in order to identify key progress levers to support healthy movement development without adding additional burden to our stakeholders who are always being asked to input.

As movement leaders, it is critical that we all take the time to pause to appreciate the evolving landscape of data and engage in a data-informed dialogue around equity across our movement organizing spaces to better calibrate both our tools and our practices. Travelling this road together will improve the validity of evaluative tools as well as the feedback stakeholders are provided (Krause, 2019; Magana, 2019; We All Count, 2019), it also can make data, its analysis, and reporting, more usefully relevant to a broader audience of stakeholders (Chicago Beyond, 2019).

The Big Idea: Provide easily accessible data on inequities within the world and our movement.

With increased data awareness of inequities, movement stakeholders can reinforce pathways for diversity, inclusion, and equity and target interventions that eliminate barriers in the Wikimedia Movement.

The Vision

Imagine a Wikimedia Foundation and movement enabled to identify disparities in outcomes across movement spaces; to track our shared progress at increasing diversity, inclusion, and equity; and to provide data and analyses by which we can direct opportunities for change and hold ourselves accountable to our 2030 movement strategy aims.

Who will use the dashboards and reports?

Foundation decision-makers, Movement organizers, and Grantees

How will it change how we work?

It will help Foundation and movement gatekeepers to use data-informed decision-making processes to direct outreach and development programs, grants & partnerships programs, as well as potential product innovations & community growth initiatives. Further, this data will foster accountability in measuring equity for the Foundation's OKRs.



How will the movement change, what impact will this have?

With increased data awareness of inequities, movement stakeholders will reinforce pathways for diversity, inclusion, and equity and target interventions that eliminate barriers in the Wikimedia Movement.

If the above happens successfully, we will see coefficients of inequity and diversity for the distribution of key resources improve to be more fair across the movement.

Image credit: [Vision by Eucalyptus from the Noun Project, CC BY-SA 3.0](#)

We envision tools and process supports (i.e., dashboards, reporting, and data use) to enable movement-wide conversation and development of a shared data framework for the social changes which will be targeted by our shared movement strategy.

We envision that, once the monitoring inputs are agreed, the data dashboards and reporting will be used movement-wide to guide investments and movement partnership opportunities.

Most importantly, we envision that identifying disparities in outcomes across movement spaces, tracking our shared progress at increasing diversity, inclusion, and equity, and providing data and analyses by which we can hold ourselves accountable to our movement strategy aims will enable us, as an organization and a movement, to prioritize efforts that achieve the impact we wish to see in 2030.

Who will use the dashboards?

Foundation decision-makers leading teams, products, or programs;
Movement organizers leading outreach and development efforts; and
Grantees seeking to identify and align to relevant movement outcome metrics.

How will it change how we work?

We envision an annual report and corresponding calls to action about notable disparities or improvement trends. The report and corresponding data will be available to be used foundation-wide to make data-informed decisions about program strategies and investments for social change via programs for outreach, grants, partnerships, product innovations, & community growth initiatives.

We also envision support to internal teams to optimize the tools for their specific and general use and ensure their ability to use the data to direct:

Outreach and development programs

Grants & partnerships programs

Product innovations & community growth initiatives

A public version will also be made available for all movement stakeholders to similarly so that movement-wide we can also identify disparities in outcomes across movement spaces, and track progress at increasing diversity, inclusion, and equity.

How will the movement change as a result?

We will track year over year progress using comparative analysis as well as via calculating the global annual inequity and diversity coefficients for each of the underlying data inputs.





It will be important to carefully coordinate with other analysts, researchers and development teams to ensure alignment and coordination (which is already happening) as well as to collaboratively identify and develop follow-up research project plans to dig deeper into key observations and qualitative contexts around identified differences.

Strategy

Mapping movement engagement

Operationalizing diversity, inclusion, and equity in data

Strategy

	How We Get There
	Timeline
	The Metrics: Wikimedia Presence & Growth
	The Mock-up & Pilot Learning & Practice Demo

We will gather and structure our data annually to track participation to understand engagement in terms of diversity & inclusion, and equity



To bring transparency to movement participation in terms of the different aspects of equity we will examine movement demographics, social climate, and experiences and we will triangulate data from across Wikimedia engagement data including:

