Case Study: EMPOWER – Engaging Museums around Problematic data On Wikimedia's Educational Resources

2019

Case Study: EMPOWER – Engaging Museums around Problematic data On Wikimedia's Educational Resources

Time of the Case study: 1 December 2018 to 30 September 2019

Project by: Swedish National Heritage Board and Wikimedia Sverige

Wikimedia/free knowledge communities involved: Wikimedia Sverige, Wikimedia Finland, Wikimedia Foundation

Edited by: Wikimedia Sverige – final version 31 January 2020

Keywords: Problematic data, outdated terminology, bias, Wikimedia Commons, Wikidata

Licens: CC BY 4.0
https://creativecommons.org/licenses/by/4.0/

Cover image: The binocular camera, Popular Science Monthly Volume 21 (1882)
https://commons.wikimedia.org/wiki/File:PSM_V21_D054_The_binocular_camera.jpg

Future image: Modified Brewster stereoscope, Popular Science Monthly Volume 21 (1882)
https://commons.wikimedia.org/wiki/File:PSM_V21_D056_Modified_brewster_stereoscope.jpg

Abstract

The collections of GLAM institutions[1] include data, images and information which can be perceived as offensive, provoking, racist or outdated. We call this data problematic, and the questions about how to handle it needs attention as materials are opened up to new platforms and made available under free licenses. We who work on open platforms have a common responsibility to create guidelines, methods and tools to make it easier for institutions to open up and allow access to data which might be considered problematic.

There is no doubt about the high value of bringing together experts in this way and being able to share experiences with one another. This work must be done together with GLAM-professionals and the community that is affected by the problematic data in question. Open platforms provide space for new interpretations, opinions and boundaries that may conflict with one another and where one does not need to agree.

The subject is complicated and concerns questions about power and who should have the right to interpret, present and publish historical material. Using an open perspective, this project has gathered experiences and used these to propose possible improvements.

Key conclusions

- Handling problematic data is an ongoing process and historic data and information should not be deleted, but instead updated and supplemented.
- In cases where we add new data and information we should use tools that support transparency in order to increase knowledge.
- We should ensure traceability in the process by presenting all the arguments on why new data and information needs to be added.
- There is a shared ownership of open platforms that makes collaborative learning and collaboration possible.
- Open platforms create opportunities for targeted interventions for affected groups.

---

[1] GLAM institutions – Galleries, Libraries, Archives and Museums

Background

Collections at GLAM institutions have been built up over a long period of time and the documentation describing the content is often a reflection of different time periods and their values. The collections' information and data can be regarded as time capsules from the past with descriptions that can nowadays be regarded as unexplained, unclear and problematic. This applies not only to Sweden, but is generally applicable to all countries where collections are being built up over long periods of time.

The discussion of sensitive material in the collections is a recurring topic in the GLAM sector. There is a lot of documentation of issues through previous seminars and work. The focus has however largely been on interpretation and management within each GLAM institution's own collection. Open platforms and free licenses mean new problems to be solved.

The book *Words Matter*[2] by The National Museum of World Cultures in the Netherlands is a research publication on potentially sensitive words in the museum sector. Words matter contains a review of some 50 problematic words and concepts, with suggestions of words to use in their place. These issues had a major impact in the media when the Terminology group at the Rijksmuseum[3] worked with a selection of titles on works of art in the collection.[4]

There is increasingly a debate within the museum community about the morals and ethics, if not legality, of museums creating and copyrighting media based on unethically acquired objects.[5] On the Wikimedia platforms, material may not carry any restrictions beyond those of CC BY-SA on Wikimedia Commons or CC0 on Wikidata.

---

[2] The book Words matter
[3] The Research and Terminology group on Rijksmusuem
[4] Media 2015 Why the Rijksmuseum Is Removing Bigoted Terms from Its Artworks' Titles
[5] See for example Copyright Cortex.

Problems

Problematic material from the institutions can provide knowledge on open platforms, but the handling requires caution and clarity in the process. What makes the open platforms different from when a GLAM publishes the material on their own website is the often drastic increase in visibility, coupled with the expressed ambition of seeing re-use of the material, expressed by highlighting the free licenses. We have defined problematic data as *"the data and information in the collections that can be perceived as abusive, provocative, racist or outdated".* We have worked on three major general areas during this project:

Questions before uploading data to open platforms
- What are the benefits to the GLAM institutions of making problematic materials available?
- How can we reduce the concerns that exist when handling problematic material?
- How can we get GLAM institutions to make more problematic material available?

