**Assessment of New Web Archive Collection Development and Appraisal Approaches and Methodologies**

Kristi Torp

NLM Associate Fellow 2023-2024

January 2024

Project Sponsor: Christie Moffatt

Co-sponsor: Caitlin Sullivan

**Table of Contents**

[Acknowledgments 3](#_Toc155617562)

[Abstract 4](#_Toc155617563)

[Introduction 5](#_Toc155617564)

[Objective 5](#_Toc155617565)

[Methods 5](#_Toc155617566)

[Outcomes 6](#_Toc155617567)

[Discussion 10](#_Toc155617568)

[Recommendations 10](#_Toc155617569)

[Conclusion 11](#_Toc155617570)

[Appendix A: Rubric Definitions 12](#_Toc155617571)

[Appendix B: Appraisal Tables 14](#_Toc155617572)

# **Acknowledgments**

I would like to thank Christie Moffatt and Caitlin Sullivan for their guidance during this project and for inviting me to the NLM Web Collecting and Archiving Working Group meetings to learn more about their process. I am also grateful for the links and discussions about digital archives and for nurturing my curiosity on the subject.

I would also like to thank Kathel Dunn for her continuous support in helping me clarify my project goals and objectives.

# **Abstract**

**Introduction**

The NLM Web Collecting and Archiving Working Group, as a significant part of their collection development strategy, preserves web content through theme-based collections using the Internet Archive’s Archive-It service. To streamline the selection process for these thematic collections, they adopted a new appraisal rubric in Spring 2023.

**Objective**

The purpose of this project is to evaluate the effectiveness of the new appraisal rubric for web archive content. This is an exploratory project to assess what is and is not working with the new rubric.

**Methods**

First, test the new rubric with randomly selected starting point URLs (“seeds”) from 2 new collections in development and compare the results to the groups’ scores, noting where there are weak points. Second, visualize how rubric category scores affect approval for inclusion in a collection. Finally, research how other organizations evaluate web content for inclusion in web archives.

**Outcomes**

This assessment determined that uniqueness is a difficult category to appraise. Additionally, it found that most of the unapproved content was determined to be out of scope and thus unaffected by the scores in the rubric. It also determined that the scores for Uniqueness and Representation/Inclusion categories have more exceptions for inclusion than other categories.

**Recommendations**

First determine if the seeds are in scope. The remaining seeds should then be evaluated with the rubric to determine the best content. Discarded content should be reconsidered for uniqueness or representation after the other seeds have been assessed.

# **Introduction**

The National Library of Medicine selectively provides permanent access to web content through the work of the NLM Web Collecting and Archiving Working Group. This group collects biomedical, theme-based content to document a particular subject or theme, preserve scientific research or diverse perspectives, and chronicle NLM institutional history. Multiple websites are generally collected as part of a broader theme, event, or topic, prioritizing content at risk of loss or believed to be of vital interest to current and future audiences. Curating large amounts of web content is often overwhelming and time-consuming as each resource must be individually evaluated. Working Group members, NLM staff, and sometimes the general public recommend seed URLs for inclusion in a collection, and subject matter experts in the Working Group review and approve them for inclusion in the collection.

# **Objective**

The Web Collecting and Archiving Group adopted a new selection rubric in the spring of 2023 to improve the selection process so reviewers could focus their attention on the highest quality recommendations that are in scope for a particular collection. They implemented the rubric after their extensive work on the NLM Global Health Events/Coronavirus 2019 (COVID-19) pandemic web archive collection (January 2020-May 2023), which includes over 20,000 archived web resources. This exploratory project aims to evaluate the new rubric's effectiveness for selecting web archive content with testing, analysis, and research. Once the evaluation is complete, the project will note the strengths and weaknesses of the new rubric, with suggestions for improvement and further discovery.

# **Methods**

I reviewed the new rubric, adapted from one created by the Library of Congress, and became acquainted with proposals to develop new web archive collections on Health Equity and Rare Diseases. The Web Archive Group had already identified content, known as seeds, for evaluation. The Health Equity collection had 128 seed recommendations, and Rare Diseases had 133 seed recommendations when the project began. Initially, the sponsors and I agreed I would identify new seeds and then evaluate them using the rubric as if I were a group member.

