**How far we plan to preserve: Do current Digital Preservation research suffices?**

Sajana, C ,. Krishnamurthy, M\*

*Research Scholar, Bharathiar University, Coimbatore,* [*sajana.drtc@hotmail.com*](mailto:sajana.drtc@hotmail.com)

*\*Assistant Professor, DRTC, Indian Statistical Institute Bangalore,* [*mkrishna\_murthy@hotmail.com*](mailto:mkrishna_murthy@hotmail.com)

##### **Abstract**

##### ***Purpose***

*Digital content is a strategic asset that has observed an exponential growth in recent years. Like the physical universe, the digital universe is large. It is doubling in size every two years, and by 2020 the digital universe i.e. all the data we create and copy is tend reach 44 trillion gigabytes. Digital data is useful if it is suitably organized, tagged and preserved. So, to better understand current research trends in Digital Preservation and create awareness among scientists and policy makers the urgent need to explore more possibilities in preserving digital content, Bibliometric techniques are applied for quantitatively evaluating the related literature in Digital Preservation research.*

##### ***Methodology/ Approach***

*Using data from the IEEE database between 2001 and 2015, a total of 139 publications were retrieved by searching keyword- “Digital Preservation”. The records were analysed for research trends in Digital Preservation.*

##### ***Conclusions/ Findings***

*The preliminary results show that the yearly publications on Digital Preservation have increased exponentially, specifically peak in publications was observed in years 2010 and 2014. Other findings are related to patterns of International research collaboration, research interests, keyword frequency/subject interests to researchers, high impact articles etc.*

##### ***Novelty/Significance***

*As observed by authors, this article signifies the vibrant growth of research in Digital Preservation. It is a priority research area that need to be explored more, as mankind in this modern world is dependent on digital technology. Also we believe there seems to be minimal work in Bibliometric mapping of publications related to Digital Preservation. Hence, this article can be considered as a fundamental work within this field of study.*

**Keywords**

Scientometrics, Digital Preservation, Bibliometrics, Research output, Digital Repositories

**Introduction**

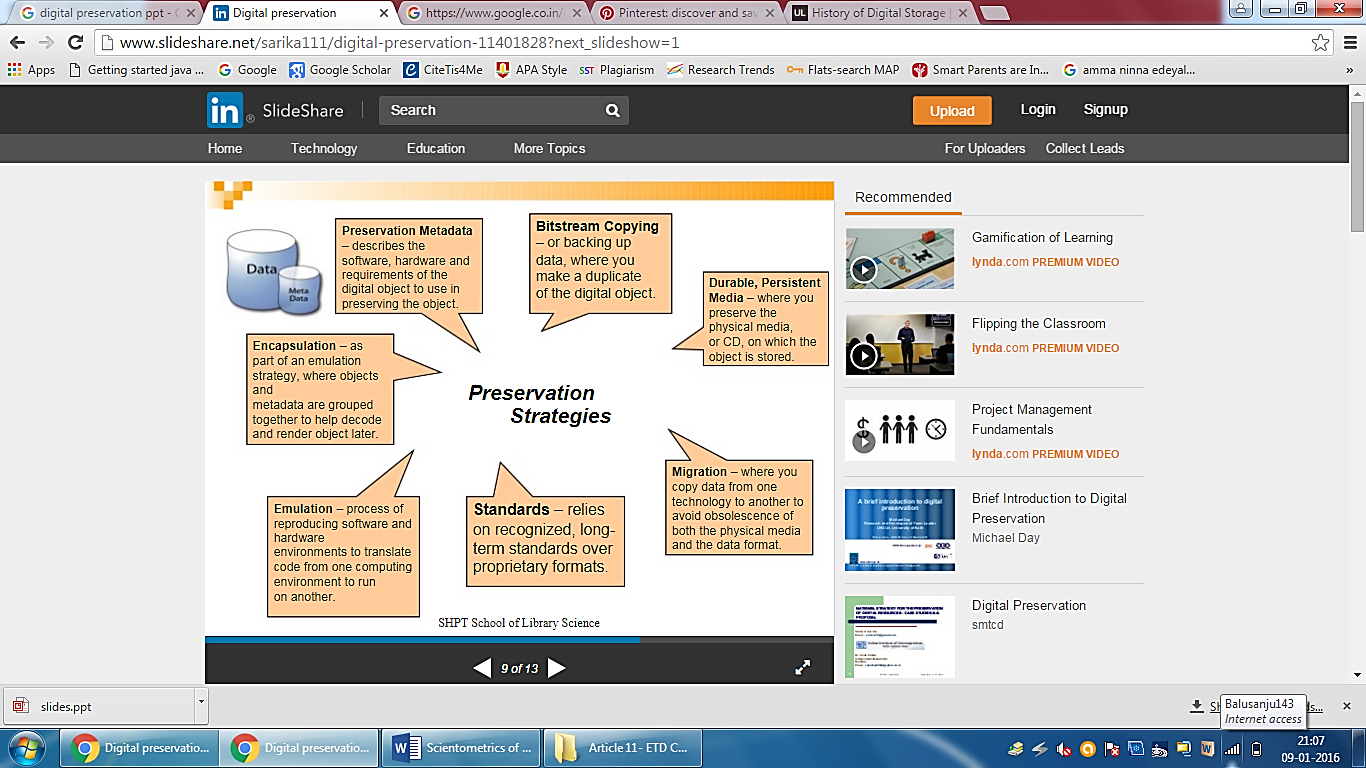
Digital objects are bound to deteriorate over long term, which can be rightly prevented by converting them to digital form. Digital objects are preserved in order to provide digital access, long term preservation and intellectual integrity. In early 1970s, Electronic Arts Intermix promoted preservation of audio and video recordings. It was in 1990’s major developments in digital preservation took off to greater heights with projects like Yale’s Open Book project, University Libraries’ Digital Preservation project and the digital preservation consortium formed in 1994.

