



Research Paper

Top web preservation programs and projects: properties and activities

Kianoosh Rashidi¹, Hajar Sotudeh²

1. PhD Student, Knowledge and Information Sciences, School of Educational Sciences and Psychology, University of Shiraz, Shiraz, Iran
2. Associate Professor at Knowledge & Information Science Department, School of Education & Psychology, Shiraz University, Shiraz, Iran.

Article Info.

Received: 2016/09/13

Accepted: 2017/05/14

Abstract

Background and Objectives: Nowadays, we are passing through an era of transition from analog to digital format. Most valuable information is either digitally born or digitized which require digital preservation to ensure their safety and survival for long-term maintenance and access for posterity. Several web preservation programs have been launched around the world, each of which having its own properties and area of activities in line with policies and goals of the user organization. The present study aimed to explore the activities and properties of the existing top web preservation projects and programs in terms of their time coverage, scopes of preservation, and types of resources preserved, access models and authorized users.

Methodology: A documentary method was used to identify and analyze the relevant available literature such as papers, handbooks, web sites, etc. The programs' people-in-charge were also questioned via a short questionnaire sent by Email. Top web preservation programs and projects were identified using Google Search, as well as analyzing the program interfaces and documents, directories and the related literature. After being verified and filtered, 61 top programs were selected to be studied.

Findings: The verification of the launching dates of the programs revealed that "Internet Archive" is the oldest one dating back to 1996. Most recent programs were "Anarchism Web Archive" and "Web Harvesting Project of the German National Library", of which the first was subject specific while the other was that of a specific nationality. While some programs cover a global scope as wide as the web, some others limit their borders to web resources published in a specific country, region, subject,

organization, and/or document type. The first and oldest digital preservation program, i.e. “Internet Archive” has selected to cover the world-wide web as its preservation scope, thus its time coverage goes back to as far as 1996. For some programs, the time coverage is very limited and covers 2-9 years prior to their launching dates; examples are: “The Cyber Cemetery”, “LAC (Electronic Collection of Library and Archives Canada)” and “Portuguese Web Archive”. However, these programs are apparently depending on macro programs such as “Internet Archive” for the web resources published prior to their launching dates. It was also revealed that 50% of these programs run at national level and 13.4 % cover a specific subject. Politics, Culture, Religion, Science, Economy, Slavery, Government, Anarchism, Human Rights, Social Issues, Computer and Information Science are among the subjects that are most frequently dealt with by the programs. Some programs selected only one or two document types while others covered a combination of document types for preservation. Access to the archived version of the preserved documents ranges on a continuum from fully open, through semi-open to restricted access. Of all the programs the majority (39.1%) apply a full open access model; next comes those adhering to a restricted access model (23/9%). The semi-open access model had the least frequency (6.7%). Some programs offer their services to people throughout the world and do not limit themselves to specific users (6.3%) of which a prominent example is the “Internet Archive” that is open to all users around the globe. For some other programs (15.2%), access is restricted just for authorized users; for example, “Web Harvesting Project of the German National Library” and “AOLA (Austrian Online Archive)” are limited to students and researchers.

Discussion: The results of the present study revealed that the importance of web preservation is duly recognized all over the world so that a wide range of countries are found to be engaged in this endeavor. The programs under study can be classified into two main groups including R&D related and operational ones. Most of them are found to have chosen their national domains for preservation; this results in the perseveration of all document types in almost all subjects available in their cyberspaces. There are also many programs found to provide open access to the preserved contents for all kinds of users throughout the world.

Keywords: *Web, Preservation, Programs, Projects*

*Corresponding author:
Email: kia_rsh85@yahoo.com

©2019 Published by Shahid Chamran
University of Ahvaz

How to Cite:

Rashidi, K., Sotudeh, H. (2019). Top web preservation programs and projects: properties and activities. *Journal of Studies in Library and Information Science*, 11(1): 41-56.



برنامه‌های برتر حفاظت وب جهان: ویژگی‌ها و زمینه فعالیت

کیانوش رشیدی^{۱*}، هاجر ستوده^۲

۱. دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه شیراز، واحدین الملل، شیراز، ایران
۲. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه شیراز، شیراز، ایران

اطلاعات مقاله

تاریخ پذیرش: ۹۷/۰۲/۲۴

تاریخ دریافت: ۹۶/۰۶/۲۳

چکیده

هدف: بررسی برنامه‌های حفاظت وب جهان به منظور تعیین زمینه‌ی فعالیت و ویژگی‌های آنها به لحاظ سابقه فعالیت، پوشش مجموعه، گستره پوشش، نوع منابع، الگوی دسترسی و نوع جامعه تحت پوشش فعالیت‌های حفاظت وبی است.

روش‌شناسی: گردآوری داده‌ها به روش متن‌پژوهی انجام شد. برنامه‌های برتر حفاظت وبی از طریق جستجو در گوگل، مستندات، مقالات و راهنماهای معتبرشناسایی شد.

یافته‌ها: بررسی برنامه‌ها نشان داد که شمار برنامه‌های حفاظت در سراسر دنیا رو به افزایش دارد. این امر نشان از اهمیت حفاظت وب و آگاهی روز افزون در این باره دارد. این برنامه‌ها در دو دسته کلی تحقیق و توسعه در حوزه حفاظت وبی و نیز اجرای عملیاتی حفاظت فعالیت دارند.