I reached out to a former colleague with an interest in rare diseases for materials recommendations. I added some of these resources to the Rare Diseases spreadsheet and used the rubric to evaluate them. The rubric categories are: Informational Value, Representation/Inclusivity, Page Level, Uniqueness, Intrinsic Value, and Updated in the last 12 months. Each category received a score of 1, 2, or 3 based on the strength of its content. A score of 1 is considered poor, 2 is average, and 3 is good, top-quality content for the collection (see Appendix A for definitions and scoring instructions). However, given my lack of experience and the subjective nature of some categories, I struggled to evaluate the new seeds. After meeting with the project sponsors, we agreed it would be more productive if I evaluated a sample of their seeds first and saw how my scores matched with the rest of the team. We would then determine the next steps based on that information. My project sponsor also gave me reading material about the origin of the rubric to give me greater context and help me understand the categories better.

I randomly selected 15 seeds recommended for each collection, copied them to a new personal spreadsheet, and deleted any previous scoring information from my copy. I then began evaluating seeds and tracked my thoughts in notes as I looked at each seed (see Appendix B). After completing the evaluation, I compared my scores with those of the group and tracked where the scores did not match. It is important to note that the group did not fully evaluate some seeds in Rare Diseases because they were deemed out of scope early on. These scores have significant discrepancies in their comparison numbers because of this, and I did not include those unscored areas in this report as being dependent on the rubric. After viewing the score comparison, and noting which categories had difficulty, the project sponsor asked if I would compare how scores affect inclusion in the collection. By comparing the scores in each category with whether it was approved for collection, we hoped to see if the rubric scores affected inclusion in a collection or not.

The project sponsors also suggested I review the collection development policies of members of the International Internet Preservation Consortium for their web archiving practices, particularly if they define uniqueness or value. **The Society of American Archivists (SAA) does not have an entry for "uniqueness" in their dictionary, making it a challenging appraisal category.** I viewed the policies for Columbia University Libraries, the University of Alberta Libraries, the National Library of Australia, the UK Web Archive, and the Library of Congress. The Library of Congress was the only institution that mentioned uniqueness or value, simply stating in their supplementary guidelines[[1]](#footnote-1), "important factors to the Library of Congress include…uniqueness and quality of information provided...." I decided to expand my search for seed appraisal in general web archiving. This search provided useful information about web archive appraisal in differing contexts and more background information about the Library of Congress rubric.

# **Outcomes**

Table 1 shows which categories had the most discrepancies in scoring. In Rare Diseases, the categories with the most discrepancies were "Uniqueness" and "Informational Value." In Health Equity, the most discrepancies were in "Representation/Inclusion" and if the site was "Updated in the last 12 months?" The project sponsors were not concerned about the latter category for this analysis, which puts "Uniqueness" as the second highest for discrepancies. When reviewing my scoring comments, 13/30 comments mentioned difficulty assessing "uniqueness" (see Appendix A). We noted that this was a problematic category in general to appraise and moved on to evaluating how scores affected inclusion.

Table 1

*Number of Scores for Recommended Seeds That Didn’t Match (out of 15 each)*

|  |  |  |
| --- | --- | --- |
| **Rubric Section** | **Rare Diseases** | **Health Equity** |
| Informational Value | 8 | 3 |
| Representation/Inclusivity  | 3 | 6 |
| Page Level | 2 | 2 |
| Uniqueness | 9 | 4 |
| Intrinsic Value | 1 | 2 |
| Updated in the last 12 months? | 1 | 6 |

When I compared the scores with inclusion in the collection, I looked at each collection separately and noticed that the scores for Representation/Inclusion did not necessarily affect if a seed was approved for inclusion in the collection. Health Equity had 128 seeds when the project began. One seed was not scored at all due to being eliminated before scoring, and two were neither approved nor disapproved at the time of this writing. Of the remaining 125 seeds, 62 were approved, and 63 were disapproved. Of the 63 disapproved, 60 were deemed out of scope but were graded in the rubric. When visualizing the scores as a percentage (see Figure 1), there is conflicting evidence about the impact of the scoring rubric on whether or not a seed is included in the collection. In this collection, approved seeds are more likely to have higher scores; however, the same is not true for the reciprocal. Disapproved seeds do not necessarily have lower scores. Disapproved seeds are simply more likely to have been deemed out of scope despite having higher scores (see Figure 2). It is important to note that Representation/Inclusion scores had an inverse effect on approval.