Questions after the upload
- How can we create value for, and give ownership to the group that has been exposed?
- What are the needs of processes, guidelines and tools when publishing on open platforms?
- How can the community of the platform aid in updating and contextualising the problematic data?

Follow-up activities
- How to follow up and maintain published material?
- How can you ensure that the material is reused, linked and disseminated in a way aligned with the intentions for making it available?

Execution

We have used an explorative investigative method and contacted organizations that have worked with sensitive data and information. There were 24 participants[6] in the workshop from institutions representing libraries, archives and museums. The selection was made through invitations to Swedish institutions that have had some previous experience of working on open platforms, or have been in contact with Wikimedia Sweden in other projects, and to a selected few international organizations. The workshop was held at Goto 10 in Stockholm.

Communication

The communication has been based on contacts and inquiries with previous partners before the workshop which was conducted on 30 August 2019. In this way, the research work was combined with a discussion of how methods and tools can be developed on open platforms.

Two blog posts, *Workshop on problematic data?*[7] and *– Börja bara! Börja med det värsta ni har!*[8], has been written on Wikimedia Sveriges website. The first blog post was made as an invitation to the workshop and this web post was shared in social media by both Wikimedia Sverige, our partner the Swedish National Heritage Board, and by Digisam, the national coordinator of digitisation. The blog post about the Problematic data workshop was also included in the September newsletter from Wikimedia Sverige.

On the 16th August the project was presented at the Wikimania conference in Stockholm in the GLAM space section *Structured Data on Wikimedia Commons for GLAM-Wiki.*[9]

This report will also be included as a case study in the white paper developed as part of the FindingGLAMs project which will be shared widely in the Wikimedia movement in late 2019 and early 2020.

Presentations

From a larger selection, we asked three lecturers with experience in dealing with problematic data, both in previous work and in current projects. The lecturers were to start the workshop on August 30 with presentations on how they and their organizations work with problematic data. The purpose of the presentations was to give all participants a basis in how problematic data can look and be handled at different institutions for the workshop part of the day. The recorded presentations and slides from the workshop are available on Wikimedia Commons.[10]

---

[6] Global metrics [Problematisk data, Stockholm](#)
[7] Blog post at Wikimedia Sverige [A workshop on Problematic data](#)
[8] Blog post at Wikimedia Sverige (in Swedish) [– Börja bara! Börja med det värsta ni har!](#)
[9] Wikimania Stockholm 2019 [Structured Data on Wikimedia Commons for GLAM-Wiki](#)
[10] Wikimedia Commons [Problematic data workshop 2019](#)

Rijksmuseum's terminology project

Bas Nederveen, an information specialist at the Rijksmuseum, has worked from the start in a special group that has been tasked with critically assessing previously used terminology, and describes this in the lecture "Today's language for today's audience".[11] Problematic terms can be of the following nature:

- Terms that are (or were) not used by a group or a people itself (bushman, eskimo or indian)
- Terms that by origin have a negative meaning (hottentot, heretic, 'mongool' (a person with Down's syndrome))
- Terms that have acquired a negative meaning over time (slave, 'zigeuner' (gypsy), 'allochtoon' (immigrant))
- Terms used from a Eurocentric perspective (native, oriental, primitive, exotic)

Terms   RIJKS MUSEUM

Terms. *An image from Bas Nederveen's lecture on terms and terminology.*

The process starts with choosing the term you want to work with. Terms mainly come from our collection database but may come from another work at the museum, such as a work in progress with a book or an exhibition. In working with terms, you do research and consult experts in the subject, both internally with the curatorial staff at the museum and externally with the groups. Based on this research you choose an alternative. New titles and descriptions are added to the collection database. As of that moment, these become the preferred titles and descriptions when searching the database and the website. All research is documented in an information sheet with a description of the problem, the solutions and which sources were consulted.

---

[11] From the lecture "Today's language for today's audience" from Annexet Goto 10 in Stockholm

Difficult person museum

> **"Who has the right to decide what to be remembered or forgotten from the past"**

Stefan Bohman is a former chairman of Swedish ICOM and has written the book *Skelett i garderoben*[12] (Skeletons in the closet) about persons that are popular for some reason, but also have a problematic history and how this is presented in the person museum. In the lecture at the workshop Stefan describes some conclusions on how difficult person museums[13] work on problematic issues.