نتیجه‌گیری: رویه‌های غالبی که در این برنامه‌ها مشاهده شد عبارتند از حفاظت از دامنه ملی و در نتیجه حفاظت از تلفیقی از انواع منابع و (تقریباً) همه موضوعات، و همچنین ارائه الگوی دسترسی آزاد به جامعه کاربری جهانی.

کلیدواژه‌ها: حفاظت دیجیتالی، حفاظت وب، طرح‌ها، برنامه‌ها

* نویسنده مسئول: kia_rsh85@yahoo.com

استناد به این مقاله:

رشیدی، کیانوش؛ ستوده، هاجر؛ (۱۳۹۸). برنامه‌های برتر حفاظت وب جهان: ویژگی‌ها و زمینه فعالیت. مطالعات کتابداری و علم اطلاعات، ۱۱(۱): ۴۱-۵۶

مقدمه و بیان مسئله

عصر حاضر، دوران گذار از قالب‌های قیاسی (آنالوگ) به قالب دیجیتال است. مزیت‌های انتشار دیجیتالی نسبت به منابع چاپی و علاقه‌مندی ناشران و نویسندگان به این نوع انتشار، باعث شده است بسیاری از اطلاعات ارزشمند، فقط در قالب دیجیتالی منتشر شوند. منابع سنتی نیز به طور روزافزونی به قالب دیجیتالی تبدیل می‌شوند. علاوه بر آن، اینترنت امکان خودانتشاری را به وجود آورده است (Gomes, Freitas, Silva, 2006). بنابراین، انتشار دیجیتالی گستره‌ی وسیعی از منابع مانند کتاب‌های الکترونیکی، مجلات الکترونیکی، انتشارات دولتی، عکس‌ها، نقشه‌ها، طرح‌ها، فیلم، موسیقی، تبلیغات، بازی‌ها و... را به خود اختصاص داده است.

گرچه وب مزایایی از قبیل دسترسی آسان و گسترده به اطلاعات و توان ذخیره‌سازی بالایی را فراهم می‌سازد، ولی دسترسی همیشگی به اطلاعات را تضمین نمی‌کند. در واقع، عمر اطلاعات دیجیتالی بسیار کوتاه - از کمتر از ۷۵ روز تا کمتر از یک سال - برآورد شده است (Ntoulas, Cho, Olston, 2006). ناپایداری و میرایی منابع وبی به دلایل مختلف از جمله منسوخ شدن زیرساخت‌های فناوری، قطع حمایت سازمانی، حوادث پیش‌بینی نشده طبیعی و غیرطبیعی، ماهیت پویای محتوای دیجیتالی و تجدید نظر در ساختار وب‌سایت‌ها و پایگاه‌های داده روی می‌دهد (Waller & Sharp, 2006; shadanpoor, 2007). عدم حفاظت از محتوای وب امروزی، نه تنها پژوهش در حوزه‌های موضوعی مختلف مانند محیط زیست، جامعه‌شناسی و حقوق را متأثر خواهد ساخت، بلکه شکافی عمیق در تاریخ بشر برای نسل‌های آینده به جا خواهد گذارد.

حفاظت دیجیتالی، به دسته‌ای از فعالیت‌های نظامند اشاره دارد که برای دسترسی مستمر به منابع دیجیتالی در بلندمدت ضروری هستند. دلایل علاقه‌مندی به حفظ منابع اینترنتی متفاوت است، از علاقه‌های تجاری و وظیفه‌ی قانونی کتابخانه‌ی ملی گرفته تا دیدگاه آرمانگرایانه‌ی جنبش منبع آزاد (Day, 2006). انگیزه‌های مالی، راهبردهای قانونی، قراردادی و کسب شهرت می‌توانند منجر به حفاظت منابع وبی گردند (Ashley et al., 2006). با آنکه مسئله حفاظت درباره منابع سنتی نیز همواره مطرح بوده است، اما تفاوت سرشتی منابع دیجیتالی و منابع سنتی و نیز فناوری‌های به کار گرفته در تولید، نمایش و استفاده از آنها باعث می‌شود حفاظت مواد دیجیتالی و به ویژه وبی از ماهیت بسیار متفاوتی برخوردار باشد. بنابراین، نمی‌توان رویه‌ها، استانداردها و طرح‌های متداول در دنیای سنتی را که برای گزینش، گردآوری، سازماندهی و حفاظت مواد به کار می‌روند در دنیای وب نیز به کار گرفت، هرچند می‌تواند الهام بخش باشد (Emmanuelle, Gildas, 2009).

