

# Comments on Accessibility of Google Print and Google's Library Project

Jim Fruchterman  
Benetech  
February 27, 2005

## Overview

Google's recent announcement of massive library digitization partnerships has generated a huge amount of interest and angst in the print disabled community, and brought focus on the Google Print program. This short white paper aims to illuminate the issues and set the stage for future discussions with Google. Google has not approved this paper, although we hope they will use it as a tool in advancing accessibility.

There are two accessibility issues involved in Google Print. The first is the accessibility of Google Print page images served to Google users. At present, users who cannot read text visually are unable to read these page images, because they are presented as a picture of text without the associated digital text, which is what is needed for the speech and Braille accessibility products to provide the text of the page to a disabled user. The second is the incredible thirst of people with print disabilities for access to the books they cannot read at the library. The prospect of a huge, fundamentally accessible library is tantalizing. It's accessible because they know that Google will have the full book scanned in digital form.

Both of these issues can be addressed by Google, but it will take effort and the explicit cooperation of the publishers, authors and libraries providing the materials to Google. This requirement makes this a significant challenge.

## Google Print Page Image Accessibility

It's important for responsible organizations to ensure that their web pages are accessible. There are both legal requirements and best practices in web coding. This paper doesn't discuss whether or not Google has a legal obligation to make these pages accessible. It simply assumes that it's the right thing to do.

There is a great deal of information about how to make web pages accessible. The [Web Accessibility Initiative of the W3C](#) comes to mind, as well as the U.S. federal government [Section 508 regulations](#). But, the issue here is not particularly complicated. Google has always been known for simple, usable design, which is fundamentally accessible. The issue is how to deliver the text associated with the image to disabled people.

In general, the solution to image accessibility is to provide an [Alt attribute on each image](#). The alt-text is a short piece of text that describes the image. For example, if you go to [www.bookshare.org](http://www.bookshare.org) and put the mouse over the "News Update" image

[http://www.bookshare.org/web/images/feature\\_newsupdate.gif](http://www.bookshare.org/web/images/feature_newsupdate.gif), there will be associated text that says “News Update, Stay Up to Date”, which is the text information presented as part of the image. This works well for short snippets of text, but is not typically seen for really long passages of text (because people generally don’t provide significant amounts of text in image form). Most implementations of Alt-text have a 255 character limit, which means it won’t work for the typical page of text from a book.

Longer text can be provided through the LONGDESC attribute on the image, which points to another file with the accessible text. So, the LONGDESC attribute is simply the URL for the longer text description. This should be an XML file using the DAISY/NISO/NIMAS standard discussed later in this paper, which in this case will look like a standard HTML page of text plus a couple of helpful tags. Since content owners don’t want the entire book to be available to web users in general, it is likely that the LONGDESC link would point to a single page of text which is equivalent to the page image being displayed. LONGDESC and alt-text are supported by the main screen readers used by visually impaired people (JAWS and Window-Eyes), which means that a blind person will be able to access this information when Google serves a page image.

So, let’s list the four options for making Google Print page images accessible:

1. Alt-text always on Google, with link in LONGDESC
  - Pros: Always accessible  
Follow existing standards
  - Cons: More data sent with each page (less than one percent, because the image is bigger)  
Content owners may object
2. Alt-text/LONGDESC by preference (default off) on Google
  - Pros: Accessible, but requires user to find preferences page  
Follow existing standards
  - Cons: Content owners may object
3. Links to accessible text of the page on Google as a separate web page
  - Pros: Easier to find for novice users
  - Cons: Takes up real estate  
Doesn’t follow standards (alt-text, LONGDESC or deprecation of separate accessible versions of web pages)  
Content owners may object
4. Accessible versions somewhere else (probably with authentication of disability)
  - Pros: Only provides text to people with proven disability  
Content owners will object less
  - Cons: Isn’t really accessible: 95+% of people with print disabilities will not get certified

Content owners must be a part of solving this problem. Google has signed contracts with publishers and libraries, the details of which are confidential. It is safe to say that these contracts commit Google to actively protecting the content so that Google doesn’t

become a vehicle for extracting the full content of the book. Also, these contracts probably prevent Google from making the text available.

Dr. Andrew Burt, a professor of computer science and Vice President of Science Fiction and Fantasy Writers of America, Inc., has demonstrated how to do this with Amazon's scanned book service and is concerned that Google makes the same thing possible. So, we're assuming that Google is addressing this issue and has a separate obligation to prevent access to the full content of the book. Still, whatever portions of books that are presented visually to sighted users should be made available in text for people with disabilities.

Accessibility opens up the possibility of providing access to the text of pages, rather than just the images. But, a sighted person can read the page image being presented and do a screen grab and print if they want a permanent copy of it. The technology to automatically create text versions from the page images already exists and works well (standard optical character recognition). It could be argued that shutting off text access mainly hurts people with print disabilities without benefiting the content owners.

Accessibility is one of those issues where public policy interests conflict with private interests, although thankfully the conflict should be minor because most private interest holders are willing to support access. Still, Google is in the middle on this one. It's a very similar issue to that covered in the paper I wrote with George Kerscher in 2002 called [The Soundproof Book](#), about the locking up of the first generation of ebooks.

Rather than make an extended analysis at this stage, I would simply recommend that Google choose option 1, and engage the content owners to make this possible. I think that accessibility advocates would make the most headway by convincing a major publisher to permit this, and then engage Google to trial this. Alt-text/LONGDESC has the benefit of being standard, supported by the adaptive technology in use by people with print disabilities, adds modest overhead to an image and doesn't materially expand the potential for abuse. I would look for feedback from users about the usability of this solution before implementing.