There are two main issues to be considered:
1. Who has the right to decide what to be remembered or forgotten from the past?
2. What stories are told and what stories are not told?

These issues will also become relevant when choosing what to upload to open platforms, and what to withhold. Stefan Bohman presents[14] several strategies for how GLAM institutions deal with difficult questions:

- **Full account:** The museum tells the visitors about the difficult questions in exhibitions and other material.
- **Omitting:** The problematic fact is not included in the museums exhibitions or in any other museum material.
- **Double bookkeeping:** The museum presents the difficulties in different ways – one for the ordinary public and one for those with a special interest.
- **Minimizing:** The problematic facts are presented, but in a minimized way.
- **Reduction of responsibility:** The museum claims that everyone did the same, the society during the time "was just like that".
- **Comparison:** The person did really do bad things – but in comparison to his or hers contribution in society it's of lesser importance.
- **Change of subject:** The museum concentrates on other subjects than the persons history and work.

---

[12] Skelett i garderoben: svåra museer, http://libris.kb.se/bib/9kktljdz7lrnxz9p
[13] Difficult issues – Difficult Person Museums
[14] The lecture "Difficult person museums" at Annexet Goto 10 in Stockholm

WikiProject Saami

Susanna Ånäs is project coordinator at Wikimedia Finland and a project leader for the WikiProject Saami.[15] The project examines and improves knowledge about the Saami representation on Wikimedia platforms.
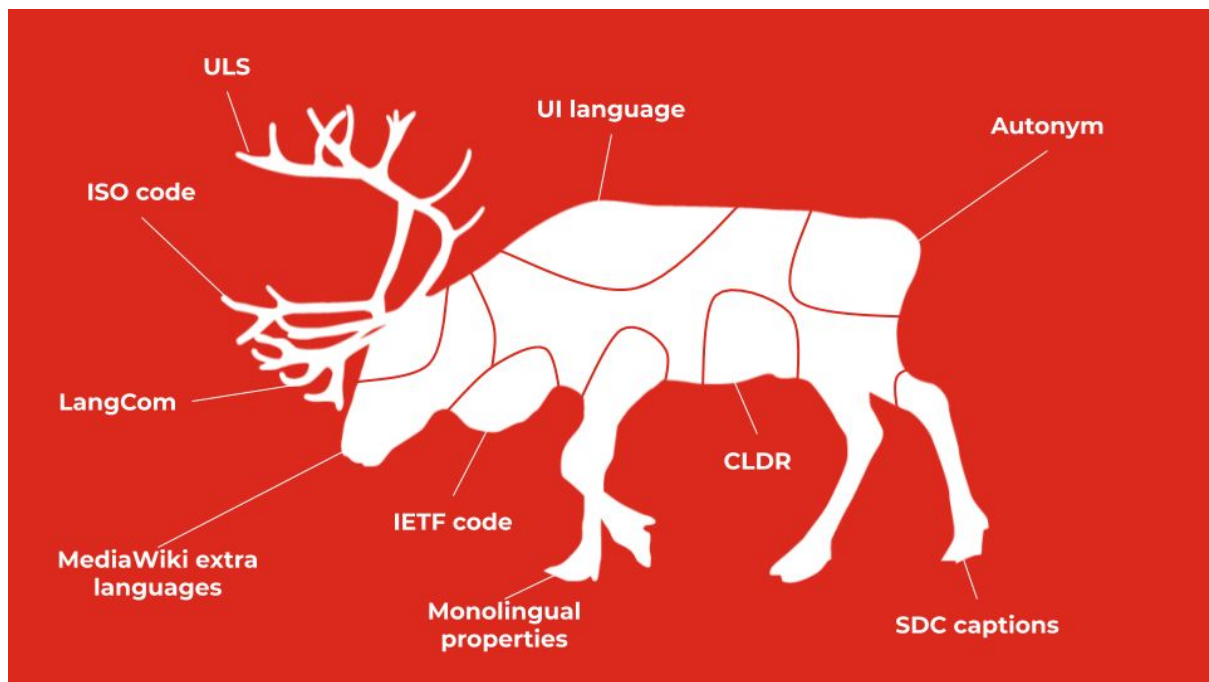
One solution does not work for every Wikimedia project. *An image from Susanna Ånäs lecture WikiProject Saami.*