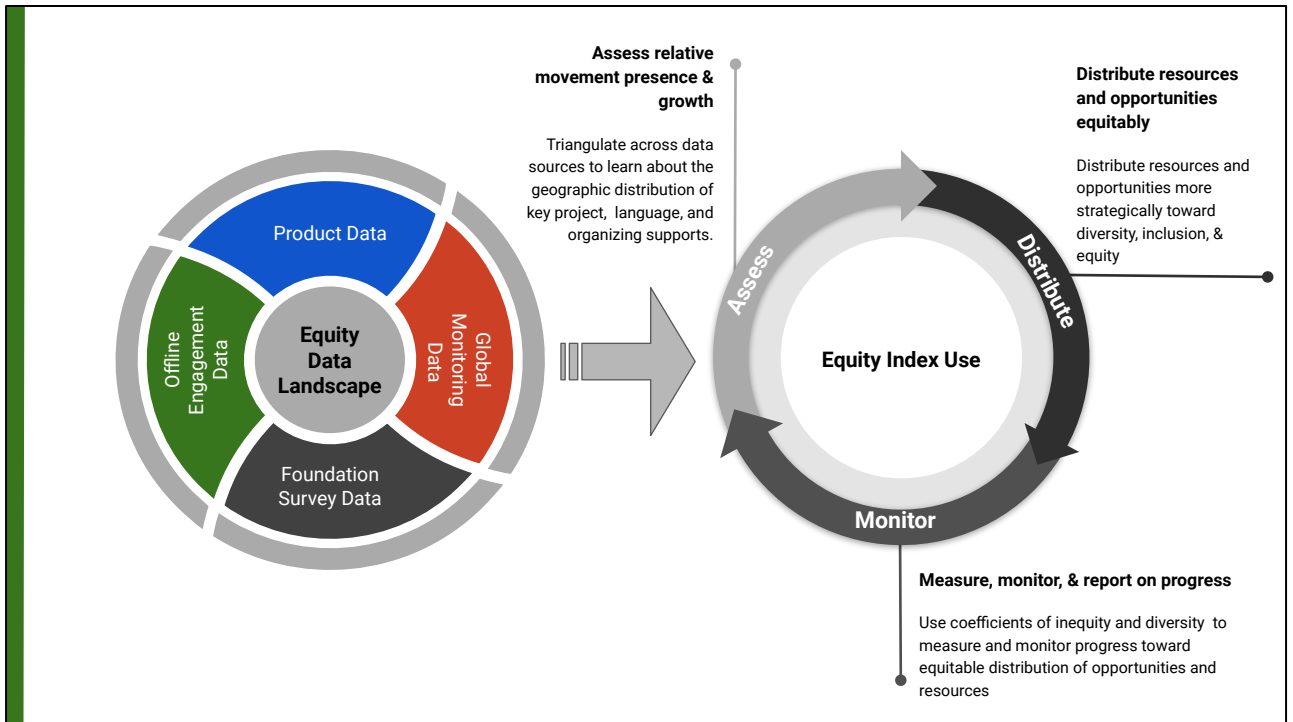
- Product Data (Wikimedia online engagement and content metrics)
- Foundation surveys
- Foundation offline engagement data

As we build each metric area we will build out our annual Global Data profile and work to report through three analytic lenses to expose each of three domains of Diversity, Equity, & Inclusion data to answer questions across the spectrum of Participation, Diversity & Inclusion, & Equity. As noted in the slide we separate out data to inform questions about participation (i.e., who is showing up, where, and how?) from Diversity (i.e., how representative is engagement in our spaces of the world; who is showing up and who is missing; how are they experiencing inclusion?), and Equity (i.e., how fairly distributed are our resources an opportunities and who faces barriers which are uneven?).

We will make comparison to world indices along with our Wikimedia participation and assessment data in order to paint an overall picture of the movement's progress in the context of each country.

We will prepare an annual reporting to highlight trends, gaps, and potential areas for improvement, and to share updates on changes to the calculated coefficients of inequity and diversity to hold ourselves accountable to improving our reach and equitable distribution of opportunities and resources.

- **Participation** will examine reach and participation in terms of demographics of our Readers, Editors, & Movement Organizers (Grantees, Event Organizers, Affiliates) from across annual surveys and research databases including key demographics to present comparative distribution data by country and engagement space, as available.
- **Diversity** will overview how we are we doing in terms of diversity & inclusion looking at the current gap in representativeness of certain demographics as well as exploring community experiences of diversity & inclusion based on the Diversity & Inclusion constructs of the annual Community Insights survey.
- **Equity** will overview how we are we doing in terms of equity. It will overview how our inclusion in opportunities and other support map to world indices for gender and developmental factors of inequity such as resource access, infrastructure access, and personal/volunteer investment to participation and outcomes. It will provide a view of the current inequities in order to set shared understanding for strategic interventions to target and attend to those inequities.