We need new storage technologies, standards, and practices to preserve modern cinema. The chemically processed film will last for 100 years or longer. In 2007, the Science and Technology Council of the Academy of Motion Picture Arts and Sciences estimated that the annual cost of preserving an 8.3-terabyte digital master is about US $12 000—more than 10 times what it costs to preserve a traditional film master15.

According American Library Association (ALA) 3, digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.

**Digital Preservation Strategies**

Scientific and heritage information resources are digitally generated, stored, maintained, and accessed. However the risk of losing the resources over the long term is very high. Since the hardware and software technology always continues to evolve rapidly, the digital media seems to be quite short lived.



**Figure 1. Digital Preservation Strategies (Sawant, S., 2012).**

According to ALA, Digital preservation strategies state about:

* **Content creation** that includes- Production of reliable master files with sufficient descriptive, administrative and structural metadata to ensure future access and detailed quality control of processes
* **Content integrity** that includes: Use of persistent/ permanent identifiers, recorded provenance and change history for all objects
* **Content maintenance** that includes: Strong computing and networking infrastructure for Storage and synchronization of files at multiple sites

**Need for Digital preservation**

The aim of digital preservation is long-term, error-free storage of digital information, with the means of retrieval and interpretation, for the period of time that information is required. Digital

Preservation is needed to ensure that future scholars will be able to access and build upon today’s research and science. According to Kirchhoff (2008)2, the main objectives of digital preservation are:

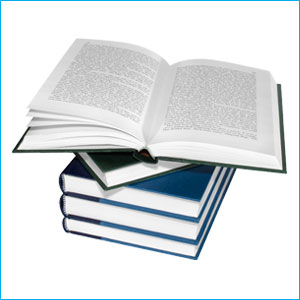
* *Usability* – the intellectual content of the item must remain usable via the delivery mechanism of current technology;
* *Authenticity* – the provenance of the content must be proven and the content an authentic replica of the original;
* *Discoverability* – the content must have logical bibliographic metadata so that the content can be found by end-users through time; and
* *Accessibility* – the content must be available for use to the appropriate community.

**Resources to be Digitally Preserved**

Preserving the resources from “Digital obsolescence” is the need of the hour. Some array of resources that need to be preserved include (UKOLN, 2008):

* ***Digitally Repackaged***: Digitized resources that are substitutes of print forms.

For Eg.



* ***Born Digital***: Resources that originate in digital form.

For Eg.



* ***Individual resources***: Texts, still and moving images, sound recordings, etc.

