



HAL
open science

How far we plan to preserve

M Krishnamurthy, C. Sajana,

► **To cite this version:**

M Krishnamurthy, C. Sajana,. How far we plan to preserve: Do current Digital Preservation research suffices?. 19th International Symposium on Electronic Theses and Dissertations (ETD 2016): "Data and Dissertations", Jul 2016, Villeneuve d'Ascq, France. . hal-01433027

HAL Id: hal-01433027

<https://hal.univ-lille.fr/hal-01433027>

Submitted on 12 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Dr. Krishnamurthy, M and Sajana, C

Abstract

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.

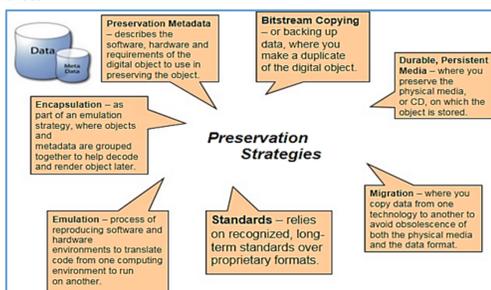
Introduction

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 master.

According American Library Association (ALA), 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.



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 that includes: Strong computing and networking infrastructure for Storage and synchronization of files at multiple sites

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):



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 2010, The Library of Congress (LOC) accumulated 170 billion tweets during 2006- 2010 totaling 133.2 terabytes 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. And, vexingly, there is still no workable plan to preserve digital 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." 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.

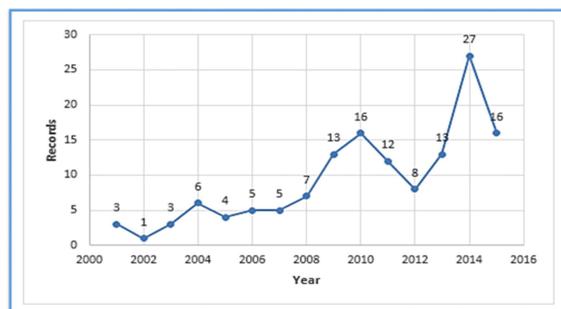
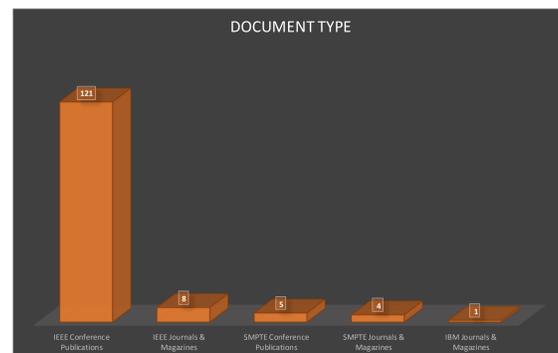


Figure 1. Year of publication

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.



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 Authors	21
2 Authors	16
3 Authors	37
More than 3 Authors	65

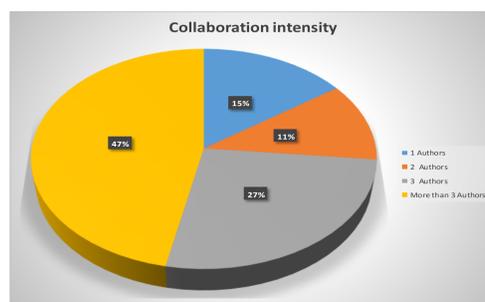


Table 2. Author collaboration

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.

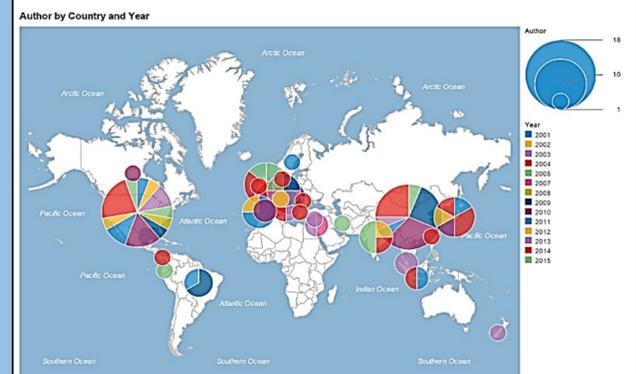


Figure 2. Country based research output

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.

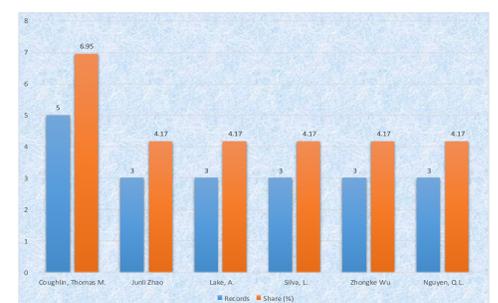
Not surprisingly, researcher Silva, L. can be considered one among leading scholars in the field of Digital Preservation research

Author	Title	Citations
Su-Shing Chen	The paradox of digital preservation	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	5
Katsura, S.; Yamanochi, W.; Yokokura, Y.	Real-World Haptics: Reproduction of Human Motion	5
Pasko, G.; Pasko, A.; Vilbrandt, C.; Ikedo, T.	Virtual Shikki and Sazae: shape modeling in digital preservation of Japanese lacquer ware and temples	4
Nikitenko, D.; Wirth, M.; Trudel, K.	White-balancing algorithms in colour photograph restoration	4
Vrubel, A.; Bellon, O.R.P.; Silva, L.	A 3D reconstruction pipeline for digital preservation	4
Rank, R.	Enterprise IT support for NOAA archives	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.



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.

Same References

- 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>
- Kirchhoff, A. J. (2008). Digital preservation: challenges and implementation. *Learned Publishing*, 21(4), 285-294.
- Ala.org. (2009). Preservation Resources | Association for Library Collections & Technical Services (ALCTS). Retrieved 12 January 2016, from <http://www.ala.org/alcts/resources/preserv>
- Sawant, S. (2012). Digital preservation Slideshare.net. Retrieved 12 January 2016, from http://www.slideshare.net/sarika111/digital-preservation-11401828?next_slideshow=1
- 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>
- NCSLU. (2016). Retrieved 14 January 2016, from <http://www.news.lib.ncsu.edu>
- Wallpaperart.com. (2016). Find your favorite wallpapers here at wallpaperart.com. Retrieved 14 January 2016, from <http://www.wallpaperart.com>
- Southfieldchristian.wikispaces.com. (2016). southfieldchristian - home. Retrieved 14 January 2016, from <http://www.southfieldchristian.wikispaces.com>