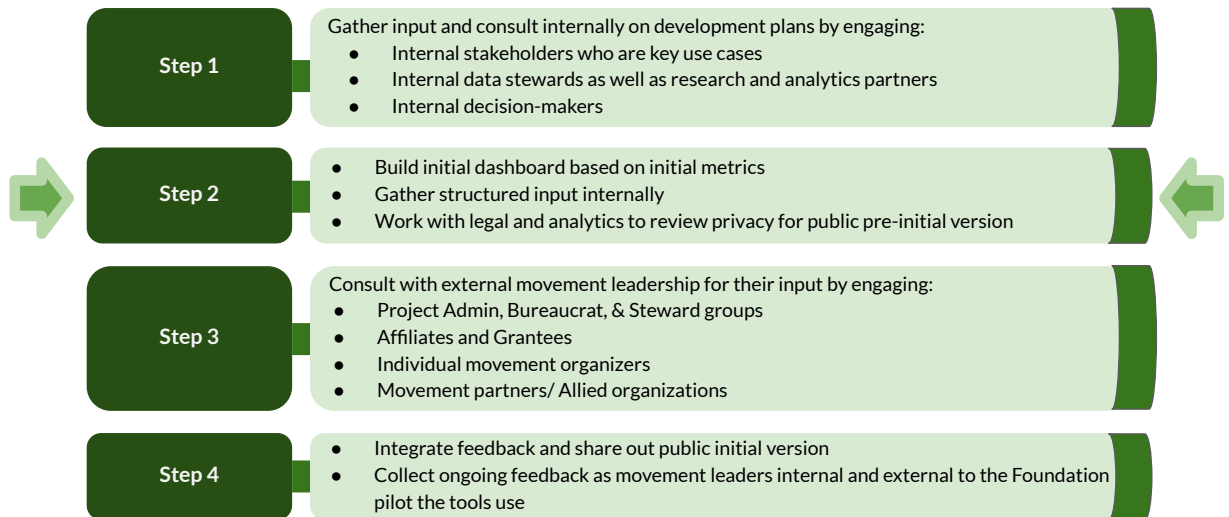
What we propose is a set of products and system that starts and ends with a process, in the middle, dashboard and reporting products.

Gathering and rolling up our available data to widen our aperture for understanding our movement:

Building structured data references and dashboards to enable access, and guiding action through strategic reporting, data-informed decision-making processes, and monitoring for accountability.

Here you see on the left the equity data landscape we propose to triangulate measures across and to the right, the process for ensuring equity index use.

We will use participatory engagement practices to shift power to community for shared sense-making



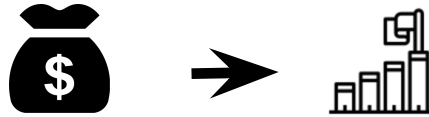
Participatory engagement to enable some power shift to community leaders for shared sense-making. Engage internal data stewards, community decision-makers, and communities represented by the data along the way to develop buy-in for the metrics and shared sense-making.

Steps 1 and 2 aimed at internal consultation and development while later steps are aimed at movement stakeholder consultation (step 3) and potential modification of designs as we iterate (step 4)

Two Use Case Examples (Briefly)

Use Case 1: Direct grant funds and resource opportunities to spaces ready for growth

Actors: Grant applicants, Grants officers, Grant-making committees.



Use Case 2: Direct non-monetary support and partnering opportunities to spaces ready for development.

Actors: Community Organizers, Affiliates, Grantees, Community engagement and support teams.

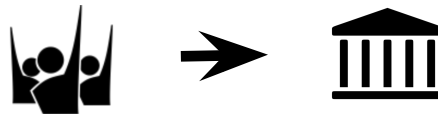


Image credits: [Fund by By OCHA Visual, US from The Noun Project, Public Domain](#); [Progress by Nhor from The Noun Project, CC BY-SA 3.0](#); [Raising hands by Gatuso from The Noun Project, CC BY-SA 3.0](#); [Acropolis by Wolf Böse, DE from The Noun Project, CC BY-SA 3.0](#).

For the past couple years as we were gathering what data we could about our movement landscape and communities, we have had incoming requests for input from strategy groups and market researchers as they were seeking to identify and understand various areas of movement growth. We consulted our available data to develop triangulated metrics of movement presence and growth along various movement indicators which might make our existing data more actionable, and through these experiences, we have begun to map some specific use cases.