For Eg.



* ***Collective resources***: Web sites, e-journals, wikis, catalogues, etc.

For Eg.



* ***Data Sets***: Scientific and cultural data comprising multiple individual pieces of data.

For Eg.



* ***Communication records***: For example, email, instant messages, etc.

For Eg.



**Methodology**

We retrieved research publications from *IEEE database* between years 2001 to 2015. A total of 139 records were retrieved by searching keyword- “Digital Preservation”. The retrieved document types included journal articles and conference proceedings. The data was analyzed using spreadsheet software to identify the research trends in the field of digital preservation. The visualizations were created using *Lumira*.

**Results and findings**

*Year of publication*

Figure 1 shows research output each year from 2001 to 2015. During this 14-year period, the overall trend has been a steady increase of yearly output, though the number of records in the year 2015 decreased slightly. It is also observed that there is a sudden steep increase in years 2010 and 2014. In support of this data researchers observed some events in these years like- in 2010, The Library of Congress (LOC) accumulated 170 billion tweets during 2006- 2010 totaling 133.2 terabytes6 and Collaborative Efforts Expand: NDSA, LOCKSS, Hathi Trust. In 2014, it became official: all of the Big Five publishers now offer their full e-book catalogues to libraries. But, there is still no workable plan for digital preservation of works, a challenge that is growing more urgent as more works are being published—and self-published—in digital-only formats. As Columbia University’s Robert Wolven told *PW* earlier this year, digital preservation is the “global warming” issue for libraries. “Everyone knows that if we don’t do something now,” he says, “we’ll be in big trouble later”7. Results also revealed that the concept of Digital Preservation was consistently the focus of researchers and has developed at an increasingly rapid rate since past few years.