The vision is to *"make the Wikimedia projects more useful to the Saami communities, help the communities control the circulation of the representation of their culture, and to make the Saami communities, languages, and cultures more visible and factual across all Wikimedia projects."* and Susanna highlights several areas where work can be done to support this vision.[16]

- **Aspects of protection:** Attention to copyright, privacy and personality rights. Culturally sensitive and sacred knowledge. Protection against commercialization, theft and vandalism.
- **Visibility of sensitive data**: Is there a need to exclude information from Wikimedia projects? Ways to describe and filter from display? Remove location data?

---

[15] The WikiProject Saami at Meta-Wiki
[16] From the lecture about WikiProject Saami at Annexet Goto 10 in Stockholm

- **Ways to correct information:** Remove or tag fake indigenous content. Add and use indigenous names and nationalities. Identify and tag personalities and locations when appropriate. Express consent and restrictions. Add knowledge and data provenance.
- **Decolonising the digital commons and terminology:** Translating and importing concepts into Wikidata enables tagging in minority languages. Initiative to import a multilingual Saami museum thesaurus and Saami place names. Propose the use of Traditional Knowledge labels for Saami communities.
- **Consent requires documentation and infrastructure:** Wikimedia environment has opportunities to store consent. OTRS[17] is used for licensing purposes, can be repurposed for consent. In events it is possible to ask for consent. Children cannot legally consent. Both opt-in and opt-out should be possible and the right to be forgotten.

Workshop

In the introduction to the workshop, a brief presentation of Structured data on Commons[18] and a couple of different variations on templates that are already used on Wikimedia Commons. Templates can be seen as alerts and extra customized additional information The purpose was to show the tools that exist and to keep in mind the platforms where development can take place.

The workshop was based on the open space method[19] which takes advantage of the participants' experience and ability to create relevant questions. What was generally desired for future work on problematic data was better discussions, more experiences, good examples, advice and support, more knowledge, how to give back to minorities, strategies and more confidence in what is problematic.

The starting point and the initial question was – *What problematic situations have you encountered or heard about when materials were made available to a larger public?*

This question and subsequent questions generated 18 proposals in areas for further discussion divided into two sessions during the afternoon. The person who suggested a problem area was also given the task of documenting the discussion in a report template. Out of the 18 proposals for in-depth discussion, the result was 10 reports that became working material for further analysis. Some examples of focus areas from the group discussions:

- **Trigger warnings:** Problematic expressions and outdated words.
- **Medical images:** Old language and abusive material.
- **Reproduction of old values:** When there are no valid facts.
- **Lack of interest from colleagues:** How we talk about and describe minorities.
- **Relatives:** Relatives or groups that do not like the museum's story.
- **Unconscientious pictures:** Children, bodies, minorities, tortured and dead people.

---

[17] OTRS – Open-source Ticket Request System
[18] The project Structured data on Commons
[19] The method Open Space Technology

- **Illegal activities:** Legal considerations.

Outcome

Thorough preparatory work should be done before GLAM institutions can share problematic data with free licenses on open platforms. The conclusions after the lectures and workshop is that you should not change historical descriptions and, if necessary, expand and add new updated information. In this way, you maintain provenance and traceability over time. The reports were analysed and grouped around five general issues, presented below.

Ethical perspectives

Ethical perspectives seeks to resolve questions of human morality by defining concepts. There is data where the initial creator might not have asked for the consent of those depicted, or where they lacked the de-facto ability to refuse consent due to asymmetric power relations. There are photographs of dead children and adults, tortured people, and minorities who do not want to be depicted after death in an online collection. This may not always be a legal issue, but an ethical one. There are already ethical guidelines for many cultural institutions, but those need to be further clarified when it comes to uploading materials to open platforms.

Legal aspects

There may be legal aspects of sharing problematic data and files on open platforms. The laws for abusive and racist material may look different in different countries. When is consent needed and how can it be expressed effectively? These issues are often difficult and have to be decided on a case by case basis. There are risks for institutions to end up in a context that the cannot be influence themselves. There is also a fear that images can be used and edited by others and where the institution can become indirectly responsible. The intention when making problematic material available is to have transparency and clarity.