برنامه‌های حفاظت وب در سراسر دنیا راه‌اندازی شده است. هر برنامه بسته به خط‌مشی و اهداف سازمان مادر، ویژگی‌ها و زمینه فعالیت متفاوتی را منعکس می‌سازد. برخی طرح‌های آرشوگری وب به صورت گزینش‌گری^۱ و برخی در مقیاس بزرگ‌تر و با دامنه‌ی^۲ جامع ملی یا جهانی اجرا می‌شوند. برخی هنوز به صورت برنامه‌ی آزمایشی یا در مراحل ابتدایی هستند، در حالی که برخی دیگر به بهره‌برداری رسیده‌اند. یکی از پیش‌شرط‌های دستیابی به برنامه حفاظت دیجیتالی موفق، انجام مطالعات مقدماتی به روش علمی به منظور سنجش نیازها، کشف موانع و بسترسازی است. در این راستا، آشنایی با تجارب جهانی موفق، ضمن آگاهی‌رسانی درباره آخرین پیشرفت‌های دانش و فناوری در حوزه حفاظت وب، زمینه دستیابی به مدلی برای پیاده‌سازی در ایران را فراهم می‌سازد. بنابراین، پژوهش حاضر بر آن است تا با تمرکز بر برنامه‌های برتر حفاظت وب

۱. Selective

۲. Domain

ویژگی ها و زمینه فعالیت آن ها به لحاظ سابقه فعالیت، پوشش مجموعه، گستره پوشش، نوع منابع، الگوی دسترسی و نوع جامعه تحت پوشش فعالیت های حفاظت و بی را بررسی کند و از این رهگذر چشم اندازی برای راه اندازی برنامه ی ملی در ایران در این زمینه به دست دهد.

مروری بر پژوهش های پیشین

آگاهی روزافزون نسبت به اهمیت راهبردی وب در سراسر جهان به رویکرد گسترده به حفاظت منابع وبی منجر شده است. با این حال، کمتر پژوهشی در این زمینه انجام شده است. آثار این حوزه بیشتر ماهیت نظری و توصیفی دارند و می گویند تا چالش های فرا راه حفاظت منابع وبی و همچنین راهکارها و ملزومات مختلف آن را مورد بحث و بررسی قرار دهند و تجارت و دستاوردهای خود را به اشتراک بگذارند (Murray, Hsieh, 2007; Rauber, Hunter, 2007; Kelly and et al., 2008; Strodl, S., Becker, C., Rauber, 2009). بسته به تجاربی که نویسندگان در امر حفاظت اطلاعات اندوخته اند، تاکید هر اثر بر گروهی خاص از چالش ها است. برای نمونه، برخی بر راهبردها و ابزارهای مجموعه سازی، مسائل مالی در ایجاد آرشیو و ذخیره و نگهداری اطلاعات میرا متمرکز شده اند (Rauber, Hunter, 2007)، برخی به چالش حقوقی (Kelly and et al., 2008)، و برخی به تأثیر فناوری ها (Strodl, S., Becker, C., Rauber, 2009) پرداخته اند. در این میان، معدودی از آثار نیز به بررسی و معرفی برنامه های حفاظت وب (Hodge, G. and Frangakis, 2004) و تجارب کشورهای امریکایی، اروپایی و آسیای شرقی پرداخته اند (Beagrie, 2003; Verheul, 2006; Connertz, 2003).

از معدود پژوهش های انجام شده می توان به پژوهش کتابخانه کنگره اشاره کرد که با همکاری متخصصان حفاظت^۱، کتابداران و آرشیوگران، محققان دانشگاهی و تولید کنندگان اطلاعات انجام شده است و نیازمندی های آرشیو وبی را شناسایی و معرفی کرده است (2005 Library of Congress). (Borbinha, Antunes, Barateiro (2009). در پژوهش خود ضمن تأکید بر نقش حفاظت دیجیتالی در دسترس پذیری بلندمدت، تضمین اعتبار و اصالت اشیای دیجیتالی، به شناسایی خطرات، آسیب ها و فنون حفاظت دیجیتالی پرداخته است. بررسی قوانین حق مؤلف و تجارب حق مؤلف حفاظت دیجیتالی در کشورهای مختلف نیز مورد توجه پژوهشگران بوده است. نتایج نشان می دهد که قوانین حق مؤلف در بسیاری کشورها با هم و با حفاظت و بناسازگار هستند و حتی در مواردی برخوردی خصمانه با منابع وبی و دیجیتالی داشته اند (2006; Charlesworth, Fitzgerald Besek, Coates &).

در ایران نخستین همایش ملی مدیریت منابع اطلاعاتی وب در سال ۱۳۹۱ توسط سازمان اسناد و کتابخانه ملی به هدف ایجاد هم افزایی ملی برای مدیریت بهینه اطلاعات وب و ایجاد بستری مناسب برای راه اندازی آرشیو ملی وب در کشور برگزار شد. در این همایش، در کنار مقالاتی پیرامون سازماندهی، ارزیابی و مدیریت حقوقی منابع وبی، مقالات ارزشمندی درباره آرشیو و حفاظت منابع وبی ارائه شد^۲.

همان گونه که پیشتر ذکر شد و از مرور پژوهش ها نیز بر می آید، متون این حوزه بیش از آن که ماهیت پژوهشی داشته باشند، نظری و توصیفی هستند. این امر شاید به دلیل نوپایی نسبی فعالیت ها و در عین حال پیچیدگی و در این زمینه باشد. همچنین، گذشته از پژوهش هوج و همکاران که به بررسی پروژه ها یا سامانه های حفاظت منابع وبی پرداخته اند و تحقیق هوج و فرانگاکیس (Hodge, G. And Frangakis, 2004)، که راهکارهای برنامه های برتر برای مقابله با چالش های فنی و مدیریتی

۱. curator

۲. <http://conferences.nlai.ir/92/abstract.pdf>

حفاظت وب را بررسی کرده اند، پژوهشی دیگر یافت نشد که به بررسی جامع برنامه‌ها و طرح‌های حفاظت پرداخته باشد. بنابراین، ضرورت پژوهشی تازه که ابعاد گوناگون برنامه‌های برتر حفاظت دیجیتال جهان را روشن سازد احساس می‌شود.