### **Google Print Book Access**

The question here is about making entire books available to individuals with print disabilities. The rights conflict issue mentioned above is still acute. Organizations such as the National Library Service for the Blind and Physically Handicapped of the Library of Congress, Recording for the Blind & Dyslexic, the American Printing House for the Blind and Bookshare.org (a project of the author's Benetech technology nonprofit) have the right to provide just about any book in the United States to print disabled people without needing to ask for permission or pay a royalty. This right is written right into federal copyright law in [Section 121](#), where [Section 106](#) is the grant of the right to control copying to copyright holders (subject to the limitations in Sections 107 through 122). Bookshare.org is the national library that concentrates on scanned books (as opposed to human recordings of books). So, if Google hadn't executed a contract that

prohibits providing of the complete scanned books in Google Print to third parties, it would have been legal to provide these books to Bookshare.org and its peers. But, they almost certainly did agree to this restriction, because the content owners wouldn't have agreed to anything without it. So, the solution is for Google (or Bookshare.org, or a combination of groups authorized under Section 121) to get the content owners to make an exception for people with print disabilities.

Such an exception would need to be drawn narrowly enough to protect the legitimate interests of content owners while providing access to people with print disabilities. Section 121 is the ideal model for this exception, since it was enacted in the last ten years with the joint efforts of the disability activists and the publishing industry. The key provisions are:

- Only authorized entities can make these accessible books available: this is limited to nonprofit organizations or government agencies with a primary mission of meeting the information access needs of the blind and other people with print disabilities.
- People with print disabilities are a limited group, representing a few percent of the population at most, with bona fide disabilities that affect print reading, such as vision impairment/blindness, significant dyslexia or a physical disability that affects reading. The groups that serve this community require a formal proof of disability before providing accessible books.
- The books provided have to be in specialized formats, which means “braille, audio, or digital text which is exclusively for use by blind or other persons with disabilities”

Together these provisions provide the safeguards that content owners need, to have comfort that their content is going to people with print disabilities and not being distributed to the mass market. The fact that the authorized entities are a limited set also provides publishers and authors with a point of contact. For example, we at Bookshare.org are in monthly contact with publishers, authors and their industry associations, and we are actively committed to upholding the social bargain in Section 121 by fighting abuses. The situation is similar with the other authorized entities, which have been labeled “trusted intermediaries.”

In today's world, issues around intellectual property are not just limited to the United States. Laws similar to Section 121 are in place in other countries in the world (such as Canada and the UK), and the European Union has a directive to its members to pass legislation in this area. Trusted intermediaries exist in most countries.

Microsoft recently hosted the heads of the libraries for people with print disabilities at a [summit](#), where the challenges of building a global library were discussed. The number one issue was reducing the duplication of effort in making the same book accessible in dozens of countries. Groups serving the disabled have relatively small resources, and reducing duplication would stretch those resources much further

Another issue is the relative importance of different titles. People with print disabilities are similar in their reading needs to people without disabilities. They want to read:

- The latest bestsellers, to be an active part of their peer group
- Textbooks and other educationally needed books, to ensure educational opportunity
- Vocationally relevant titles, to ensure employment opportunities including advancement
- Religious titles

For the great majority of people with print disabilities, the books in demand are current copyrighted titles. Public domain titles are useful, but represent a tiny fraction of what people want to read. The incremental value of adding the full content of public domain titles beyond the first few thousand core literature titles is relatively low, especially if the texts are already searchable and accessible as part of Google Print. It's worth doing in the interest of full inclusion. But, the core priority is access to the most commercially valuable titles, which are the most challenging in terms of permissions!

On the technical side, the direction for ebooks for people with print disabilities is towards two closely related XML standards: the [DAISY 3/ ANSI/NISO Z39.86](#) standard, and the [National Instructional Materials Accessibility Standard](#) (a subset of DAISY/NISO adopted by the U.S. Department of Education for textbook accessibility). These standards both support full text ebooks (such as scanned books). DAISY/NISO also supports digital audio versions of books (by human narrators and/or by synthetic speech). The organizations in the field have already created extensive tools for turning books in different formats into these XML formats for people with print disabilities. In addition, the authorized entities implement digital rights management solutions to restrict access to people with bona fide disabilities. It is likely that Google's internal text and image formats for these scanned books can be turned into this XML standard through inexpensive automated processes. The images may need to be put through the optical character recognition process again if the quality of OCR could be improved: the disabled community has higher quality requirements for text than raw OCR often provides. For example, most books submitted to Bookshare.org have been proofread after OCR.

The recommendation to Google is to make these scanned books available to authorized entities through the following actions:

- Develop a standard, short permission form/letter with the following core provisions:
  - Google may provide scanned books to authorized entities for redistribution only to people with bona fide disabilities
  - Authorized entities can also share the scanned books with other authorized entities, including international trusted entities that are the equivalent of U.S. authorized entities
- Add this permission letter to the agenda of future publisher and library discussions
- Work with authorized entities to get existing publishers and libraries to sign the letter (those groups already with a contract with Google)

- Sponsor these efforts to make the books more accessible, with a focus on current copyrighted materials

A commitment in this direction would ensure that those who most need access to electronic books, people with print disabilities, would benefit from the vast vision that Google has for Google Print to expand the accessibility of books for everyone.