Terminology

Terminology serves to facilitate communication between people who are familiar with a subject area. Certain words and phrases used in the past may be perceived as offensive, unintentional or not, to a more contemporary audience. There are several different glossaries but they are not jointly created. Translation is an important issue when the terminology is specialised and concepts do not overlap perfectly in different languages. Working together with terminology on open platforms let institutions mix and influence the interpretation and the shared meaning of words and how they are used.

Labels

Problematic and sensitive material is often about people, specifically from disenfranchised communities. What is the significance of the description of these communities when made

by an outside party, and how can this change over time? How do we name groups and what happens when we use outdated concepts or names? There are problems with power and interpretability that are manifested throughout the communication of society.

Label systems can be seen as a method for tagging different kinds of problematic material. The project Traditional Knowledge Labels[20] is an example of such a system. To add a information from a label system can provide guidance in the reuse and access to culturally sensitive content. This extra information from a conceptual system can be used for indigenous communities to add existing local protocols for access and use to record cultural heritage that is circulating digitally outside community contexts.

Labels can be seen as a basis for further discussion. Responsiveness and cooperation are two keywords in the work to agree on commonly agreed descriptions. Tools that support different languages are critical to success here.

Reuse

The point of knowledge on open platforms is reuse and it's part of the concept of open that you can't control how this is done. This can be extra sensitive when it comes to problematic data. High-resolution digital images of materials can be reused in commercial contexts, which can be perceived as tasteless, ignorant, and/or inaccurate. Image agencies can place their own restrictions on free material on open platforms, with uncertainty as a result.

The advantages of working with problematic data on open platforms is that it creates a transparency and visibility that shows that these problems are actively taken seriously. It may be better to be the one who devotes resources to this work at an early stage and shows possible solutions to difficult questions.

It is difficult to generalize problematic data as it deals with different types of problems and each case has to be dealt with separately. But open platforms nevertheless provide opportunities to make materials accessible and to be able to influence how the material is curated. Ownership and curation becomes a joint commitment and responsibility.

Two tools to reduce the negative impact of reuse are; increased public understanding of how freely licensed material can be reused, and that the context in which material is encountered does not necessarily reflect its origins. Increased awareness on the side of content providers about potential reuse, so as to be prepared in the case of unwanted reuse.

---

[20] Traditional knowledge labels
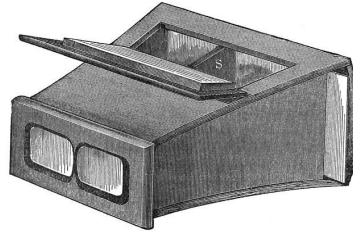
Future

The number of uploads will increase as more material is given free licenses. There will continue to be a need for advice and support as the background, material and process will be different in each upload. We see an opportunity to take advantage of all experiences by documenting good examples in the work of developing the management of sensitive material.

Documentation and processes to develop

Wikimedia has the capability to create a knowledge bank around problematic data and can support collaboration between institutions working on transparent open methods and assist in disseminating results. We recommend further work, development and projects in the three areas described below. For a good result we suggest working with institutions that have unpublished collections that contains problematic data.

Ethical perspectives

Existing ethical guidelines should be updated, to also include open platforms. This can help institutions when preparing a collection with problematic content or when looking at sharing content on open platforms for the first time. We can release a new code of ethics under a free license and focus on the part that affects publishing on open platforms. We can write and organize a *code of ethics for open platforms* with inspiration from four areas:

- Code of ethics for **galleries**.
- Code of ethics for **libraries**.[21]
- Code of ethics for **archives**.[22]
- Code of ethics for **museums**.[23]

---

[21] Professional Codes of Ethics for Librarians.
[22] The ICA code of ethics for archives.
[23] The ICOM code of ethics for museums.

Legal aspects

There are several different laws that are relevant when dealing with problematic data. There is a need for more legal knowledge in this area with an international perspective. It is also an area where changes are constantly taking place so continuous education and platforms where this knowledge is easy to update are desirable. This work should preferably be done by country but can be compared to each other in the form of a table.[24] Working in a project on an open platform can be used for four areas of legal aspects.

- **Harassment, discrimination** and other abusive treatment.
- The law and rules of **copyright**.
- **Integrity** and general data protection regulation.
- Indigenous and minority rights.