Figure 1

The Web Collecting & Archiving Group scored each seed with 1, 2, or 3 in each category based on the strength of its content. 1 = poor content, 2 is average, and 3 is good. Note that the category Representation/Inclusion has a high amount (nearly 70%) of poor scoring content approved for inclusion.

Figure 2

The Rare Diseases Collection had 133 seeds at the start of the project. Eighty-three were approved, with 50 disapproved. All the disapproved seeds were also marked as out of scope. In this collection, the rubric has a more substantial impact on whether a seed is approved or not for the collection (see Figure 3). Higher scores are more likely to be approved, and lower scores are more likely to be disapproved. However, both Uniqueness and Representation/Inclusion had more than 60% of their low-scoring seeds approved. Additionally, Figure 4 shows that the scores of the out-of-scope seeds in the Rare Diseases collection were lower than those in Health Equity.

Figure 3

The Web Collecting & Archiving Group scored each seed with 1, 2, or 3 in each category based on the strength of its content. 1 = poor content, 2 is average, and 3 is good. Note that the categories Uniqueness and Representation/Inclusion have a high number of poor scoring content approved for inclusion.

Figure 4

NLM adapted its rubric from another created by the Library of Congress, so I looked for information to understand that origin better. I found a presentation that detailed the creation of their appraisal rubric. Noting the lack of definition from the SAA, the Library of Congress presentation states that "uniqueness" is included in the rubric as it is part of the definition of what makes something archival in nature: "The characteristics that are inherent in archival documents resulting from the circumstances of their creation, including naturalness, organic nature or interrelationship, impartiality, authenticity, and uniqueness."[[2]](#footnote-2)Also worth noting is that the Library of Congress also determined to use an average total score as the bottom threshold for inclusion. They determined the rubric worked for them by starting with 2200 seed nominations, 235 of which were eligible records, and narrowing it to 124 for collection. These numbers are difficult to compare to these NLM collections, which only started with around 130 seeds.

# **Discussion**

 I will note some limitations of this project along with suggestions for future investigations. Three of the sites reviewed independently were deemed out of scope before being fully assessed by the Web Collecting and Archiving Group, leaving incomplete final scores. I was unaware I would need to account for this, and in hindsight, I would have excluded those sites from my random evaluation. Additionally, it is worth considering that some categories may be more difficult to score due to the collection topic or theme. For example, Representation/Inclusivity had the most mismatched scores for the Health Equity collection, which is about greater inclusivity, and the Uniqueness category had the highest mismatched scores in the Rare Diseases collection, which is about a unique topic. Perhaps certain collections are more complex to evaluate if their theme is about that category.

In the future, it would be beneficial to determine if the rubric works differently for theme-based versus event-based collections when content is rapidly changing and at high risk for loss, and more expansive in the types of resources collected. Furthermore, I recommend evaluating if placement on the recommendation spreadsheet affected how seeds are scored, particularly regarding uniqueness. For example, is the 20th site on the list overlooked because the resources listed before were evaluated earlier? I attempted to find a pattern for this project by color-coding the scores. However, the results were inconclusive and could be investigated more in future evaluations.

# **Recommendations**

After completing this evaluation, I recommend that the Web Collecting and Archiving Group utilize the rubric differently. This evaluation shows the rubric as an effective resource for determining if seeds are in scope, though the scope may change as the understanding of the breadth and depth of a topic evolves. This can include decisions about whether to include local authorities or only national ones and how broad or narrow the topic will be. Using the categories Informational Value, Top-Level Domain, and Intrinsic Value, will check for scope with greater detail and find the top-scoring seeds. The low-scoring seeds can then be checked for Uniqueness or Representation/Inclusion as possible exceptions. This should be done after scoring the other categories, as there will be a better understanding of what is already in the collection and what is missing (see Figure 5).

Figure



Additionally, I have included adjusted wording for the definition and criteria of uniqueness with my additions in italics.