روش‌شناسی پژوهش

برنامه‌های حفاظت وب برتر دنیا، جامعه این پژوهش را تشکیل می‌دهند. منظور از برنامه‌های برتر دنیا برنامه‌هایی است که در فهرست نتایج گوگل در صدر واقع شده‌اند یا در مستندات، مقالات و راهنماهای معتبر به عنوان برنامه‌ای مهم معرفی شده‌اند. نمونه مورد بررسی، به روش هدفمند و به شیوه ذیل شناسایی شد:

واژه‌های program, project یا initiative همراه با عبارات web, "web curation", "web archive", "web archiving", "preservation" در موتور کاوش گوگل مورد جستجو قرار گرفت. در مجموع ۳۰۰۰ نتیجه در صدر فهرست گوگل بررسی شد و پس از حذف پیشینه‌های تکراری، فیلتر شده یا غیرفعال، ۵۰۰ نتیجه‌ی برتر برای شناسایی و استخراج فهرست برنامه‌ها انتخاب گردید. این نتایج متشکل بودند از (۱) رابط خود برنامه‌ها و طرح‌های حفاظت؛ (۲) مستندات برنامه‌ها و طرح‌ها و دیگر متونی که به معرفی آنها پرداخته بودند؛ (۳) راهنماهای موجود مانند پدی^۱ که به معرفی برنامه‌ها می‌پرداختند و (۴) پیوند به برنامه‌های دیگر در یک برنامه. در نتیجه برای هر برنامه، چندین پیشینه یافت شد. نام برنامه‌ها از این متون استخراج گردید و پس از آن برنامه‌های تکراری حذف گردید.

بررسی‌های بیشتر نشان داد که برخی از برنامه‌های شناسایی شده با آن که در صدر فهرست گوگل واقع شده‌اند، در حد اندیشه باقی مانده بودند، و برخی نیز در نیمه راه از رسیدن به اهداف خود باز مانده بودند. همچنین، برخی برنامه‌ها مستندات بسیار اندکی را ارائه کرده و برخی نیز مستنداتی به زبانهای غیرانگلیسی داشتند. پس از حذف این گونه برنامه‌ها از جامعه پژوهش، در پایان شمار کل برنامه‌های شناسایی شده^۲ به ۸۵ برنامه رسید. زمینه فعالیت و ویژگی‌های برنامه‌های برتر حفاظت وب با بررسی مستندات برنامه‌های حفاظت دیجیتال و نیز ادبیات نظری یا پژوهشی مربوطه و تماس با مسئولان برنامه شناسایی شد. ۱۵ برنامه از اقلام شناسایی شده اهداف تحقیق و توسعه را دنبال می‌کردند. بدین معنا که ابعاد مختلف حفاظت را به هدف تحقیقاتی صرف مورد مطالعه قرار می‌دادند، یا به هدف زمینه‌سازی اجرای پژوهش‌های عملیاتی در آینده صورت گرفته بودند. برنامه‌های تحقیق و توسعه به دلیل سرشت و حوزه فعالیت‌های متفاوت در این پژوهش مد نظر قرار نمی‌گیرند. شمار برنامه‌های عملیاتی شناسایی شده به ۷۰ مورد رسید. دستیابی به اطلاعات کامل درباره ۲۴ برنامه، به دلیل کمبود مستندات ارائه شده در اینترنت، در عین حال عدم واکنش به فرمهای بازخورد ارسالی یا ارائه اطلاعات به زبانهای غیرانگلیسی و به‌روز نشدن وب سایت و خدمات میسر نبود. با حذف این برنامه‌ها، شمار برنامه‌های عملیاتی قابل بررسی به ۴۶ رسید (جدول ۱).

۱. PADI (<http://www.nla.gov.au/padi/topics/43.html>)

۲. به جهت رعایت اختصار برای همه‌ی طرح‌ها و برنامه‌ها از واژه «برنامه» استفاده می‌شود.