Terminology

Defining problematic terminology and preferable alternatives to it is an important step in recognising and addressing problematic data. It is crucial to work on this together on open platforms so that more institutions can influence the interpretation and, by extension, the shared meaning of the terminology. When more actors join the chosen platform becomes an authority. It is important to work directly on open platforms with free licenses so that there are opportunities for several players to contribute. If the terminology is to be used and remain relevant, it must also be maintained and updated. We can work with the terminology directly on three global projects with support for hundreds of languages.

- Add words at **Wiktionary**[25]
- Add information at **Lexicographical data**[26]
- Add knowledge and sources on articles at **Wikipedia**

Wikimedia tools for development

There are specific tools and processes that can be developed to facilitate work in databases and on open platforms. This can be about using the tools in new processes or in new ways that are suitable for uploading and handling problematic data. Our proposal is to continue with processes and development around these three areas, each with an associated set of tools. **Preferably with an unpublished collection that contains problematic data** where the purpose is to release it on an open platform. The tools can be developed in different ways depending on the system and platform. On Wikimedia platforms and projects, development takes place continuously with the aim of collaborating and benefiting from each other.

---

[24] Example of comparative table Freedom of panorama
[25] Wiktionary is a free-content multilingual dictionary.
[26] Lexicographical data – words, phrases and sentences, described in many languages.

Structured data

A common use of structured data is to use it on web pages so that search engines can easily find and understand your data. This requires some kind of agreement between the transmitter (the web page) and the receiver (the search engine) so that the data is interpreted in the same way by both parties. One advantage of using structured data is that you can make clearer and better searches. You can also open your data so that others can do searches from outside of your web pages. This is called open linked data, and there are many opportunities for development in this area. For problematic data, one such opportunity for development is how to transfer information from text to structured data without losing the nuances from the text.

The tool Structured Data on Wikimedia Commons[27] aims to allow structured and machine-readable metadata to be associated with the free media files on Wikimedia Commons, to make them easier to view, search, edit, organize and re-use. This is one way in which data could be marked as potentially problematic using different statement and properties. The strength of Structured Data on Wikimedia Commons is that it is an open system, meaning anyone can add information. This opens up a discussion about individual interpretations that fit the character of problematic data very well. The result is a combination of domain knowledge of the content and how the technology supports the dissemination of this curated knowledge.

Structured Data on Wikimedia Commons is fairly new and more development is needed to make this mechanism easily usable for the case of problematic data.

Templates

One way to highlight important information can be to work with templates using visual elements, such as warning texts or other labels. This can be done easily in many systems and can be a way to make visible selected problematic parts at an early stage. These templates can be designed in different ways depending on the situation.

Templates[28] on Wikimedia Commons can also be further developed in both form and content. There are opportunities for specialized solutions based on presuppositions. These specialized solutions can be investigated and, above all, tested in real situations. There are ethical guidelines and laws in most areas and one way to add knowledge about problematic data is to actively link to them. Referring to a guideline or law can be a way to initiate a discussion about whether or not certain problematic data and images should be on free platforms. It is also a way to have contact with organizations that are a part of creating the practices in the area. A template can have one or more organizations as senders and refer

---

[27] Structured Data on Wikimedia Commons is tool for adding structured data on media files
[28] Templates on Wikimedia Commons

to one or more relevant authorities in each case. We can also explore more about whether a labeling system as, for example, *Traditional knowledge labels*[29] work on open platforms.

It should be noted that templates, and metadata in general, often gets divorced from media representations when these are published or reused on other platforms.

Properties

Properties[30] in Wikidata are a way to describe an object. The use of properties can in a longer perspective become an important aspect in handling problematic data. Wikidata properties that describe terms as problematic might be used in a similar way to what dictionaries do. A description in dictionaries for certain words and concepts can for example be *outdated* when they want to show that the word is no longer in use. Properties for describing objects can be developed by addressing the need for similar solutions on open platforms. Developing new properties on Wikidata follows a process where several users can participate and in this way the property becomes embedded within the community even before it is used.

Having many properties that describe an object can be a way to increase the granularity of the information. Having the involved actors agreeing on terms and concepts makes it easier to collaborate. Along with reaching an agreement, you can also link to lexicons, sources, and discussions that describe more about the background of why the property is used for the problematic data.

---

[29] Traditional knowledge labels
[30] Proposal for a Property at Wikidata