**Uniqueness**

*(note: evaluate a unique author on Intrinsic Value)*

• The content or resource is unique in nature (*content, reputation, location*).
• We have not already collected on this topic *in this collection*.
• This resource provides a new perspective on an already collected topic.
• This resource contains a significant update on a previously collected topic.
• The source itself is unique, *or at risk of being lost*, and therefore worthwhile to collect.
• This resource presents original content not duplicated across other web-based content. Is the information on this website already represented in previously nominated content?
• This resource is THE source itself, not ABOUT the source.

I also recommend repeating this assessment as more collections are completed using the rubric and for collections with over 200 in-scope seeds, as collection size could affect outcomes. Future assessments could also look for patterns with placement on the spreadsheet and whether or not there are differing trends between event-based and topic-based collections.

# **Conclusion**

 Currently, the appraisal rubric effectively eliminates seeds that are out of scope for the collections. Analyzing the appraisal rubric demonstrated that uniqueness is a difficult category to appraise and may be best assessed separately from the total score along with Representation/Inclusion. By understanding how the rubric is and is not working, the NLM Web Collecting and Archiving Working Group can use it to set collection goals and find the strongest seeds for future collections.

# **Appendix A: Rubric Definitions**

**Informational Value**

• The information throughout the seed is substantial.
• There is breadth and/or depth to the information.
• The information appears to be accurate (without fact-checking or significant vetting) and professional.
• The seed contains information that is directly within scope of the collecting needs as defined in the collection proposal.
• The seed is a resource that its audience might visit many times - i.e., for blog updates, event updates, new resources, etc. It is not a "one and done" seed.

3= The seed contains substantial, accurate, and professionally presented information that is directly within scope of the collection proposal. There is breadth and depth to the information.
2= The seed contains substantial and accurate information that is somewhat within the scope of the collection, or might be tangentially related to collection scope. Information may be represented elsewhere in the collection. There is limited breadth or depth to the information.
1= The seed contains accurate information but is not very substantial, or it is not a resource that an audience might visit many times.

**Representation/Inclusion**• The content reflects underrepresented voice(s) or perspective(s) that are not currently present in the collection, including dissenting or opposing views.
• The content reflects underrepresented voice(s) or perspective(s) from historically marginalized groups, communities, and individuals, and/or furthers the diversity and inclusion of the collection. These include:
1. Race and ethnicity - Individuals who identify as other than White and non-Hispanic (including but not limited to Asian, Asian American and Pacific Islanders; Black, African or African American; Hispanic, Spanish, or Latino; Native Hawaiian or Pacific Islander; Middle Eastern or North African)
2. Access to care - Individuals who have not needed a medical visit in the past 12 months or who cannot easily obtain or pay for medical care as needed
3. Age - Children 17 or younger and adults 65 or older
4. Annual Household Income - Individuals with household incomes equal to or below 200% of the Federal Poverty Level
5. Disability - Individuals with either a physical or cognitive disability
6. Educational attainment - Individuals with less than a high school degree or equivalent
7. Gender identity - Individuals who identify as gender variant, non-binary, transgender, or something else
8. Geography - individuals who reside in rural and non-metropolitan areas
9. Sex assigned at birth - individuals who are neither male nor female (i.e. intersex)
10. Sexual orientation - individuals who identify as asexual, bisexual, gay or lesbian, or something else

3= The seed is solely dedicated to a historically underrepresented group, community, or perspective, and furthers the diversity and inclusion of the collection
2= The seed contains some information on a historically underrepresented group, community, or perspective, but is not the seed's sole focus
1= The seed does not contain information specific to any historically underrepresented group. It provides general information

**Top Level Domain**• The seed is an entire website/resource devoted to the subject matter
• The seed is NOT a single article, blog post, video, press release, etc.

3= Top level domain seed
2= Sub-domain seed
1= Sub-page or individual article, blog post, video, press release, etc.

**Uniqueness**

• The content or resource is unique in nature.
• We have not already collected on this topic.
• This resource provides a new perspective on an already collected topic.
• This resource contains a significant update on a previously collected topic.
• The source itself is unique, and therefore worthwhile to collect.
• This resource presents original content not duplicated across other web-based content? Is the information on this website already represented in previously nominated content?
• This resource is THE source itself, not ABOUT the source

3= The seed provides a new perspective, addresses a topic that we have not already collected on, provides a significant update on a previously collected topic, presents content not already duplicated elsewhere, or fills a collection gap.
2= The seed is somewhat unique, but content may already be reflected in other seeds in this collection.
1= There is valuable information here, but this is not a very unique resource.

