Association of Learned
and Professional
Society Publishers

ORIGINAL ARTICLE

The state of scientific PDF accessibility in repositories: A survey in Switzerland

Alireza Darvishy ⬤,[1]* Rolf Sethe,[2] Ines Engler,[2] Oriane Pierrès,[2] and Juliet Manning[2]

[1]Institute of Applied Information Technology, Zurich University of Applied Sciences, Winterthur, Switzerland

[2]Faculty of Law, University of Zurich, Zurich, Switzerland

ORCID:
A. Darvishy: 0000-0002-7402-5206

*Corresponding author: Institute of Applied Information Technology, Zurich University of Applied Sciences, Obere Kirchgasse 2, Winterthur, 8400, Switzerland.
E-mail: dvya@zhaw.ch

Abstract: This survey analyzes the quality of the portable document format (PDF) documents in online repositories in Switzerland, examining their accessibility for people with visual impairments. Two minimal accessibility features were analysed: the PDFs had to have tags and a hierarchical heading structure. The survey also includes interviews with the managers or heads of multiple Swiss universities' repositories to assess the general opinion and knowledge of PDF accessibility. An analysis of interviewee responses indicates an overall lack of awareness of PDF accessibility, and shows that online repositories currently have no concrete plans to address the issue. This paper concludes by presenting a set of recommendations for online repositories to improve the accessibility of their PDF documents.

Keywords: accessibility, accessible science, education, online repositories, PDF documents, visual impairment

INTRODUCTION

The United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) from 2006, which was signed and ratified by more than 180 countries as well as the European Union, institutes specific accessibility goals for different areas of life, such as work, healthcare, and education (UN General Assembly, 2007). The primary purpose of this convention was to ensure equal rights and to protect the integrity of people with disabilities at an international level. Similarly, the Disability Discrimination Act (DDA) in Switzerland, which came into force in 2004, introduced guidelines for the general accessibility conditions of public institutions and transport, residential buildings, workspaces, as well as research, education, and educational facilities (Federal Assembly of the Swiss Confederation, 2002). Despite these efforts to create equal rights and opportunities for people with disabilities, many obstacles are still encountered, specifically in tertiary education. To date, fewer people with disabilities pursue higher education compared with people without disabilities (Federal

Statistics Office of Switzerland, 2021). To change this, institutions of higher education must become more inclusive in all areas of operation: buildings and websites must become more accessible, awareness must be raised, communication should be barrier-free, and accessible teaching materials and educational resources must be provided, among other important changes. A guarantee of accessible teaching materials and library media, such as accessible portable document format (PDF) documents, must be established.

PDF is by far the most common file format used for scientific and academic publications. But PDFs are often created with visual layout in mind. This presents a challenge to persons with visual impairments, many of whom rely on the use of assistive technologies, such as screen readers, to read digital content aloud. PDF documents often lack sufficient structural information to be properly interpreted by screen readers. This means that users with visual impairments may struggle to navigate the document, understand its organization, or locate specific content.

The international standard for PDFs, 'PDF/UA' (short for PDF Universal Access), establishes the norm for the quality of

accessibility in PDFs (Drümer & Chang, 2013). For PDF documents to be accessible, they must entail certain minimal accessibility features. PDF creators can include or add these accessibility features manually when making their documents, or can choose from a number of tools which automate or partially automate this process (Darvishy, 2018; Darvishy et al., 2012; Darvishy & Hutter, 2013; Doblies et al., 2014). Some new research is also investigating the potential of artificial intelligence to automate document accessibility (Darvishy et al., 2016; Schmitt-Koopmann, Huang, & Darvishy, 2022; Schmitt-Koopmann, Huang, Hutter, et al., 2022). For PDFs created by scanning physical pages, OCR (Optical Character Recognition) technology can convert them into a standard machine-readable PDF format.