جدول ۱. برنامه‌های برتر حفاظت منابع وبی^۱

Table 1. Top Web Preservation Programs

ردیف	نام برنامه	گستره	موضوع / دامنه / نوع منبع
	Program name	Scope	Subject/domain/Document type
1	Anarchism web archive	موضوعی subject	آنارشیزم Anarchism
2	AOLA (Austrianonline archive)	دامنه‌ای Domain	فضای وب ملی National web space
3	Archipol	موضوعی؛ دامنه‌ای Subject; Domain	وبسایت‌های سیاسی Politics websites
4	Archiv (Archive of the Czech web)	دامنه‌ای Domain	فضای وب ملی National web
5	Archived Harvard's Website	سازمانی	وبسایت‌های دانشگاه هاروارد Harvard university websites
6	BNF (national library of france)	دامنه‌ای Domain	فضای وب ملی National web
7	British Slave Trade Legacies	موضوعی subject	تجارت برده Slave Trade
8	CiteSeer	موضوعی subject	علوم رایانه و اطلاع‌رسانی Computer and Information Science
9	DACHS (digital archive for Chinese studies)	موضوعی subject	مسائل سیاسی و اجتماعی Social and politic Issues
10	Dept. of labor web archiving project	منابع خاص؛ دامنه‌ای document type, Domain	وبسایت‌های دولتی، وزارت کار Government websites, Department of Labor
11	Diva	منابع خاص، دامنه‌ای document type, Domain	گزارش‌های تحقیقاتی Research reports
12	European archive	-	-
13	Finnish Web Archive	دامنه‌ای Domain	فضای وب ملی National web
14	GCWA (Government of Canada Web Archive)	دامنه‌ای Domain	دامنه وب دولت فدرال federal websites
15	Ina (institut national de l'audiovisuel)	منابع خاص؛ دامنه‌ای document type, Domain	منابع دیداری شنیداری Audiovisual media
16	Internet Archive	نامحدود Unlimited	فضای وب جهانی World Wide Web
17	k-12 web archiving	-	منابع وبی انتخابی دانش‌آموزان web documents select by students
18	Kb web archiving	دامنه‌ای Domain	فضای وب ملی National web
19	Kulturarw	دامنه‌ای Domain	فضای وب ملی National web
20	LAC (Electronic collection of library and archives canada)	دامنه‌ای Domain	فضای وب ملی National web
21	Lawap (Latin American web archiving project)	منابع خاص؛ موضوعی؛ دامنه‌ای Subject, document type	اسناد دولتی و بیانات سیاسی Government document
22	Mellon Project on Web Resource Collection Program Development	موضوعی Subject	موضوعات مهم برای سازمان Important subject for organization
23	Minerva (Library of Congress Web Archives)	23	موضوعات متنوع Various subjects

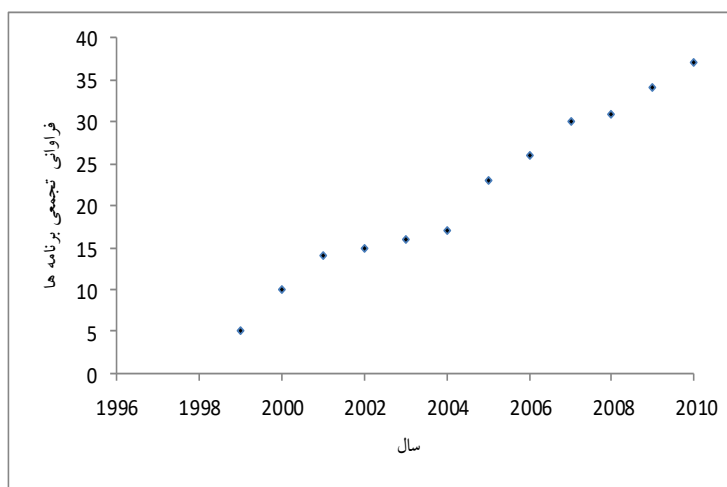
برای رعایت اختصار، از این پس از تکرار نام برنامه‌ها خودداری می‌شود و در صورت نیاز تنها به شماره ردیف آنها در جدول ۱ ارجاع داده خواهد شد.

ردیف	نام برنامه	گستره	موضوع / دامنه / نوع منبع
	Program name	Scope	Subject/domain/Document type
24	National Library of Iceland- Electronic Legal Deposit	دامنه‌ای Domain	فضای وب ملی National web
25	Netarkivet. dk (Royal library and the state and university library)	دامنه‌ای Domain	فضای وب ملی National web
26	New Zealand Domain harvest	دامنه‌ای Domain	فضای وب ملی National web
27	OASIS (online archiving & searching internet sources)	دامنه‌ای؛ موضوعی Domain, subject	فضای وب ملی و همچنین موضوعات سیاست، فرهنگ، مذهب، علوم، اقتصاد National web, Politics, Culture, Religion, Science, Economy
28	Our Digital Island	دامنه‌ای Domain	فضای وب تاسمانیا Tasmania's web space
29	PADICAT (Patrimoni digital de Catalunya)	دامنه‌ای Domain	فضای وب ملی National web
30	PANDORA Australia's Web Archive	دامنه‌ای Domain	فضای وب ملی National web
31	Portuguese Web Archive	دامنه‌ای Domain	فضای وب ملی National web
32	Program e-Helvetic of the Swiss National Library	دامنه‌ای Domain	فضای وب ملی National web
33	Slovenia	دامنه‌ای Domain	فضای وب ملی National web
34	STORS	دامنه‌ای Domain	فضای وب تاسمانیا Tasmania's web space
35	The CyberCemetery	منابع خاص	وبسایت‌های دولتی Government websites
36	The National Library of Wales Web Archiving Program	دامنه‌ای Domain	فضای وب ملی National web
37	UK Government Web Archive	منابع خاص document type	وبسایت‌های دولتی Government websites
38	UK Web Archive	دامنه‌ای Domain	فضای وب ملی National web
39	University of Michigan Web Archives	سازمانی organization	وبسایت‌های مرتبط با دانشگاه Websites related to the university
40	VERS (Victorian electronic records strategy)	نوع منبع خاص، دامنه‌ای Domain	منابع وب دولت ویکتوریا Government of Victoria web document
41	WAS (The Web Archive Singapore)	دامنه‌ای Domain	فضای وب ملی National web
42	Web Archive Norway	دامنه‌ای Domain	فضای وب ملی National web
43	Web Archiving @ the University of Melbourne	سازمانی Organizational	وبسایت دانشگاه The university website
44	web archiving project	نوع منابع خاص document type	وبسایت‌های و منابع دولتی Government websites and documents
45	Web harvesting project of the German National Library	دامنه‌ای Domain	فضای وب ملی National web
46	WebCite	نامحدود Unlimited	منابع مورد استناد در مقاله‌ها Citations in articles