**Intrinsic Value**

Can the content creator be identified?​
Regardless of content, is the creator significant enough to the collection scope for the website to be archived?​ (Federal Government, for example)
Is this a federal site? If so, it has intrinsic value to us.

3= The identity of the content creator (individual, corporate body, collective group) is significant on its own and the resource includes novel documents/media or hard to find records. OR, it is a federal site.
2= The identity of the content creator (individual, corporate body, collective group) is somewhat significant on its own and the resource includes some novel documents/media or hard to find records.
1= The identity of the content creator (individual, corporate body, collective group) is not significant on its own and embedded documents/media are available via other resources

# **Appendix B: Appraisal Tables**

RARE DISEASES

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Informational Value** | **Representation/Inclusivity**  | **Page Level** | **Uniqueness** | **Intrinsic Value** | **Updated in the last 12 months?**  | **RUBRIC SCORE** |
| **URL** | <https://rarechromo.org/> |
| **WAG** | 3 | 3 | 3 | 3 | 3 | 1 | 16 |
| **Kristi** | 2 | 3 | 3 | 2 | 3 | 1 | 14 |
| **Kristi’s notes** | I have no idea if this is unique. I feel like I have to look at EVERYTHING on the site and in the collection first. Also, note this is a UK site. |
| **URL** | <https://rarediseases.info.nih.gov/> |
| **WAG** | 3 | 2 | 2 | 2 | 3 | 1 | 13 |
| **Kristi** | 3 | 2 | 3 | 3 | 3 | 1 | 15 |
| **Kristi’s notes** | I think it's unique because it's federal and there probably isn’t much else out there that the government has. This seed felt easy to evaluate.  |
| **URL** | <https://everylifefoundation.org/> |
| **WAG** | 2 | 2 | 3 | 2 | 3 | 1 | 13 |
| **Kristi** | 3 | 2 | 3 | 3 | 3 | 1 | 15 |
| **Kristi’s notes** | This one felt easier overall, too. Still not sure how unique it is.  |
| **URL**  | <https://rare-x.org/> |
| **WAG** | 2 | 2 | 3 | 2 | 3 | 1 | 13 |
| **Kristi** | 2 | 1 | 3 | 3 | 3 | 1 | 13 |
| **Kristi’s notes** | It's definitely unique! Is it too narrow though? |
| **URL** | <https://nymacgenetics.org/> |
| **WAG** | 2 | 2 | 3 | 2 | 3 | 1 | 13 |
| **Kristi** | 2 | 1 | 3 | 1 | 3 | 1 | 11 |
| **Kristi’s notes** | Only for one geographic location, doesn't seem like anything unique however. |
| **URL**  | <https://www.sicklecelldisease.org/> |
| **WAG** | Null | Null | 3 | 3 | 3 | 0 | 9 |
| **Kristi** | 2 | 3 | 3 | 2 | 3 | 1 | 14 |