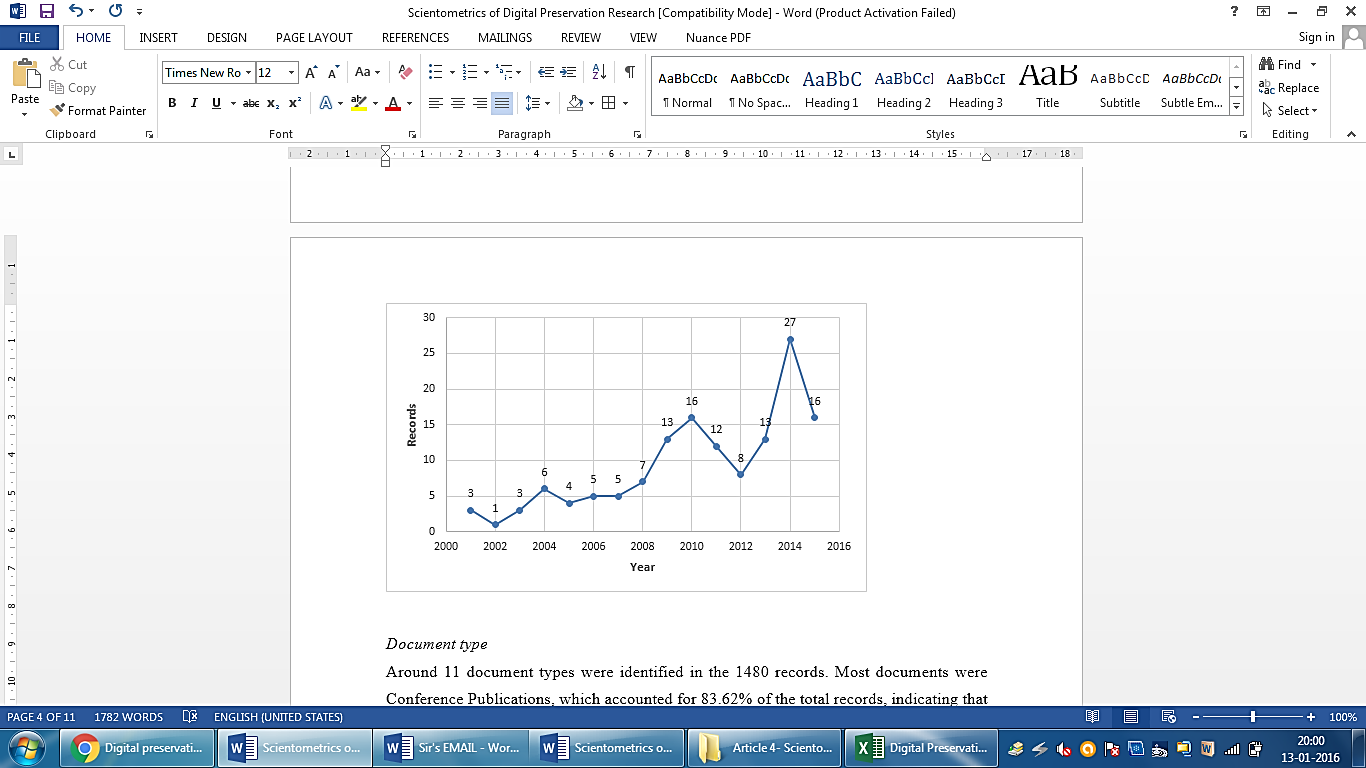


Figure 1. Year of publication

*Document type*

Around 5 document types were identified in the 139 records. Table 1. indicates that most documents were IEEE Conference Publications, which accounted for 87.05% of the total records, indicating that Conferences are the main approach for scientific communication among researchers in Digital Preservation domain. Researchers in this study opine that the conferences are one of the vital verbal channel in any scholarly scientific communication process. Journals and Magazines were the other important modes to publish academic achievements in this field of research.

|  |  |
| --- | --- |
| **Document Type** | **Records** |
| IEEE Conference Publications | 121 |
| IEEE Journals & Magazines | 8 |
| SMPTE Conference Publications | 5 |
| SMPTE Journals & Magazines | 4 |
| IBM Journals & Magazines | 1 |

Table 1. Document Type

*Co-authorship pattern*

Scientific collaboration – as measured by means of co-authorship patterns – has considerably increased during the last decades at all levels of aggregation7. With the increase in communication network, scientific progress and interdisciplinary research the collaboration trends has seen a radical change. As represented in Table 2, Collaboration intensity was high in three and more authors. Perhaps, the authorship linkage decreased relatively in dual authorship. Thus, Digital Preservation research and development can be considered as a collaborative research efforts of a team.

|  |  |
| --- | --- |
| **Number of Author(s)** | **Collaboration intensity** |
| 1 | 21 |
| 2 | 16 |
| 3 | 37 |
| More than 3 Authors | 65 |

Table 2. Author collaboration

*Country based distribution of research output*

Analysis of the country-based research in Digital Preservation, help us to recognize the scholarly research output of a country and explore the differences among them. Approximately more than 90 countries, contributed to the research output of digital preservation, indicating that this topic as an important concept that has attracted world- wide attention. From Figure No. 2 it can be seen that the top five countries were USA, China, USA, Japan, United Kingdom and France. Since contemporarily it is a technological era, perhaps, it seems that India (in 9th position) need to keep pace towards research in this field in order to explore the future needs of digital preservation.

Nonetheless, as a part of digital preservation initiative, Department of Electronics & Information Technology, Government of India (DeitY) has taken initial steps towards envisioning a National Digital Preservation Program. The steps taken in actualizing this vision include “National Study Report on Digital Preservation Requirements of India” and setting-up of “Centre of Excellence for Digital Preservation” at C-DAC, Pune and C-DAC, Noida. Also, the National Digital Preservation Programme (NDPP) of India was launched by Ministry of Communications & Information Technology in 20089.