Two of the most crucial accessibility features in PDFs are tags and hierarchical heading structures. A tag is like a label that provides semantic information about an element in a PDF document; for example, 'bullet list', 'new paragraph', 'table', and so forth. Hierarchical heading structure refers to the organization of headings and subheadings in a structured manner, typically following a nesting format (e.g., heading levels like H1, H2, and H3). A screen-reader user can jump from heading to heading, which provides a quick overview of the document and allows them to navigate easily within it. It should be noted that PDFs can also have other accessibility features, such as alternative text for images, content design (e.g., colour contrast, font, table design), and metadata. However, tags and hierarchical headings could be considered the most basic, and most critical of the accessibility features. Clearly structured headings enhance the accessibility of documents for everyone. However, documents lacking these two features become extremely impractical and nearly unusable for screen-reader users. Additionally, documents that are missing these simple features are also usually unlikely to contain other, more complex accessibility features. For this reason, and for the sake of simplicity, the authors have focused on tags and hierarchical structure as representations of the most basic PDF accessibility.

In the past, researchers have analysed the state of accessibility in international repositories, such as Semantic Scholar and Web of Science; the analyses of PDF document accessibility in such repositories yielded results between 2 and 15% of all papers having certain minimal accessibility features (Nganji, 2018; Wang et al., 2021). However, no such evaluation has been carried out in the Swiss repositories. As Switzerland has pledged to guarantee access to higher education, it is critical to assess where its universities stand on that topic. For that reason, this study raises the two following research questions:

Research question 1: What is the percentage of accessible scientific PDF documents in Swiss repositories?
Research question 2: What is the general opinion and knowledge about PDF accessibility of managers of Swiss repositories?

A mixed-method approach combining quantitative and qualitative analyses was chosen to answer these two questions. First, interviews with the managers of multiple Swiss online repositories were conducted to assess their knowledge of PDF document

---

Key points

- In order to be accessible, portable document format (PDF) documents require tags and hierarchical heading structures.
- We performed a quantitative and qualitative analysis of PDF document accessibility in Switzerland.
- Fewer than 11% of documents in Swiss repositories have the minimum accessibility features.
- Repository owners and managers were generally unfamiliar with PDF accessibility.
- PDF accessibility was not considered a priority by most repositories.

---

accessibility, their specific measures to create a more accessible service, and their plans for improving accessibility. Second, a statistical analysis was conducted with 2,500 papers from Swiss online repositories, examining their accessibility for people with visual impairments. Two minimal accessibility features were analysed: the PDFs had to have tags and a hierarchical structure. Finally, a set of recommendations for online repositories and institutions of higher education in general is provided to facilitate their path to more accessibility.

METHODS

Semi-structured interviews

In seeking interviewees, the authors decided on conducting interviews with the managers or employees of nine Swiss university repositories. The sample included the five German-speaking repositories from the quantitative analysis, as well as three from the French-speaking region of Switzerland and one from the Italian-speaking region. The interviews were conducted in German, French, and English. No interview was longer than 35 min. The results of the quantitative analysis were not mentioned in any of the interviews. All interviews took place online and were recorded. After all interviews were conducted, auto-generated transcripts were created and analysed.

Initially, ten interview questions were developed by the authors. After deciding on four main topics of interest, the authors revised the questions, finally deciding on 12 questions for the interview, some with sub-questions. The four topics were *Knowledge and Opinions*, *Priority and Awareness*, *Accessibility Measures*, and *Wishes and Future Plans*. All categories comprised three questions, except the last one. The first question about the role and function of the interviewee was not included in any of the categories. All questions were posed in each interview. The complete list of questions can be found in Appendix A.

Before starting an interview, the participants were informed about the purpose of the interview, data usage and storage,

recording, and automatic transcription. They were assured of anonymity and had the option to skip any questions. After the opportunity to ask any questions, they consented to the participation conditions.

For the analysis of the interviews, all answers for every question were extracted, put in a grid and, where suitable, coded (for example, for yes/no questions, a 'yes' answer was coded with a one, and a 'no' answer with a two). This provided a quick overview of which questions were routinely answered negatively. Keywords were extracted from qualitative answers. Further comments from the interviewees that could not be classified under any of the posed questions were added as separate content. Additionally, the words and concepts that were mentioned most frequently across all interviews were gathered.