| **Kristi’s notes** | Easy to evaluate because it’s a single disease, not within scope |
| **URL**  | <https://training.mmlearn.org/blog/are-you-caring-for-someone-with-a-rare-disease> |
| **WAG** | 2 | 1 | 1 | 2 | 1 | 0 | 7 |
| **Kristi** | 1 | 1 | 1 | 1 | 1 | 0 | 5 |
| **Kristi’s notes** | Low scores, but easy to evaluate |
| **URL**  | <https://www.ultragenyx.com/> |
| **WAG** | 2 | 1 | 3 | 2 | 3 | 1 | 12 |
| **Kristi** | 1 | 1 | 3 | 2 | 2 | 1 | 10 |
| **Kristi’s notes** | Hard to evaluate the value and uniqueness. People might be interested in gene therapy, but probably have broader resources elsewhere to look for it. I have no idea. |
| **URL** | <https://www.encephalitis.info/rasmussens-encephalitis> |
| **WAG** | Null | Null | 2 | 2 | Null | 0 | 4 |
| **Kristi** | 1 | 1 | 1 | 2 | 3 | 0 | 8 |
| **Kristi’s notes** | Easy to evaluate because it’s a single disease and not within scope. Note, UK |
| **URL** | <https://knowrare.com/blog> |
| **WAG** | Null | Null | Null | 2 | Null | Null | 2 |
| **Kristi** | 2 | 2 | 2 | 2 | 3 | 1 | 12 |
| **Kristi’s notes** | I noticed a 2023 on the site, but the blog posts aren't dated, so is it really updated? |
| **URL**  | <https://www.hrsa.gov/opa/program-requirements/orphan-drug-exclusion> |
| **WAG** | 2 | 1 | 1 | 2 | 3 | 1 | 10 |
| **Kristi** | 2 | 1 | 1 | 2 | 3 | 1 | 10 |
| **Kristi’s notes** | Rubric really helped on informational value on this one, unique was still hard to determine |
| **URL**  | <https://raregenomes.org/> |
| **WAG** | 3 | 1 | 3 | 2 | 3 | 0 | 12 |
| **Kristi** | 2 | 1 | 3 | 2 | 3 | 0 | 11 |
| **Kristi’s notes** | no depth or breadth, but shows what is being done, so some value. Unique is hard on this one. |
| **URL**  | <https://alliancerm.org/> |
| **WAG** | 3 | 1 | 3 | 3 | 3 | 0 | 13 |
| **Kristi** | 2 | 1 | 3 | 1 | 3 | 0 | 10 |
| **Kristi’s notes** | too narrow, I think, for informational value. Could be valuable to show what is happening in the private sector. Not sure if it’s unique though. |
| **URL** 97,  | <https://www.rareiscommunity.com/blog/> |
| **WAG** | 3 | 1 | 2 | 3 | 3 | 1 | 13 |
| **Kristi** | 2 | 2 | 2 | 2 | 3 | 1 | 12 |
| **Kristi’s notes** | Sometimes I think evaluating value and uniqueness would be easier if it was the main domain and not the blog. |
| **URL** 108 | <https://travere.com/> |
| **WAG** | 3 | 1 | 3 | 1 | 3 | 1 | 12 |
| **Kristi** | 2 | 1 | 3 | 1 | 3 | 1 | 11 |
| **Kristi’s notes** | Another private sector company. How many do you include? Any? Not unique.  |
|  |

