Analysis of achievements of Nov 2017 event of Sanitation Wikipedia project (comparison with indicators)

Elisabeth von Muench and Diane Kellogg
22 January 2018

This report compares the results achieved to the indicator target values which were set by ourselves. These indicators are explained in a separate document called: “Establishing Indicators for Sanitation Wikipedia Output of Grant: A Discussion Document” by Elisabeth von Muench and Diane Kellogg from 21 Sept 2017, see here.

Summary:

In August and September 2017 we had thought long and hard about the right indicators to chose, how to determine their baseline values and how to determine targets. It is not so easy to quantify Wikipedia work or to quantify the quality of Wikipedia articles. So we had set ourselves a certain number of targets. The analysis afterwards showed that we achieved only some of the targets but not many. We think that in hindsight, some of the targets were actually way too ambitious and we need to change them for the next event (which could take place in March or May).

Details on how the indicators were set and what the targets were are available in the tables below.

Timeline:

- **August to September 2017**: Setting up indicators, determining their baseline values (for 70 articles) and setting up targets
- **19 September to 19 November 2017**: The official period of our Sanitation Wikipedia World Toilet Day Drive, during which time we tried intensively to promote participation and worked with a range of people to achieve our targets. All details of the event, and the results from the Dashboard monitoring are available here.
- **1 December 2017**: This was the date used to determined the values of the indicators on that particular day (give and take a week or two)

Discussion:

The project was a big success, even though we were far, far too ambitious in setting our goals.
--we picked too many indicators
--we picked too many articles
--we were unrealistic about how quickly we could get people excited about Wikipedia work.

It takes time, and our experience has proven that where we do invest the time, we get results over time. UN-Water and the CBS Association are perfect examples.

1. We had great support from UNWater. Elisabeth started cultivating that relationship with Anna Nylander last March (probably well before that) and Diane had a face to face meeting with her and her team in Stockholm. She gave us great visibility and kept offering ideas for promoting it more. She is making plans for helping us find the PhD candidates of the world for future Wikipedia work. We have a long-term partner now, and with THE key organization for World Toilet Day and World Water Day.

2. We made initial contacts with a high percentage of organizations that had booths at World Water Day. Even though we didn't end up getting a lot of participation from many of these, we think of this as the beginning of an education and motivation process that could end up getting the same results we got from UNWater. But it takes time.

3. Some of the work was done offline, using Microsoft Word, and that doesn't show up in the numbers because multiple edits and changes eventually showed up on the Dashboard as one edit, by one person.

Now what we have learned these lessons, the goals or target values we could set for a possible future drive could include:

1. Select ten articles only and identify what needs to be done to improve each.
--Ask a SuSanA member to be a team leader for each article, responsible for making just that one article better.
--Support them as team leader, if asked, but remain hands-off so they are the editor in charge of doing the Wikipedia work.

2. Select five organizations to work with to cultivate long-term commitment to public education via Wikipedia.
--Work with a key person in that organization to identify ONE article that organization should care about.
--Be their Wikipedia editor, for that one article (like we did for Anna for the World Toilet Day article).

3. Cultivate one-to-one relationships with specific SuSanA members we think might be highly likely to continue doing Wikipedia work over the long-term.

Reminder: The Context: Output 2.1.3

The commitment to improving sanitation entries on Wikipedia, the 5th most visited website globally (Source: http://www.alexa.com/siteinfo/wikipedia.org), is in the grant as **Output 2.1.3**:

> *Wikipedia pages that relate to sanitation are improved and serve as an entry point for sharing and collaboration amongst those working in the WSH sector, and to provide orientation to non-specialists and the general public.*

Wikipedia work also contributes to **Outcome 2.1:**

> *Improved thematic content and curated KM tools for the Forum and Website.*

Methodology for setting and quantifying indicators

We came up with three types of indicators for the Sanitation Wikipedia project. These are:

1. Indicators related to what individuals do (as editors)
2. Indicators related to what partner organisations do
3. Indicators related to what key articles look like (originally we planned for 100; in the end we chose only 70; in future we might select even less)

The baseline values, target values and actual values for these indicators are shown in the table below.