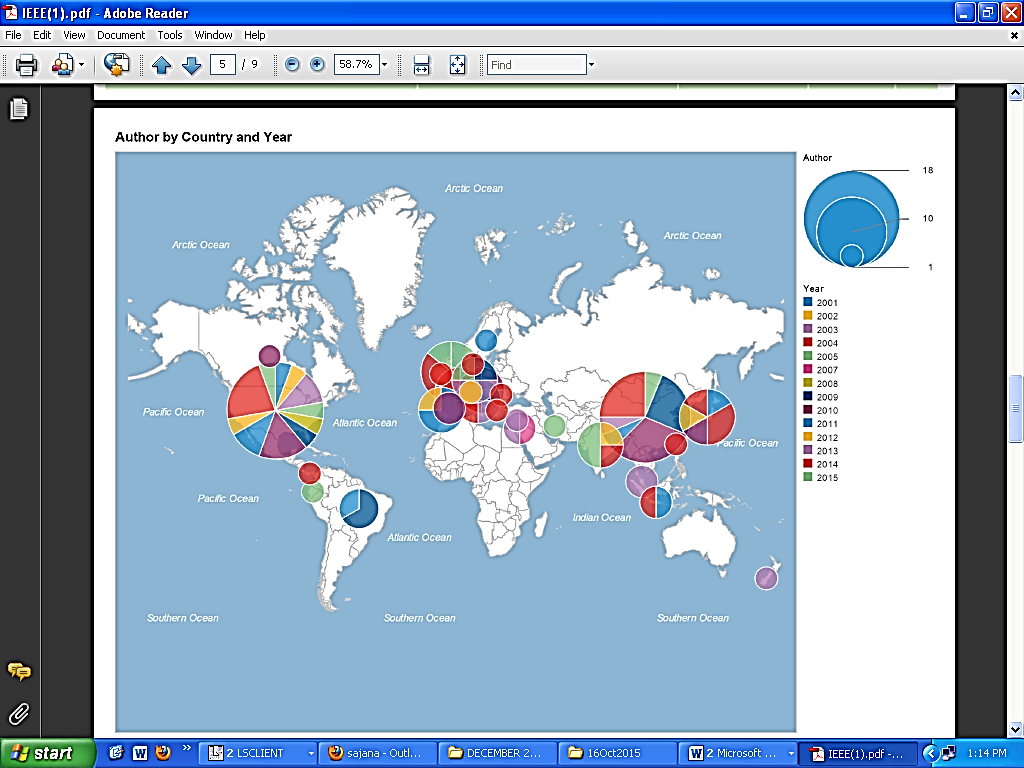


Figure 2. Country based research output

*Keyword Analysis*

Results of keywords analyzed from the data is presented in Table No. 3, which indicates the term “Digital Preservation” is the most extensively used keyword. It is followed by OAIS, cloud computing, Cultural Heritage, Digital Archives, Metadata, Web archiving, Content Server.

|  |  |
| --- | --- |
| **Keyword** | **Count** |
| Digital preservation | 44 |
| OAIS | 6 |
| Cloud computing | 5 |
| Cultural Heritage | 5 |
| Digital Archives | 4 |
| Metadata | 4 |
| Web archiving | 4 |
| Content Server | 3 |

Table 3. Keywords

*High-impact articles and authors*

High-impact articles were selected based on citation count. From Table 4 it can be seen that the top ten high-impact articles were written by 21 authors, of which Su-Shing Chen has authored the High impact article. The highest impact articles on Digital Preservation were published in two journals and five different conference proceedings. It is evident from the table that researchers from field of Computer Science, Electronics, Pattern Recognition, Geoscience and Remote Sensing are focusing on Digital Preservation research.

Not surprisingly, researcher Silva, L. can be considered one among leading scholars in the field of Digital Preservation research.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author** | **Title** | **Journal/Conference** | **Year** | **Citations** |
| Su-Shing Chen | The paradox of digital preservation | Computer | 2001 | 6 |
| Chao Li; Chunxiao Xing; Huang, M.B.; Li Dong; Mi Zhang | A Practice of Web-Based Sustainable Preservation Risk Management System for Digital Libraries in China | Web Information Systems and Applications Conference (WISA), 2010 7th | 2010 | 5 |
| Katsura, S.; Yamanouchi, W.; Yokokura, Y. | Real-World Haptics: Reproduction of Human Motion | Industrial Electronics Magazine, IEEE | 2012 | 5 |
| Pasko, G.; Pasko, A.; Vilbrandt, C.; Ikedo, T. | Virtual Shikki and Sazaedo: shape modeling in digital preservation of Japanese lacquer ware and temples | Computer Graphics, Spring Conference on, 2001. | 2001 | 4 |
| Nikitenko, D.; Wirth, M.; Trudel, K. | White-balancing algorithms in colour photograph restoration | Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on | 2007 | 4 |
| Vrubel, A.; Bellon, O.R.P.; Silva, L. | A 3D reconstruction pipeline for digital preservation | Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on | 2009 | 4 |
| Rank, R. | Enterprise IT support for NOAA archives | Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International | 2007 | 3 |