Use Case 1: Direct grant funds to spaces ready for growth

Here you see the actors include Grant applicants, Grants officers, Grant-making committees.

The action will be to move grants decision-making from inconsistent and somewhat lacking vetting information to systematic data on relative social asset distributions to make data informed decisions to drive toward our desire strategic change.

Use case 2: Direct partnership opportunities to spaces ready for development and with capacity to support the work.

Here you see the actors include Community Organizers, Affiliates, Grantees, Community engagement and support teams.

We will dive into the details in an applied demonstration later in the presentation.

We will examine the collection of data in light of global data to see how our movement progress stack's up.

Global Data sources to be triangulated:

- World Development Indicators (World Bank)
- Human Development Index (United Nations)
- Global System for Mobile Communications (GSMA)
- Worldwide Governance Indicators (World Bank)
- Freedom Index (Freedom House)
- Women's Business & Law Index (World Bank)
- Population and Gender Inequality Index (United Nations)
- Economic Freedom Index (Heritage)
- Press Freedom Index (Reporters without Borders)
- Social Progress Index

The FY21-22 Development Timeline

Jul-Aug 2021

Mapping external and internal data pipelines.

Initiate language data linkages in the build plans for cross-referencing between languages and geospaces

Sep - Oct 2021

Communications & Privacy review planning

Communications & Privacy review planning

Disseminate documentation to update stakeholders on progress

Jan - Feb 2022

Update for 2021 data, including equity metrics, and analyze for YoY progress for reporting

Promote Wikimedia Movement metrics use in Annual Planning

Nov - Dec 2021

Deliver MVP in superset and onboard teams

Initiate data privacy review for public release plans

Announce project and consultation/pilot plans for participant recruitment

Develop pilot index scoring and analyze for YoY progress or decline & share initial reporting with 2-3 original pilot teams

Feb - Mar 2021

Host community stakeholder pilots and onwiki discussions to consult in optimizing to a few key external use cases around advancing movement DE&I with the dashboard metrics

Apr 2021

Integrate functional improvements and document changes.

Develop priorities for potential future dashboard data iterations.

May 2021

Project review to see if we have met the goals of providing supplemental research meaningful to critical planning needs for the Foundation.

Jun 2021

Complete initial public dashboard build & release.



Underlying proxy indicators and indices

Domain	Measures	Domain	Measures
Readership	Average monthly unique devices Average monthly pageviews	Affiliate Leadership	Highest Governance Type Affiliate grants Affiliate count, size, and tenure Count new recognitions Count organizing hubs engaged
Editorship	Average monthly editors Average monthly % active editors	Access	Population Accessing Internet GSMA Mobile Connectivity Score Social Progress: Access to Basic Knowledge & Information and Communications
Program Leadership	Count Education Events Count GLAM Events Self-reported capacities	Freedom	Freedom Index World Press Freedom Index Control of Corruption Score
Grants Leadership	Annual grants FY Historic grants Growth in Grants	Population	World Population Population Growth

Which measures are best to use and where better measures need to exist is another area key for consultation with all stakeholders both internal and external to the Foundation.

As you can see in the table, there are several domains already identified and integrated into the metrics. However, the domains do not yet include outcome measures such as content which will be added later this FY, or grantee and affiliate impact, which we hope to integrate into next year's reference, and programs impact data, as data become available for integration through language mapping integrations and/or advancement on geo-data for new content, interaction, and other quality metrics in development this year.

In the readership and editorship domains, content and interaction metrics will eventually also be integrated as capacity develops.

How the measures are combined into domain-specific and other high-level metrics are outlined further detail in the appendix and supplemental project write-up

The domain-level metrics: The triangulation of measures for presence and growth across key measurement domains.