The indicators for the key articles were rather time consuming to determine and involved a mixture of quantitative (objective) and qualitative (subjective) analysis.

Here we split the analysis in two parts:

- the entire article (which had 6 quality parameters and a possible maximum score of 13)
- only the lead section (which had 3 quality parameters and a possible maximum score of 8)

We set up a quality assessment system that was relatively complex but which tried to give a fair weighting towards different aspects of quality. This was inspired by a Wikipedia article about article quality metric here.
Our metric were as follows (the maximum points reflects the weighting of that parameter; the higher the maximum, the more important we regard that parameter):

Table 1: Parameters used for quality metric for entire article

	Parameter	Maximum possible value	Average at baseline	Average after event	Comments on how the values were derived
1	Readability	3	0.4	0.4	Based on Flesch score and scaling back to value between 0 and 3 using linear cut-off values, see Appendix.
2	Illustrations	2	1.3	1.3	Manual assessment
3	Comprehensiveness	3	2.4	2.5	Manual assessment
4	Sourcing	3	1.3	1.3	Based on ratio of number of words per reference and scaling back to value between 0 and 3 using linear cut-off values, see Appendix.
5	Neutrality	1	0.9	1.0	Manual assessment, e.g. PR type language would give a lower value (but in general very similar, so could be dropped?)
6	Formatting	1	0.9	0.9	Manual assessment, mainly checking if headings conform with standard headings style
	Sum	13	7.2	7.4	

Observations from table above and in relationship to this indicator that deals with the quality of the entire article:

- The overall score of the articles was hardly improved.
- Only the parameters on neutrality and comprehensiveness were very slightly improved.
- The biggest improvements could be made in future by focusing on the parameter “readability”.

- Working on 70 articles in parallel with a range of new colleagues proved to be too difficult and time consuming to have much impact on the average value across 70 articles.
- If only those 28 articles are considered which we worked on quite intensively, the new average value would be 7.8 (up from a baseline value of 7.4) - which is also not a big improvement.
- There were 28 articles that we worked on either a fair bit (14 articles) or a lot (also 14 articles); very few edits were done on 27 articles; no edits were done on 14 articles - the sum of this is 69); in addition, one article was newly created (the SDG 6 article) - its score went from 0 to 5.7.
    - It is easier to improve the score of a new article or a poorly written article than for an article that is already in fairly good shape.
- If we decide in future to focus on fewer articles, should we take those that were not even touched in this past event, or rather take those with the highest view rates or those with the lowest scores.

Table 2: Parameters used for quality metric for lead section only

	Parameter	Maximum possible value	Average at baseline	Average after event	Comments on how the values were derived
1	Readability	3	0.5	0.6	Based on Flesch score and scaling back to value between 0 and 3 using linear cut-off values, see Appendix.
2	Illustrations	2	1.7	1.6	Manual assessment
3	Comprehensiveness	3	1.1	1.0	Based only on the length of the lead using linear cut-off values, see Appendix.
	Sum	8	3.2	3.3	

Observations from table above and in relationship to this indicator that deals with the quality of the lead:

- The overall score of the leads was hardly improved.
- The parameter on comprehensiveness was slightly reduced as we made some of the leads shorter..
- The biggest improvements could be made in future by focusing on the parameter “readability”.
- Working on the leads of 70 articles in parallel with a range of new colleagues proved to be too difficult and time consuming to have much impact on the average value across 70 articles.
- If only those 28 articles are considered which we worked on quite intensively, the new average value would be 3.4 (up from a baseline value of 3.2) - which is also not a big improvement. (here were 28 articles that we worked on either a fair bit (14 articles) or a lot (also 14 articles))

Table 3: Values for indicators of the three types (baseline, target and actual values)

Indicators	**Baseline date: 12 August 2017**	**Target 1: after WTD 30 Nov 2017**	**Actual values on 1 Dec 2017**	**Comments**
1 - Related to Individuals (people, editors)				