Table 4. High-impact articles

*Most Prolific Authors*

The most prolific authors were considered based on the count of records output. Table no. 5 shows that Coughlin, Thomas M. is the top most scholar in Digital Preservation research with 6.95% share of total publication. Junli Zhao, Lake, A., Silva, L., Zhongke Wu, Nguyen, Q.L. are the other prolific authors in this field of research.

|  |  |  |
| --- | --- | --- |
| **Authors** | **Records** | **Share (%)** |
| Coughlin, Thomas M. | 5 | 6.95 |
| Junli Zhao | 3 | 4.17 |
| Lake, A. | 3 | 4.17 |
| Silva, L. | 3 | 4.17 |
| Zhongke Wu | 3 | 4.17 |
| Nguyen, Q.L. | 3 | 4.17 |

Table 5. Prolific authors

**Conclusions**

Digital assets have entered in our daily lives and hence it is a necessity to suitably organize and preserve them for eternity. Digital Preservation strategies in India should be enhanced, since our country has a culturally rich heritage. Research in Digital Preservation should be supported and augmented.

**References**

* 1. UKOLN,. (2008). An introduction to digital preservation. Retrieved 12 January 2016, from <http://www.ukoln.ac.uk/cultural-heritage/documents/briefing-31/briefing-31.doc>
  2. Kirchhoff, A. J. (2008). Digital preservation: challenges and implementation. *Learned Publishing*, *21*(4), 285-294.
  3. Ala.org,. (2009). Preservation Resources | Association for Library Collections & Technical Services (ALCTS). Retrieved 12 January 2016, from <http://www.ala.org/alcts/resources/preserv>
  4. Sawant, S. (2012). Digital preservation. Slideshare.net. Retrieved 12 January 2016, from <http://www.slideshare.net/sarika111/digital-preservation-11401828?next_slideshow=1>
  5. PublishersWeekly.com,. (2016). The Top 10 Library Stories of 2014. Retrieved 14 January 2016, from <http://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/64975-the-top-10-library-stories-of-2014-ala-midwinter-spotlight.html>
  6. Digitalpreservation.gov,. (2016). The Top 10 Digital Preservation Developments of 2010- News Archive - Digital Preservation (Library of Congress). Retrieved 14 January 2016,from <http://www.digitalpreservation.gov/news/2010/20101229news_article_top10stories.html>
  7. PublishersWeekly.com,. (2016). The Top 10 Library Stories of 2014. Retrieved 14 January 2016, from <http://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/64975-the-top-10-library-stories-of-2014-ala-midwinter-spotlight.html>
  8. Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometrics, 60(3), 421-432. <http://dx.doi.org/10.1023/b:scie.0000034384.35498.7d>
  9. Digital meets Culture,. (2016). India for digital preservation. Retrieved 14 January 2016, from <http://www.digitalmeetsculture.net/article/india-for-digital-preservation/>
  10. Spokaneprinting.net,. (2016). Printing in Spokane Print Business Cards Brochures Flyers Banners Custom Apparel T-shirt Screen Printing Company Embroidery in Spokane - The Print Center. Retrieved 14 January 2016, from <http://www.spokaneprinting.net>
  11. NCSU. (2016). Retrieved 14 January 2016, from <http://www.news.lib.ncsu.edu>
  12. Wallpapereast.com,. (2016). Find your favorite wallpapers here at wallpapereast.com. Retrieved 14 January 2016, from <http://www.wallpapereast.com>
  13. Southfieldchristian.wikispaces.com,. (2016). southfieldchristian - home. Retrieved 14 January 2016, from <http://www.southfieldchristian.wikispaces.com>
  14. Namami.org,. (2016). National Mission for Manuscripts. Retrieved 14 January 2016, from <http://www.namami.org>
  15. Maltz, A. (2015). Will Today’s Digital Movies Exist in 100 Years?. Spectrum.ieee.org. Retrieved 14 January 2016, from <http://spectrum.ieee.org/consumer-electronics/standards/will-todays-digital-movies-exist-in-100-years>