Readership = Average of two ranks:

- Avg. monthly unique devices * Annual growth in avg. monthly unique devices
- Avg. monthly pageviews * Annual growth in avg. monthly pageviews

Editorship = Average of three ranks:

- Wikipedia "presence" in avg. unique monthly editors * annual "growth" in avg. unique monthly editors.
- Avg (wikipedia presence * growth, wikidata presence * growth, wikisource presence * growth, commons presence * growth, and mediawiki presence * growth)
- Percentile rank of % editors active or very active

Program Leadership = Average of five ranks:

- Historic Grants (weighted by GDP, PPP), Annual Grants (weighted by GDP, PPP), Education Events, GLAM Events, Self-reported capacities for programs & events

Grants Leadership = Average of three ranks :

- Historic Grants (weighted by GDP, PPP) * (5-year growth/5)
- Annual Grants (weighted by GDP, PPP) * (5-year growth/5)
- Avg. rank (Historic grants count, Annual grant count)

Affiliate Leadership = Average of six ranks:

- Highest governance type, Affiliate tenure, Affiliate grants, Affiliate count, Affiliate size,, count organizing hubs engaged

Access = Average of four ranks:

- Population accessing internet * Annual growth in population accessing the internet
- GSMA Mobile Connectivity Score * Growth in GSMA Mobile Connectivity Score
- Access to Basic Knowledge (from Social Progress Index)
- Access to Information and Communications (from Social Progress Index)

Freedom = Average of three ranks:

- Freedom Index * Freedom Index growth rate
- World Press Freedom Index * World Press Freedom Index growth rate
- Control of Corruption Score * Control of Corruption growth rate

Editor Population Penetration = Average monthly editor count / Population

- Avg. monthly active Wikipedia editor count / Population

Population Presence & Growth = Rank of single metric combination:

- UN population * UN Population growth rate

This list of output metrics is not exhaustive, for some metrics there are transitional computations made in the process, and use of the input measures is not exclusive in all cases. The ideal computational model would allow for a somewhat predictive calculation for the domain-level rankings taking into account a country's current domain presence and it's growth rate, to synthesize the two through multiplication. Where the data are robust enough this model is applied, for Program and Affiliate leadership, the data are not yet viable to support this model. As multiple teams and organizations across the movement seek to identify measurement domains and define monitoring and outcome metrics, along with their inclusion and exclusion criteria and nomenclature need consultation both internally and externally as the movement works together to share and standardize key movement metrics.

The stakeholder level metrics: The remaining Presence & Growth metrics are a set of four facets calculated across our different community data domains (Affiliates, Grantees, Editors, and Readers) and rolled up for cross-comparison.

- **Presence:** The comparative annual data presence across countries based on underlying measures relevant to the stakeholder group.
- **Growth:** The comparative rank in annual growth in a countries data presence based on underlying **Presence** measures relevant to the stakeholder group.

The high-level metrics: The combined metrics are then calculated based upon the set of developed metrics

- **Overall Engagement:** Average of ranks for **Readership, Editorship, Program Leadership, Grants Leadership, Affiliate Leadership**
- **Overall Enablement:** Average of ranks in **Access, Freedom, and Population Penetration**
- **Rank in under-representation:** A combined rank based on gap between rank in **Population Presence and Growth** and rank in **Overall Engagement**

References

(see slide notes)

Chicago Beyond (2019). *Why am I always being researched: A guidebook for community organizations, researchers, and funders to help us get from insufficient understanding to more authentic truth*. Chicago Beyond Equity Series: Volume 1. Retrieved online: August 2019:

https://chicagobeyond.org/wp-content/uploads/2019/05/ChicagoBeyond_2019Guidebook.pdf

Magana, C. (2019) *Tools for mapping power and privilege: Advancing equity in evaluation*. Workshop presented at the American Evaluation Association Summer Institute, June 10, 2019, Atlanta, GA, USA. ([Synthesized here](#))

Krause, H. (2019): *How not to data like a sexist: Feminist Data Evaluation*. Presentation at the American Evaluation Association Annual Meeting, November 14, 2019. Minneapolis, MN, USA. Retrievable online:

<https://weallcount.com/resources-from-live-talks/>

PolicyLink and the University of Southern California Program for Environmental and Regional Equity (2019). *National Equity Atlas*. Policy Link. Oakland, CA, USA.

Retrieved online: July 18, 2019: <https://nationalequityatlas.org/>

We All Count. (2019) *The Data Life Cycle*. Data Assist, Inc. Retrieved online December 8, 2019: <https://weallcount.com/the-data-process/>

Wikimedia Foundation (2017). *Our strategic direction: Service and Equity*. Wikimedia Foundation. Retrieved online: March 10, 2020: https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017

See also:

https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20

Image credits (in order of use in deck):

Slide 3: [A blacktip reef shark \(Carcharhinus melanopterus\), at aquarium museum Liège \(Belgium\) by I, Luc Viatour, CC BY-SA 3.0](#)

(<http://creativecommons.org/licenses/by-sa/3.0/>)

Slide 4: [1971 Dodge Dart Swinger by Greg Gjerdingen CC BY-SA 2.0](#)

(<http://creativecommons.org/licenses/by-sa/2.0/>).