Number of people registered as editors for “Sanitation Wikipedia” (available in Outreach Dashboard tool)	72	100	111	exceeded target
Total number of editors (with at least 1 lifetime edit)	43	60	58	almost met target
% of editors with 5 or more edits in last 7 day period (see here)	8%	40%	8.1%	didn’t meet target but looking at 15-21 Dec (perhaps too late)
Number of people on our Sanitation Wikipedia mailing list where we have an e-mail address or a Wikipedia login name	125	160	154	just below target
% of people on mailing list who have a Wikipedia login	58% (72 of 125)	80%	not assessed, doesn’t seem relevant anymore, not easy to know	
2 - Related to Partner Organizations				
Number of organizations that named a contact person to monitor custom-made watchlist of articles	0	5	0	Not achieved, no uptake on this at all!
Number of organizations whose staff member(s) did make edits for Sanitation Wikipedia	1	4	4 (?)	See note **
3 - Related to Key-100 articles (list of 100 is here or below)				we settled on only 70 in the end
Average value of score for **lead sections** across list of Key-100 articles is approaching maximum value of 8 (see here)	3.2	6	3.3	Target not reached at all, see detailed notes below Table 2.
Average value of score for **entire article** across list of Key-100 articles is approaching maximum value of 13 see here)	7.2	9	7.4	Target not reached at all, see detailed notes below Table 1.
Percentage of articles who cite the JMP 2017 Update report	12% (6 out of 48)	80% (38 out of 48)	19% (9 out of 48)	Target not reached at all

compared to those who should cite it (48 articles determined who should cite it)				(but I would now reduce the 48 figure, on second thoughts)

***** Note: these values were set at the time that the indicators were prepared in August 2017. With the experience from the November event, we will likely revise them now.

** These were: Dannyboi886 (IRC?), Korrigan (WaterAid), Diane, Tom Burgess, Elisabeth, Pouchak, Maureen, Isabelle Blackett (all independent), Arno and Caspar (SEI), Mmekidmfon (small NGO), Kris (Red Cross), Carol (PHLUSH)

Appendix: Cut-off values to convert certain scores to values in range 0-3

Converting readability flesch score to score between 0-3				
A=3	B=2	C=1	D=0	
65	55	40	<30	the higher the better
Converting score for comprehensiveness of lead to between 0-3				
A	B	C	D=0	
500	300	150	<150	the higher the better
Converting score for sourcing to between 0-3				
A	B	C	D=0	
30	60	120	>140	the lower the better

Parameters to Monitor (but not treated as indicators)	**Explanations**
Median number of people watching an article for the Key-100 articles (by adding it to their watchlist)	We would expect this figure to go up over time.
Appearance in Google ranking	Monitor anyway to see if trends emerge. We did observe a trend with UDDT and FSM when those articles were new: it did not take long and they made it to the top of the Google rankings. We will also check if there is an info box at the top right for the Google search, drawing content from Wikipedia.
Popularity of the Offline Medical Wikipedia App for smartphones provided by Kiwix (currently in Top-6)	More information about this interesting Offline Medical Wikipedia App for smartphones that Kiwix is providing, see here. It contains all medical content

	and the entire content from WikiProject Sanitation (about 500 articles), in around a dozen languages – for reading where there is no internet connectivity. About 1 Gb of data.
View rates of Key-100 articles	Graph of page views is available within Wikipedia going back 2 years.
New articles created	Creating new articles is not our focal area. It is usually more impactful to edit existing articles on broad topics rather than creating a new ones that are likely to be rather specialised. Creating new articles can however be a rallying point for small teams, e.g. the new article recently created on container based sanitation.
Images uploaded to Wikimedia Commons	Encourage SuSanA members to provide their images and include them in articles (all photos from SuSanA's flickr photo stream are already included in Wikimedia Commons). We usually upload to SuSanA's flickr account first and then from there to Wikimedia Commons.
General Wikipedians (= people who are not SuSanA related) involved in editing sanitation-related articles	We will monitor this for our Key-100 articles

Table: Baseline values and targets for indicators (as set up in August 2017)