Quantitative analysis of PDFs

The repositories of the five German-speaking universities in Switzerland were chosen for an in-depth quantitative analysis of their PDF contents. All Open Access PDF documents were pulled from each repository for 2018 to 2022. In total, over 50,000 documents were downloaded. A random selection of 100 papers for every year from every repository was chosen, which resulted in a sample size of 2,500 documents in total. Using the 'Actions' command on Adobe Acrobat Pro, which allows a user to automatically apply the same command for multiple documents simultaneously, the chosen documents were scanned for two accessibility features: the existence of tags and a hierarchical heading structure. The resulting HTML files, which Adobe Acrobat Pro exports after finalizing the set command, were imported into the programming language R, which counted the features automatically. The count included a sum of all papers containing tags, all papers containing hierarchical headings, and all papers containing both features, separately.

RESULTS

Semi-structured interviews

The responses of the interviewees differed in degree of precision and amount of information, while some questions were answered by only stating 'I don't know'. No interviewee chose to omit any questions. Concerning their role at the repositories, most interviewees were heads of the repository, of the university library, or of the Open Science department.

Quantitative results of the interviews

Some of the questions asked in the interviews were quantifiable 'yes/no' questions; the responses to these questions are laid out in Table 1. An additional table showing the individual responses of each repository can be found in Appendix B.

None of the nine repositories surveyed reported keeping internal statistics on accessibility (Question 4a). Two repositories reported that training/conferences on accessibility were available

TABLE 1 Numbers of 'Yes' and 'No' responses to quantifiable questions from nine repositories.

Quantifiable questions	Responses	
	Yes	No
Do you keep internal statistics on the accessibility of accessible publications? (Question 4a)	0	9
Are there training or conferences on the topic of 'accessibility' for the employees? (Question 6a)	2	7
Is attendance in such trainings or conferences mandatory? (Question 6b)	0	9
Is there any exchange between the repositories of universities in Switzerland – through associations, conferences, etc.? (Question 7)	7	2
Do you check the accessibility of new incoming documents before publishing them on the repository? (Question 9)	0	9
Do you offer a service of providing alternative document formats? (Question 10a)	1	8

(Question 6a), but in no cases were such trainings mandatory (Question 6b). Likewise, no repository reported checking the accessibility of new incoming documents (Question 9). Most repository representatives reported having an exchange between Swiss repositories in (Question 7). Only one interviewee responded that a service to provide alternative documents was available (Question 10a).

Knowledge and opinions

Regarding their knowledge about accessible PDFs, most interviewees provided very rudimentary answers, ranging from no knowledge of the international PDF/UA standards, to some knowledge of their existence but without the know-how to implement them. Analogously, when asked about what an accessible PDF document is in their opinion, the answers mostly remained at a base level. Most mentioned the importance of readability, formatting (such as colours and font size), and structure, but did not further elucidate what a 'good document structure' entailed. Some alluded to assistive technologies and machine readability. In some cases, 'accessibility' was confounded with 'attainability', as accessible PDFs were described as having to be traceable and have no paywalls or logins necessary. Estimates on the accessibility of the documents in the repositories ranged from 'low' to 'mixed' and 'moderate'. Most interviewees mentioned that an accurate estimate is not possible, as they do not keep an internal statistic on document accessibility.

Priority and awareness

The interviewees reported a very low to low priority of the topic in their repositories. As an exception, one interviewee alleged to

a very high priority of the topic, as they have an external department dedicated to accessibility at their university. Most repositories mentioned that they currently focus on long-term archiving of PDF documents. Almost all repositories mentioned how they are currently working on converting all documents to PDF-A to guarantee their long-term availability. One repository alluded to not having any monetary or workload capacities to give accessibility a higher priority.

Some repository representatives mentioned that their employees are sensitized on the topic of accessibility of teaching materials and documents in general, and that they participate in conferences or university-internal training to further their knowledge and skills. Participation in such conferences and training was not compulsory at any repository. Concerning communication and cooperation between repositories, most interviewees indicated their participation or membership in the Open Access working group in Switzerland. This working group was described as the best portal of communication between libraries and repositories, but also that accessibility had never been a topic of discussion.