Slide 11: [Vision by Eucalyptus from the Noun Project, CC BY-SA 3.0](#)

(<http://creativecommons.org/licenses/by-sa/3.0/>).

Slide 13: Top left: [Fund by By OCHA Visual, US from The Noun Project, Public Domain](#); Top right: [Progress by Nhor from The Noun Project, CC BY-SA 3.0](#); Bottom left: [Raising hands by Gatuso from The Noun Project, CC BY-SA 3.0](#); Bottom right [Acropolis by Wolf Böse, DE from The Noun Project, CC BY-SA 3.0](#)

(<http://creativecommons.org/licenses/by-sa/3.0/>).

Slide 14: Left: Equity Data Landscape by JAnstee (WMF), CC BY-SA 4.0; Right: Equity Index Use by JAnstee (WMF), CC BY-SA 4.0

(<https://creativecommons.org/licenses/by/4.0/>).

Slide 25: Left: [Fund by By OCHA Visual, US from The Noun Project, Public Domain](#); Right: [Progress by Nhor from The Noun Project, CC BY-SA 3.0](#)

(<http://creativecommons.org/licenses/by-sa/3.0/>).

Slide 26: Left: [Raising hands by Gatuso from The Noun Project, CC BY-SA 3.0](#); Right [Acropolis by Wolf Böse, DE from The Noun Project, CC BY-SA 3.0](#)

(<http://creativecommons.org/licenses/by-sa/3.0/>).

Slide 30: Equity Data Landscape Embedded in Use Process Diagram by JAnstee (WMF), CC BY-SA 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

Slide 31: Bar Chart Example 1: Comparing within embedded geographical context by JAnstee (WMF), CC BY-SA 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

Slide 32: Example Global Heat Map by JAnstee (WMF), CC BY-SA 4.0; Sortable

Table Example 1: Looking at who is on top by JAnstee (WMF), CC BY-SA 4.0;

Sortable Table Example 2: Looking at who is on bottom by JAnstee (WMF), CC BY-SA 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

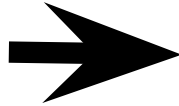
Slide 33: Bar Chart Example 2: Comparing year over year change by JAnstee (WMF), CC BY-SA 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

Slide 38: Sustainable Development Goal icons by United Nations Department of Communications

(https://www.un.org/sustainabledevelopment/wp-content/uploads/2019/01/SDG_Guidelines_January_2019.pdf)

Identify areas ready for growth through grant funds and opportunities

Actors: International Grant-makers, Grant-making committees, Grant applicants, Regional hubs.

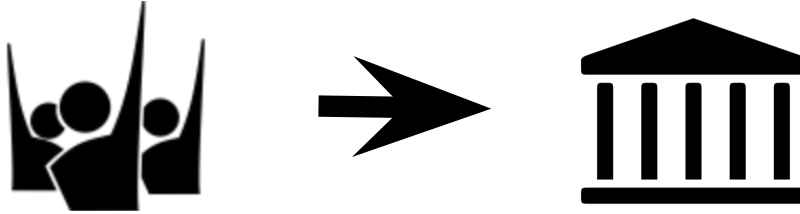


Potential reflection questions for decision-making:

- Is the grantee or partner(s) proposing a project in a country, language, and project space with an adequate reader or editor presence and/or high growth potential, to accomplish its goal(s)?
- Is the grant or partnership request in line with the country's Wikimedia presence, and is the grant's presence proportionate to their base?
- Might the grantee or partner(s) face potential barriers to freedom or access which may prevent reaching goals in the targeted geographic space(s)?
- How many other affiliates may be operating in the same for resources space?

Direct non-monetary support and partnering opportunities to spaces ready for development.

Actors: Community Organizers, Affiliates, Grantees, Governance and recognition committees, as well as International development, engagement, and/or support staff.



Potential reflection questions for decision-making:

- Which countries have a strong affiliate presence and reader base but are lacking in editorship?
- Does a country have a well-balanced movement organizer ecosystem to support an international event or extensive collaborative partnership?
- Which countries have the most potential for growth in different domains?
- To what extent does a country engage in various languages and projects?
- What countries need capacity support to enable movement coordination?

Image credits: [Raising hands by Gatuso from The Noun Project, CC BY-SA 3.0](#); [Acropolis by Wolf Böse, DE from The Noun Project, CC BY-SA 3.0](#).