یافته‌های پژوهش

سابقه فعالیت و تاریخ پوشش مجموعه

به منظور آگاهی از سابقه فعالیت برنامه‌ها، تاریخ راه‌اندازی آنها مورد بررسی قرار گرفت. یافته‌ها نشان داد «اینترنت آرکایو» قدیمی‌ترین و تنها برنامه‌ی حفاظت وب است که تاریخ آغاز به کار آن به سال ۱۹۹۶ باز می‌گردد. در خلال سال‌های ۱۹۹۷ تا ۱۹۹۸ هیچ برنامه‌ی دیگری راه‌اندازی نشده است. از آن پس در هر سال دست کم یک برنامه در هر سال آغاز شده است. جدیدترین آنها «آرشیو وب آنارشیزم»^۱ و پروژه‌ی دروگری وب کتابخانه ملی آلمان است. تاریخ راه‌اندازی چند برنامه از جمله «یو کی وب آرکایو»^۲، برنامه‌ی کتابخانه ملی ایسلند، اسلونیا، «وب سایت»^۳، آرشیوگری وب دانشگاه ملبورن، «واس»^۴ و «سایت سی یر»^۵ مشخص نشده است.



نمودار ۱: روند رشد فراوانی تجمعی برنامه‌های فعال حفاظت وب بر حسب زمان^۶

Chart 1. Growth of cumulative frequency web preservation programs based on time

چنانچه درصد تجمعی برنامه‌های فعال در هر سال مد نظر قرار گیرد، مشاهده خواهد شد که شمار برنامه‌ها رو به افزایش دارد (نمودار ۱). این یافته علاوه بر این که روند روزافزون توجه به ارزش حفاظت منابع وبی را نشان می‌دهد، نشان از نوپایی این جنبش در جامعه‌ی علمی دارد. همچنین، می‌تواند تا حدودی نشانگر آن باشد که هر چه منابع جدیدتر باشند، بخت حفاظت آنها از سوی برنامه‌های موجود بیشتر است و منابع قدیمی‌تر بیشتر در معرض خطر هستند. تاریخ راه‌اندازی برنامه‌ها تا حدودی می‌تواند پوشش زمانی مجموعه‌ها را نیز مشخص سازد. نخستین تلاش‌ها برای حفاظت وب، که خوشبختانه گستره‌ای در سطح جهانی را برگزیده است، «اینترنت آرکایو» است که قدیمی‌ترین آثار حفاظت شده در آن به سال ۱۹۹۶ باز می‌گردد. با آن که مجموعه‌ی این آرشیو نیز بنا به گواه محققان چندان کامل و قابل اطمینان نیست، اما این نوید را می‌دهد که بخش قابل توجهی از منابع قدیمی وب حفاظت شده است. در برخی برنامه‌ها، مانند «سایبرسیمتری»^۶، آرشیو وب پرتغال و آرشیو وب کانادا، تاریخ

۱. Anarchism web archive
۲. UK Web Archive
3. WebCite
۴. WAS (The Web Archive Singapore)
۵. Citeseer

۶. سال ۱۹۹۶ به دلیل وجود شکاف دو ساله در این تحلیل دخالت داده نشده است و نخستین نقطه داده‌ای مربوط به سال ۱۹۹۹ است.

مجموعه‌های تحت پوشش به دو تا نه سال پیش از راه اندازی آنها باز می‌گردد. این برنامه‌ها منابع مورد نظر در سال‌های قبل از راه‌اندازی خود را از دیگر برنامه‌های وب به ویژه «اینترنت آرکایو» بدست می‌آورند. حفاظت در چند برنامه‌ی مختلف، نویدی برای حفاظت بخشی گسترده‌تر از منابع وبی و افزایش شانس بقای منابع است، با این حال این برنامه‌ها بر خلاف «اینترنت آرکایو» گستره‌ای محدود را پوشش می‌دهند.

گستره حفاظت

هر برنامه، با توجه به اهداف و توانمندی‌های خود گستره‌ای خاص را برای حفاظت منابع وبی برمی‌گزیند. برنامه‌ها، ممکن است با انتخاب کل وب، گستره‌ای جهانی را پوشش دهند یا گردآوری منابع وبی را به دامنه ملی یا منطقه‌ای، موضوع، سازمان یا نوعی خاص از مواد محدود کنند یا آمیزه‌ای از این گستره‌ها را انتخاب کنند. می‌توان چنین استنباط کرد که پوشش موضوعی برنامه‌های منطقه‌ای، ملی و جهانی تقریباً نامحدود و به گستردگی موضوعاتی است که در فضای سایبری تحت پوشش وجود دارد. پوشش موضوعی تحت حفاظت در گستره‌ی سازمانی نیز به حیطه فعالیت آن سازمان محدود می‌گردد. نیمی از این برنامه‌ها دامنه‌ای ملی را مورد حفاظت قرار می‌دهند. آشکار است که برنامه‌های برتر، حاصل کار سازمان‌هایی با گستره فعالیت وسیع و سطح برخورداری بالا است. از این رو، به دور از انتظار نیست که این گستره بیش از همه در برنامه‌های برتر پوشش داده شده باشد. به ویژه آن که دروی کل دامنه به روش‌های خودکار مستلزم صرف وقت، هزینه و نیروی انسانی کمتری است (جدول ۲).