Accessibility measures

Pertaining to the concrete measures the repositories undertake in their efforts to be more accessible, the accessibility of the repository platform itself was the biggest point of discussion. Most repositories were in the process of migrating to a new software or digital platform. With this change, many interviewees also mentioned the hope for more possibilities to make the repository more accessible. No one mentioned specific measures to improve the accessibility of the documents in the repository. Most repositories reported not having a guideline to ensure the accessibility

of the publications, except the repository of one university which insists on students and employees submitting a barrier-free version of their theses and dissertations. No repository checked the accessibility of newly incoming documents before publication. However, almost all repositories checked the metadata of all documents before publication. Almost no repository employee had had the experience of a student asking for an alternative version of a document, but almost all added that they would happily aid the student on a case-by-case basis. No repository provides this type of help as an official service.

Wishes and future plans

In the last block of questions, the interviewees provided a wide variety of ideas and concepts. Most interviewees mentioned the need for broader accessibility measures and more education of employees. It was mentioned that accessibility should be a natural part of scientific work and not something to be conceptualized a-posteriori. The benefits of having a diversity office in universities were alluded to, and the hopes for more concrete guidelines and laws were stated. Finally, concerning plans to improve the accessibility of their own repository, almost no repository could name concrete projects. Most mentioned the migration to another platform and the improvement of the user interface again, and a few said that there were no definite plans to improve accessibility whatsoever.

Interviewees provided several reasons that explained the potential challenges in prioritizing PDF accessibility in the future. An emphasis was put on the lack of resources. Most repositories mentioned time and money to be the biggest hurdle towards more inclusivity and accessibility. They also felt that there were

TABLE 2 Number of PDFs with tags, headings, or both from 2018 to 2022 for each of five examined repositories.

Repository number	Accessibility Feature	2018	2019	2020	2021	2022	Total
1	Tags	0	0	0	1	0	**1**
	Headings	0	0	0	1	0	**1**
	Both	0	0	0	1	0	**1**
2	Tags	14	15	10	17	14	**70**
	Headings	17	18	15	24	16	**90**
	Both	13	15	9	17	12	**66**
3	Tags	5	5	6	7	12	**35**
	Headings	6	9	9	12	15	**51**
	Both	5	5	6	7	11	**34**
4	Tags	0	0	0	0	0	**0**
	Headings	0	0	0	0	0	**0**
	Both	0	0	0	0	0	**0**
5	Tags	27	29	41	40	44	**181**
	Headings	29	35	39	39	43	**185**
	Both	27	28	34	37	42	**168**

almost no suitable courses of action for them, as they saw the responsibility of PDF accessibility in the hands of the publishers who provide the publications. For some interviewees, a need for more laws and stricter guidelines was considered necessary for them to also take accessibility measures into consideration.

Quantitative analysis of PDFs

All 2,500 downloaded documents were scanned for tags and headings. Table 2 provides an overview of the results. The number of documents containing tags or headings or both features ranged between 0 and 44 out of 100 through all years and repositories. The mean number of documents that had tags was 11.48 per year; for headings, it was 13.08, and for both features, it was 10.76 documents.

The count in two repositories rendered few to no documents with either tags or headings. Two other repositories had 5 to 20% of documents that contained the accessibility features, and one repository had 20 to 45% of documents with tags, headings or both features.

Across all years and repositories, about 11.5% of the downloaded documents had tags, 13.1% had a hierarchical heading structure, and 10.8% had both (Table 3).

COMBINED RESULTS

Taking the results from the quantitative and qualitative analyses into account, some relevant relationships became evident. In repositories in which the interviewee's knowledge of PDF accessibility surpassed a certain rudimentary level, there were a higher total number of PDF documents with tags or headings in the repository. Generally, there was a lack of knowledge on what an accessible PDF is, how to create an accessible PDF, and what tools one can use to check its accessibility. Nevertheless, the more knowledge, the likelier it was for the repository to be willing to have more accessible PDFs. This willingness was assessed by the number of counter questions from interviewees about what options they had to make documents accessible and how they could provide a service that was more accessible altogether. However, awareness of the issue and willingness to create a more accessible service alone did not automatically translate into actual accessibility. A gap between willingness to adapt and actual actions was found, as most repositories could not provide any concrete plans to increase document accessibility in the future.