Indicators	**Baseline date: 12 August 2017**	**Target 1: after WTD 30 Nov 2017**	**Tentative Target 2: after WWD 31 May 2018 ***
1 - Related to Individuals (people, editors)			
Number of people registered as editors for "Sanitation Wikipedia" (available in Outreach Dashboard tool)	72	100	150
Total number of editors (with at least 1 lifetime edit)	43	60	80
% of editors with 5 or more edits in last 7 day period (see here)	8%	40%	50%
Number of people on our Sanitation Wikipedia mailing list where we have an e-mail address or a Wikipedia login name	125	160	190
% of people on mailing list who have a Wikipedia login	58% (72 of 125)	80%	100%
2 - Related to Partner Organizations			

Number of organizations that named a contact person to monitor custom-made watchlist of articles	0	5	10
Number of organizations whose staff member(s) did make edits for Sanitation Wikipedia	1	4	8
3 - Related to Key-100 articles (list of 100 is here or below)			
Average value of score for **lead sections** across list of Key-100 articles is approaching maximum value of 8 (see here)	3.2	6	8
Average value of score for **entire article** across list of Key-100 articles is approaching maximum value of 13 see here)	7.2	9	12
Percentage of articles who cite the JMP 2017 Update report compared to those who should cite it (48 articles determined who should cite it)	12% (6 out of 48)	80% (38 out of 48)	100% (all 48 articles)

***** Note: these values were set at the time that the indicators were prepared in August 2017. With the experience from the November event, we will likely revise them now.
** These were: Dannyboi886 (IRC?), Korrigan (WaterAid), Diane, Tom Burgess, Elisabeth, Pouchak, Maureen, Isabelle Blackett (all independent), Arno and Caspar (SEI), Mmekidmfon (small NGO), Kris (Red Cross), Carol (PHLUSH)

Methods to collect data

- For the figures on article quality we have to do an assessment for each article, see here.
- For the figures on editors we can use the Outreach Dashboard for "Sanitation Wikipedia", see here.

Key-100 articles (the first 10 are top priority; total number currently only 69):

	Article	Views per day for one-year period starting 13 August 2016
1	Swachh Bharat Abhiyan	3883
2	Sustainable Development Goals	2596
3	Toilet	1190
4	Hygiene	1032

5	Sanitation	913
6	Open defecation	632
7	Wastewater	523
8	Reclaimed water	315
9	WASH	166
10	World Toilet Day	121
11	Water pollution	5004
12	Gastroenteritis	3771
13	Cholera	3612
14	Diarrhea	2758
15	Sewage treatment	1773
16	Malnutrition	1429
17	Feces	1292
18	Schistosomiasis	1260
19	Menstrual cup	1155
20	Urine	1103
21	Public health	1074
22	Drinking water	1069
23	Compost	1054
24	Waterborne diseases	977
25	Emergency management	942
26	Human feces	796
27	Wastewater treatment	784
28	Composting toilet	760
29	Antimicrobial resistance	671
30	Ascariasis	623
31	Sewage	608
32	Helminths	601
33	Greywater	418
34	Hand washing	410
35	Public toilet	371
36	History of water supply and sanitation	370
37	Pit latrine	368
38	Manual scavenging	342
39	Human waste	336

40	Stunted growth	330
41	Fecal–oral route	292
42	Child mortality	264
43	Helminthiasis	261
44	Groundwater pollution	255
45	Malnutrition in children	254
46	Social marketing	245
47	Neglected tropical diseases	192
48	World Water Day	153
49	Blackwater (waste)	139
50	Dry toilet	113
51	Vacuum truck	111
52	Urine-diverting dry toilet	107
53	Fecal sludge management	81
54	Reuse of excreta	80
55	Menstrual hygiene day	77
56	Behavior change (public health)	75
57	Mass deworming	61
58	Community-led total sanitation	60
59	Improved sanitation	56
60	Human right to water and sanitation	54
61	Global Handwashing Day	49
62	Septic tank	36
63	Omni Processor	26
64	Decentralized wastewater system	23
65	Vermifilter	22
66	Emergency sanitation	20
67	Self-supply of water and sanitation	17
68	Sustainable sanitation	13
69	Container-based sanitation	7
	Sum	50575

Possible articles to add to the article list next time:

- SDG 6 (this article was created by Diane and Elisabeth during this event)
- Microplastics
- Resource recovery