Hello. I am writing with regards to the Copyright Office's study of mass digitization, specifically, the concern that any prospective legislation will ignore the role of individuals not affiliated with large organizations such as universities and libraries. I speak as someone who has both digitized books for my own private use and as a witness/benefactor of other people's efforts to digitize and distribute certain books and other media online.

Arguably, individuals engaging in unauthorized digitization have been the primary source of media preservation in the last 20 years. They have been scanning, ripping, transcoding, and distributing media on their own, for several reasons:

- First, in many cases, there are literally no legal versions available for some media. For example, *Big Guy and Rusty the Boy Robot*, a cartoon that aired on Fox from 1999 to 2001, has never been released on DVD, as a digital download, or on streaming services. The only way to see it is to download video files from piracy sites.
- Second, media going out of print with no apparent hope for reprinting or rerelease in digital formats. For example, I obtained *Project Orion: The True Story of the Atomic Spaceship* for $0.25 at a college library book sale two to four years ago. Because it is out of print, new copies have skyrocketed in price to $95.90 for paperbacks, while used copies cost at least $30.71. Since twelve years have passed since it was originally published, there is little hope of a reprint or legal eBook release. I discovered that someone has already released a digitized version of this book because of this.
- Third, publishing companies digitize things only when it is financially expedient for them. For example, Simon & Schuster has digitized practically all of the *Star Trek* novels and reference books published to date, because *Star Trek* is a brand that will make them money in direct sales to consumers and in licensing deals with companies that provide eBooks to libraries. In a more modern example, Insight Editions may **not** publish a digital version of the *XCOM2* novelization[1] unless sufficient interest is established. If there is not enough interest, people who would want a digital edition would be forced to make one themselves or download one off the internet.
- Fourth, individuals as a whole have a broader set of interests and access to a more diverse catalog of titles than formal organizations. This is because formal organizations like universities and libraries have budget and space limitations that limit their ability to possess media, not to mention the hardware and personnel required to digitize it. While individuals have technological and budget limitations of their own, the sheer number of them willing to preserve media allows them to overcome those limitations.
- Fifth, there are practically no benefits to asking for permission to digitize from publishers. At best, they will deny permission or send a cease and desist. At worst, they will either sue for exorbitant amounts of money or charge exorbitant amounts of money for the rights to digitize something, possibly without access to the original materials for

[1] http://www.amazon.com/XCOM-Official-Novelization-Greg-Keyes/dp/1608877124/

things like illustrations and photographs, leading to a lower quality digitization. These limitations extend to formal organizations like universities and libraries, which are bound by law and cannot sidestep these limitations at all, unlike an individual.

- Sixth, there are no or few pathways to legality for individuals who want to preserve media. The groups that focus on preserving media digitally tend to be small to lower costs and preserve overhead for obtaining media rights, or focus on public domain works that are of less interest to individual preservationists. That means that individuals who wish to become legal preservationists and distribute digitized works have to compete in a tiny job market or engage in piracy to do so.

Note: this is merely a summary of some of the systemic issues that have pushed piracy into the leadership position of mass digitization, and not a comprehensive examination of why piracy exists.

Another factor that has pushed the individual into the forefront of digital preservation is technology. Whole suites of hardware and software allow people to record TV programming, audio, video streams, and digitize books. Focusing specifically on books, a particularly potent combination of hardware and software is making book digitization more practical for the individual. The integration of scanners into computer mice, such as the LG LSM-100[2], make it easier than ever to produce high quality scans of a book's text without damaging the book. Before this, books would often have to taken apart for scanning on flatbed scanners, destroying the binding in the process. The only weakness of this technology is multi-page image spreads, where one image spans two pages; the binding produces massive image distortion that can only be dealt with by destroying the binding to scan the pages individually or by sinking large quantities of time into photoshopping the image to remove the distortion.

Technology has also made preserving books in a wider variety of formats easier. While TXT and PDF files were the most common formats for years, the spread of What You See is What You Get (WYSIWYG) HTML editors and the EPUB format have made it easier to make accurate digital copies of books. Where PDF was once the leader in accurate formatting, EPUB can just as easily reproduce the important formatting (italics and other text effects), while giving users the ability to scale text to fit their device's display. It is now easier than ever for the average person to produce a high quality digital copy of a book. It merely requires a large investment of time. In my experience, three or four chapters of a paperback novel can be scanned and converted to HTML in about four to five hours, depending on the amount of pages and the amount of text on each page. The entire book, depending the amount of chapters, if there are illustrations, the amount of footnotes, etc..., can be digitized in a few days to a few weeks.

As a result, there is an untapped resource pool that can provide much needed labor and access to materials worth preservation. However, the "extended collective license" concept does

[2] http://www.newegg.com/Product/Product.aspx?Item=9SIA0ZX0TN5396

not seem to be designed to take advantage of the wide range of individuals who would be willing to help larger institutions preserve our cultural history. While the initial pilot program is limited only to research and educational materials, much of the wording implies that only large organizations such as universities and libraries will be involved. This is wholly impractical, for a variety of reasons, including the fact that those organizations are limited in their resources, especially if they have to contribute money to a piece of media's ECL. Money is a finite resource for all organizations; if money is being spent on ECLs, then there is less money to employ people to digitize the materials these organizations already own.

Therefore, there should be some third party organization, not affiliated with the copyright holders, nor the organizations seeking the rights to orphaned/out of print works, that is empowered to collect and catalog digitized media until all copyright issues have been resolved. This organization should, at the very least, collect anonymous submissions of digitized content from individual users that is either uploaded to a database or shipped on a piece of physical media like a flash drive or DVD-ROM. Anonymous submission is a vital necessity, because even if there are legal protections for people who submit materials to this organization, they may have pirated the material in order to obtain it, making them targets for litigation by copyright holders. This would allow private citizens to meaningfully contribute to and speed up the mass digitization and preservation of out of print works, which will likely take decades. However, publishers should at least be required to contribute higher quality illustrations and pictures for all works that feature illustrations and pictures. This is due to the aforementioned difficulties in scanning multi-page illustrations, as well as the fact that some books feature low quality prints of those pictures and illustrations; scanning low quality images means that the books will result in equally low quality reproductions of those images.