TABLE 3 Total number and percentage of PDFs with tags, headings, or both.

Accessibility feature	Total (no. of PDFs)	Total (% of PDFs)
Tags	287	11.48%
Headings	327	13.08%
Both	269	10.76%

The repositories of universities with accessibility guidelines had more PDFs with tags and headings. A caveat, though, is that concrete guidelines are needed to make this positive relationship work. For example, in one university it was compulsory that students submit all theses and dissertations using an official document template, which contained headings as well as tags. Additionally, it was found that certain common practices of some universities, for example, using LaTeX, hinder the creation of accessible PDFs.

DISCUSSION AND RECOMMENDATIONS FOR FUTURE

Similar to the findings in other databases (Nganji, 2018; Wang et al., 2021), a majority of PDF documents in Swiss repositories are not accessible. This may be partly due to the fact that the repositories themselves do not require documents to be accessible before submission. The responsibility of PDF accessibility is a shared one – authors, publishers, and repositories can all play a role. The best solution would be to oblige publishing companies to produce accessible documents right from the beginning. It is much easier to consider the accessibility requirements when producing documents than to laboriously adjust them afterwards. From an economic perspective the publishing company is the 'cheapest cost avoider'. Placing the burden of accessibility on the shoulders of publishers has a second advantage. Published documents would generally be accessible because it would not make sense to produce a non-accessible version for the publication and an accessible one for the repository. However, reality shows that such a solution remains a future dream. In the long run we need either agreements between publishing houses and universities/repositories or even a legislative approach. Until that time comes, the second-best solution is only that repositories provide for PDF remediation tools and establish rules for PDF submission.

As this study revealed, repository managers do not see it as their responsibility to ensure the accessibility of PDF documents. Instead, their priority is to ensure the long-term availability of PDF documents. Nor are they empowered to provide accessible PDF documents: the interview responses make it clear that there is a widespread lack of knowledge of PDF accessibility among repositories, and few to no resources are directed towards improving it. Repository owners generally did not know what an accessible PDF is nor how to create one, and were unaware of existing tools for checking or improving PDF accessibility. This lack of awareness is consistent with other recent findings on the topic (Jembu Rajkumar et al., 2020). The identified issues, such as displaced responsibility, lack of knowledge, and the low proportion of accessible PDFs in Swiss repositories contradict the country's commitment to provide fair access to higher education to everyone.

Hence, to aid repositories and institutions of higher education to make science more accessible, the authors have developed a set of recommendations:

1. **Seek out knowledge and raise awareness on accessibility, and specifically on the accessibility of PDF documents.** A broader understanding of the issues and their possible solutions provides a better foundation to translate ideas into actions. Repository representatives should participate in workshops, training, and conferences on the subject, and look for experts in the field. It is also important to actively include people with disabilities in this process, as they may identify problems that aren't evident otherwise, while also providing a different perspective that can lead to more well-rounded solutions.

2. **Consult international accessibility guidelines.** There are a number of existing guidelines and standards, such as WCAG 2.1 (Web Content Accessibility Guidelines, 2023) and ISO 14289 (ISO Standard 14289-1:2014, 2014), which define the steps required to make documents accessible. These guidelines can serve as a useful starting point and could help institutions to later define their own internal accessibility goals and guidelines.

3. **Collaborate with other universities to establish accessibility goals together.** Institutions can help each other in the path to reach those goals, as many universities use the same or similar software for their repositories, and most are in a similar position with regards to accessibility.

4. **Make accessibility an institutionally binding requirement**. As long as publishing companies do not provide accessible documents right from the start, authors and students should be required to submit documents only in an accessible format and repository managers should be obliged to accept only such documents. If templates are provided, ensure that these include at least the basic accessibility features. Institutions should also regularly measure and monitor the accessibility of their repositories to maintain an overview of their progress and identify areas for future improvement. Further, the accessibility of incoming documents should be checked, and feedback should be provided to publishers to also make them aware of the problems with the documents they provide.