Health Equity

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **URL** | **Informational Value** | **Representation/Inclusivity**  | **Page Level** | **Uniqueness** | **Intrinsic Value** | **Updated in the last 12 months?**  | **RUBRIC SCORE** |
| **URL**  | <https://www.healthequity.com/> |
| **WAG** |  |  |  |  |  |  | 0 |
| **Kristi** |  |  |  |  |  |  | 0 |
| **Kristi’s notes** | I'm not even going to bother. It's an HSA card |
| **URL** 26,  | <https://scorh.net/> |
| **WAG** | 3 | 3 | 3 | 2 | 3 | 0 | 14 |
| **Kristi** | 2 | 3 | 3 | 2 | 3 | 1 | 14 |
| **Kristi’s notes** | It’s only for 1 state, so kind of out-of-scope, but it’s rural health. It’s updated, but not with anything substantial (just event info) |
| **URL** 28,  | <https://www.commonwealthfund.org/publications/scorecard/2023/jun/2023-scorecard-state-health-system-performance> |
| **WAG** | 3 | 2 | 2 | 3 | 3 | 1 | 14 |
| **Kristi** | 3 | 3 | 1 | 3 | 3 | 1 | 14 |
| **Kristi’s notes** | This one was pretty easy to evaluate |
| **URL** 64,  | <https://www.apha.org/> |
| **WAG** | 3 | 1 | 3 | 2 | 3 | 0 | 12 |
| **Kristi** | 3 | 2 | 3 | 3 | 3 | 1 | 15 |
| **Kristi’s notes** | American public health association, so that’s unique. Feels easy when it’s lots of threes. |
| **URL** 71,  | <https://thinkculturalhealth.hhs.gov/> |
| **WAG** | 3 | 2 | 3 | 2 | 3 | 1 | 14 |
| **Kristi** | 3 | 3 | 3 | 3 | 3 | 1 | 16 |
| **Kristi’s notes** | Do I evaluate inclusion for content or for the intended audience? I’m not sure. |
| **URL** 79,  | <https://www.niehs.nih.gov/research/supported/translational/justice/index.cfm> |
| **WAG** | 2 | 2 | 1 | 2 | 3 | 1 | 11 |
| **Kristi** | 2 | 3 | 1 | 2 | 3 | 0 | 11 |
| **Kristi’s notes** | Federal, kind of hard to evaluate 1 page of a whole site. What all am I considering with that? Just this page by itself? |
| **URL** 80,  | <https://www.niehs.nih.gov/research/supported/centers/ehd/index.cfm> |
| **WAG** | 2 | 3 | 1 | 2 | 3 | 0 | 11 |
| **Kristi** | 2 | 3 | 1 | 2 | 3 | 0 | 11 |
| **Kristi’s notes** | federal. But only about 1 grantee, so I think not super unique. |
| **URL** 82,  | <https://www.phrases.org/story/prioritizing-health-equity-in-mississippi-communities/> |
| **WAG** | 3 | 2 | 1 | 3 | 1 | 1 | 11 |
| **Kristi** | 2 | 3 | 1 | 3 | 1 | 0 | 10 |
| **Kristi’s notes** | Main domain might be worth evaluating. Hard to gauge info value when it’s only about 1 thing. Is it within scope since it’s Mississippi? That might help with informational value. |
| **URL** 88,  | <https://www.niehs.nih.gov/research/supported/centers/ehd/grantees/unm/index.cfm> |
| **WAG** | 2 | 3 | 1 | 2 | 3 | 1 | 12 |
| **Kristi** | 2 | 3 | 1 | 2 | 3 | 0 | 11 |
| **Kristi’s notes** | I think it’s the phrase “within scope”. Good grantee example, but I’m thrown off by the “scope” so I put 2 |
| **URL** 92,  | <https://beam.community/> |
| **WAG** | 3 | 3 | 3 | 3 | 2 | 1 | 15 |
| **Kristi** | 3 | 3 | 3 | 3 | 3 | 1 | 16 |
| **Kristi’s notes** | Felt easy to evaluate.  |
| **URL** 93,  | <https://lgbtqequity.org/> |
| **WAG** | 3 | 3 | 3 | 3 | 2 | 1 | 15 |
| **Kristi** | 3 | 3 | 3 | 3 | 3 | 0 | 15 |
| **Kristi’s notes** | unsure if updated, otherwise easy. |
| **URL** 98,  | <https://health.ri.gov/programs/detail.php?pgm_id=1106> |
| **WAG** | 2 | 3 | 1 | 3 | 3 | 1 | 13 |
| **Kristi** | 3 | 3 | 1 | 3 | 3 | 1 | 14 |
| **Kristi’s notes** | State level, but really good informational value. Hard to say if it’s within scope since it’s 1 state. |
| **URL** 118,  | <https://health.hawaii.gov/healthequity/> |
| **WAG** | 1 | 1 | 2 | 2 | 3 | 1 | 10 |
| **Kristi** | 1 | 1 | 2 | 1 | 3 | 1 | 9 |
| **Kristi’s notes** | state level, "The Office of Health Equity is currently being established". No info. |
| **URL** 122,  | <https://www.cdc.gov/minorityhealth/index.html> |
| **WAG** | 3 | 2 | 2 | 2 | 3 | 1 | 13 |
| **Kristi** | 3 | 3 | 2 | 3 | 3 | 1 | 15 |
| **Kristi’s notes** | blog and newsletter are updated, not sure about main page.  |
| **URL** 125 | <https://minorityhealth.hhs.gov/Default.aspx> |
| **WAG** | 3 | 3 | 2 | 3 | 3 | 1 | 15 |
| **Kristi** | 3 | 3 | 3 | 3 | 3 | 1 | 16 |
| **Kristi’s notes** | Link needed to be fixed. Got rid of "default". Graded this on main domain site after I did that. |
|  |

1. Library of Congress. Collections Policy Statements Supplementary Guidelines. https://www.loc.gov/acq/devpol/webarchive.pdf [↑](#footnote-ref-1)
2. Library of Congress. 2022. “When Time and Resources are of the Essence: Archival Appraisal and the Library of Congress Coronavirus Web Archive” International Internet Preservation Consortium.” Slide 21. [↑](#footnote-ref-2)