پوشش موضوعی خاص، در رده‌ی بعدی قرار دارد (۱۳/۰۴ درصد). بعضی از برنامه‌ها به جمع‌آوری منابع وبی در موضوع یا موضوع‌های خاصی می‌پردازند، در این صورت حیطه دروگری منابع وبی در آن موضوع، محدود به دامنه‌ی ملی یا بین‌المللی می‌شود. شماری از برنامه‌ها نیز هر دو گستره دامنه و موضوع را پوشش داده‌اند، بدین معنی که در کنار دروگری کل دامنه، به دروگری گزینشی پرداخته و مجموعه‌های موضوعی خاصی را به وجود می‌آورند از این جمله‌اند برنامه آرشیو وب چکوسلواکی و آرشیو وب اسپانیا.

شمار بالایی از برنامه‌ها نیز رویکرد آمیخته را در پیش گرفته‌اند که در همه‌ی موارد، گستره‌ی دامنه‌ی ملی را شامل می‌شود. ممکن است این برنامه‌ها رویکرد تلفیقی را در مجموعه‌سازی برگزیده باشند. یعنی پس از دروی کل دامنه، به مجموعه‌سازی گزینشی برخی مواد با معیارهای مشخص مثلاً موضوع یا نوع خاص می‌پردازند. یا این که موضوع یا منابع مورد نظر را تنها در دامنه‌ی ملی گردآوری کنند. چنانچه رویکرد آمیخته را در نظر بگیریم، گستره جهانی که کل وب را برای حفاظت برگزیده است از کمترین فراوانی برخوردار است. تنها «وب سایت» و «اینترنت آرکایو» دامنه بین‌المللی را پوشش می‌دهند. این امر به احتمال قوی به توانایی‌های کمتر و در عین حال مأموریت‌ها و رسالت‌های متفاوت سازمان‌های دیگر دخیل در حفاظت منابع وبی باز می‌گردد. دلیل دیگر به سرشت خود برنامه‌ها باز می‌گردد. برای مثال، پوشش بین‌المللی در برنامه «وب سایت» به آن دلیل است که بر مشارکت همه‌ی پدیدآوران از سراسر جهان تکیه دارد. در این پروژه، نویسندگان نسخه‌ای از صفحات وب مورد استناد و نشانی مورد استناد را در وب‌سایت آرشیو می‌کنند (جدول ۲).

جدول ۲: گستره حفاظت

Table 2. Scope of Preservation

درصد Present	فراوانی Frequency	گستره (Scope)
50	23	دامنه ملی National level
13/04	6	موضوعی Subject
8/7	4	دامنه ملی / نوع منابع خاص National domain / Document type
6/5	3	سازمانی Organizational
6/5	3	نوع منابع خاص Document type
4/3	2	دامنه ملی / موضوعی National domain /subject
4/3	2	کل وب World wide web
2/2	1	دامنه ملی / موضوعی / نوع منابع خاص National domain/Subject/Document type
4/3	2	نامعلوم Unknown
100	46	جمع Total

موضوعات و نوع منابع تحت پوشش

موضوعاتی که به طور خاص در رویکرد موضوعی مورد توجه قرار گرفته‌اند عبارتند از: سیاست، فرهنگ، مذهب، علوم، اقتصاد، برده‌داری و برده‌فروشی، دولت، آنارشیزم، حقوق بشر، مسائل اجتماعی، رایانه و علوم اطلاعات (جدول ۳).

جدول ۳: موضوعات و نوع منابع تحت پوشش نمونه‌هایی از برنامه‌های حفاظت وب

Table 3. Subjects and Types of documents preserved

شماره برنامه Program's number	نوع منابع Document type	شماره برنامه Program's number	موضوع Subject
11	گزارش‌های تحقیقاتی Research reports	27	اقتصاد Economy
15	رسانه‌های دیداری و شنیداری Audiovisual media	1	آنارشیزم Anarchism
14,21,34,35,37,40,44	منابع دولتی Governmental Documents	7	برده‌داری و برده‌فروشی Slavery and slave trade
2,3,5,6,10,12,21,25,27,29,39,41	وبسایت‌ها Websites	22	حقوق بشر Human Rights
6,12	وبلاگ‌ها Weblogs	14,37,44	دولت Government
7,13,18,22,24,28,30,33,36,38,45	تلفیقی از انواع منابع A combination of document types	3,21,27	سیاست Politic
26,32,42	تمام انواع منابع All document types	27	علوم Science
		27	فرهنگ Culture
		27	مذهب Religion
		9	مسائل اجتماعی Social Issues

بعضی برنامه‌ها در کنار دروغ‌گری کل دامنه، رویکرد گزینش‌گری را در پیش گرفته‌اند و مجموعه‌های موضوعی خاصی را راه‌اندازی کرده‌اند مثل «اینترنت آرکایو» که در دل مجموعه‌های خود، مجموعه‌های موضوعی خاص به وجود آورده است. برنامه‌ی آرشیو وب نروژ مجموعه‌هایی مانند «انتخابات ۲۰۰۶» و «سیل ۲۰۰۲» را گرد آورده است و برنامه‌ی پدیکت مجموعه‌هایی مانند «موزه‌های کاتولونیا»، «انتخابات اروپا» و «انتخابات ۲۰۱۰» را به وجود آورده است. بعضی از برنامه‌ها، همانند «مینروا» گستره‌ی موضوعی خاصی را تعریف نکرده‌اند. بلکه آن دسته از منابع وبی را که به رویدادها و موضوعات «مهم» می‌پردازند مد نظر قرار می‌دهند.