5. **Provide document accessibility services.** For documents that remain inaccessible, institutions should provide a remediation service for their students and researchers. Upon request, this service should check and improve the accessibility of a given document, and convert it into an accessible PDF or other accessible formats such as HTML or Word.

CONCLUSION

In this paper, we presented a quantitative and qualitative analysis of the state of PDF document accessibility in Swiss repositories. The quantitative analysis of five Swiss repositories rendered sobering results—with less than 11% of the sample having both minimal accessibility features. Ensuring a PDF document has at least these features means the document is more usable for people with reading disabilities, for example, visual impairments, thereby reinforcing their equal right to educational materials and scientific documents. The numeric results are somewhat explainable by the answers rendered in the interviews, as a lack of knowledge and resources was the apparent source of the low number of accessible documents.

In summary, the repositories of Swiss universities still have a lot of potential to be more inclusive and accessible. Still, the authors are optimistic that the repositories will continue to strive to broaden their knowledge and offer a more accessible service in the future.

Following the same procedure from this survey, an additional survey is planned, in which international publishers will be analysed on their accessibility standards and also invited to participate in interviews.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

Darvishy, A. (2018). PDF accessibility: Tools and challenges. In *Computers helping people with special needs: ICCHP 2018 proceedings, part I* (pp. 113–116). Springer. https://doi.org/10.1007/978-3-319-94277-3_20

Darvishy, A., & Hutter, H. P. (2013). Comparison of the effectiveness of different accessibility plugins based on important accessibility criteria. In C. Stephanidis & M. Antona (Eds.), *Universal access in human-computer interaction. Applications and Services for Quality of life. UAHCI 2013. Lecture Notes in Computer Science* (Vol. 8011). Springer. https://doi.org/10.1007/978-3-642-39194-1_36

Darvishy, A., Leemann, T., & Hutter, H. P. (2012). Two software plugins for the creation of fully accessible PDF documents based on a flexible software architecture, Computers helping people with special needs. In K. Miesenberger, A. Karshmer, P. Penaz, & W. Zagler (Eds.), *ICCHP 2012. Lecture Notes in Computer Science* (Vol. 7382). Springer. https://doi.org/10.1007/978-3-642-31522-0_94

Darvishy, A., Nevill, M., & Hutter, H. P. (2016). Automatic paragraph detection for accessible PDF documents. Computers Helping

People with Special Needs: 15th International Conference, ICCHP 2016, Linz, Austria, July 13-15, 2016, Proceedings.

Doblies, L., Stolz, D., Darvishy, A., & Hutter, H. P. (2014). PAVE: A web application to identify and correct accessibility problems in PDF documents. In K. Miesenberger, D. Fels, D. Archambault, P. Peňáz, & W. Zagler (Eds.), *Computers Helping People with Special Needs: ICCHP 2014 Proceedings, Part I.* Springer. https://doi.org/10.1007/978-3-319-08596-8

Drümer, O., & Chang, B. (2013). PDF/UA in a Nutshell. PDF Association. Online eBook. https://www.pdfa.org/wp-content/until2016_uploads/2013/08/PDFUA-in-a-Nutshell-PDFUA.pdf

Federal Assembly of the Swiss Confederation. (2002). Federal Act on the Elimination of Discrimination against People with Disabilities, 13. https://www.fedlex.admin.ch/eli/cc/2003/667/en

Federal Statistics Office of Switzerland. (2021). Persons with Disabilities. https://www.bfs.admin.ch/bfs/en/home/statistics/economic-social-situation-population/equality-people-disabilities/disabilities.html

ISO Standard 14289-1:2014. (2014). Document management applications. https://www.iso.org/standard/64599.html

Jembu Rajkumar, A., Lazar, J., Jordan, J., Darvishy, A., & Hutter, H.-P. (2020). PDF accessibility of research papers: What tools are needed for assessment and remediation? In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 4185–4194). Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2020.512

Nganji, J. T. (2018). An assessment of the accessibility of PDF versions of selected journal articles published in a WCAG 2.0 era (2014–2018). *Learned Publishing, 31*, 391–401. https://doi.org/10.1002/leap.1197

Schmitt-Koopmann, F., Huang, E., & Darvishy, A. (2022). Accessible PDFs: Applying artificial intelligence for automated remediation of STEM PDFs. In J. Froehlich, K. Shinohara, & S. Ludi (Eds.), *ASSETS '22 proceedings: 24th international ACM SIGACCESS conference on computers and accessibility* (p. 90). ACM. https://doi.org/10.1145/3517428.3550407

Schmitt-Koopmann, F., Huang, E., Hutter, H.-P., Stadelmann, T., & Darvishy, A. (2022). FormulaNet: A benchmark dataset for mathematical formula detection. *IEEE Access, 10*, 91588–91596. https://doi.org/10.1109/ACCESS.2022.3202639

UN General Assembly. (2007, A/RES/61/106). Convention on the Rights of Persons with Disabilities: Resolution adopted by the General Assembly. https://www.refworld.org/docid/45f973632.html

Wang, L. L., Cachola, I., Bragg, J., Cheng, E., Haupt, C. H., Latzke, M., Kuehl, B., van Zuylen, M., Wagner, L. M., & Weld, D. S. (2021). Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users. *ArXiv*, abs/2105.00076.

Web Content Accessibility Guidelines. (2023). WCAG 2 Overview. https://www.w3.org/WAI/standards-guidelines/wcag/

APPENDIX

A.1. INTERVIEW QUESTIONS

1. Could you please briefly introduce yourself and your role in the repository?

Knowledge and Opinions

2. What do you know about the topic of accessible PDF documents?
3. In your opinion, what is an accessible PDF document?
4. How would you rate the accessibility of the publications in your repository?
   a. Do you keep internal statistic on the accessibility of accessible publications?

Priority and Awareness

5. With what priority do you address the issue of accessibility in your repository?
6. How are the employees at the repository sensitized on the topic of 'accessibility'?
   a. Are there trainings or conferences on the topic of 'accessibility' for the employees?
   b. Is attendance in such trainings or conferences mandatory?
7. Is there any exchange between the repositories of universities in Switzerland – through associations, conferences, *etc.*)?
   If yes...
   a. To what extent is 'accessibility' addressed in the discussions?

Accessibility Measures

8. What are specific measures underway in your repository to keep it as accessible as possible (in terms of accessibility of publications)?
   a. Do you have internal guidelines to ensure accessibility of published publications (mostly in PDF format)?
9. Do you check the accessibility of new incoming documents before publishing them on the repository?
   If yes...
   a. How do you check the accessibility of the documents (manually, *etc.*)?
   b. Since when do you check the accessibility of the documents?
10. How do you handle it when, if for example, a visually impaired person requests an accessible version of a document?
    a. Do you offer a service of providing alternative document formats?

Wishes and Future Plans

11. In your opinion, what fundamental change is needed at universities and research institutions for science in general to become more accessible?
12. What do you intend to do in the future to improve accessibility in your repository?

APPENDIX B

B.1. REPOSITORY'S ANSWERS TO QUANTIFIABLE INTERVIEW QUESTIONS

Repository number*	Question 4a	Question 6a	Question 6b	Question 7	Question 9, 9a, 9b	Question 10a
1	No	Yes, internally	No	Yes	No	Yes
2	No	No	No	Yes	No	No
3	No	No	No	Yes	No	No
4	No	No	No	Yes	No	No
5	No	Yes	No	No	No	No
6	No	No	No	Yes	No	No
7	No	No	No	No	No	No
8	No	No	No	Yes	No	No
9	No	No	No	Yes	No	No

*Repositories 1 through 5 correspond to the ones from the quantitative analysis (as shown in Table 1). Repositories 6 through 9 likewise participated in the interviews, but their PDF contents were not analysed concerning accessibility.