انواع متفاوتی از منابع وبی در فضای سایبری وجود دارد که طیف وسیعی از وب‌سایت‌ها و وب‌لاگ‌ها گرفته تا مجلات، روزنامه‌ها، اخبار، منابع دولتی، گزارش‌های تحقیقاتی را در بر می‌گیرند. برخی برنامه‌ها مجموعه‌ی خود را به یک یا دو نوع منبع محدود کرده‌اند اما برخی دیگر آمیزه‌ای از انواع منابع را پوشش می‌دهند. اکثر برنامه‌هایی که نوع منبع را محدود نکرده‌اند یا طیف وسیعی از منابع را تحت پوشش خود اعلام کرده‌اند، در زمره‌ی برنامه‌هایی با پوشش حفاظتی ملی، جهانی یا تلفیقی از آن هستند (جدول ۳).

الگوی دسترسی و نوع کاربران

دسترسی به آرشیو منابع ممکن است آزاد، نیمه‌آزاد یا محدود باشد. در الگوی دسترسی آزاد، هیچگونه محدودیتی برای دسترسی کاربران به آرشیو منابع وبی خود قائل نمی‌شوند. دسترسی نیمه‌آزاد منوط به اخذ مجوز و مذاکره با ناشران و مالکان وب‌سایت‌هاست. دسترسی می‌تواند به لحاظ محل دسترسی، زمان دسترسی، نوع کاربر مشخص یا بخشی از مجموعه محدود شود. بیشینه برنامه‌های حفاظت (۳۹/۱ درصد) الگوی دسترسی آزاد را انتخاب کرده و پس از آن دسترسی محدود از فراوانی بالایی برخوردار است (۲۳/۹ درصد). دسترسی نیمه‌آزاد به منابع وبی آرشیو شده از کمترین فراوانی در میان برنامه‌ها برخوردار است (۶/۱۷ درصد) (جدول ۴).

جدول ۴: وضعیت دسترسی کاربران برنامه‌ها به منابع وبی حفاظت شده

Table 4. Models of access to the preserved documents

درصد Present	فراوانی Frequency	نوع کاربران مجاز Types of Users	درصد Present	فراوانی Frequency	الگوی دسترسی Access Models
65/3	30	جهانی World	39/1	18	آزاد Fully open
15/2	7	کاربران خاص Authorized users	23/9	11	محدود Restricted access
19/6	9	نامعلوم Unknown	19/6	9	نیمه‌آزاد Semi-open
100	46	جمع Total	17/4	8	نامعلوم Unknown
			100	46	جمع Total

بیشینه‌ی برنامه‌ها (۶۵/۳ درصد) خدمات خود را به جامعه‌ی جهانی عرضه می‌کنند و دسترسی به آرشیو به نوع کاربر خاصی محدود نمی‌شود. نمونه بارز این نوع برنامه‌ها «اینترنت آرکایو» است. شمار اندکی نیز (۱۵/۲ درصد) دسترسی را به نوع کاربران خاصی محدود می‌کنند. برای نمونه آرشیو پیوسته اتریش و پروژه‌ی دروغ‌گری وب کتابخانه ملی آلمان دسترسی را تنها برای دانشجویان و محققان امکان‌پذیر کرده‌اند. محدود کردن نوع کاربر می‌تواند به دلایل مختلفی از جمله رعایت مسائل حقوقی

و حق مؤلف، پیشگیری از آسیب به کارآیی سامانه در نتیجه ازدیاد کاربران و همچنین خاص بودن اهداف یا موضوع مجموعه باشد (جدول ۴).

بحث و نتیجه گیری

در این پژوهش تجربیات موفق جهانی در زمینه حفاظت وبی معرفی شد و ویژگی های مختلف برنامه های برتر حفاظت وب به لحاظ سابقه فعالیت، گستره پوشش، نوع منابع، الگوی دسترسی، و جامعه تحت پوشش مورد بررسی قرار گرفت. نتایج نشان داد که شمار روزافزونی از برنامه های حفاظت وبی مشغول فعالیت هستند. این امر نشان از اهمیت حفاظت وب و آگاهی روز افزون در این باره دارد و ضرورت افزایش توجهات داخلی به این مهم را برجسته می سازد.

در ایران، تجربه ی «پروژه حرم» و «کنسرسیوم محتوای ملی» از معدود تلاشهای جدی است که از سوی کتابخانه ملی و با مشارکت کتابخانه ها و مراکز اطلاعاتی سراسر کشور به هدف حفاظت از میراث معنوی دیجیتال آغاز شده است. با آن که این تلاشها بسیار ارزشمند و در خور تقدیر است، اما هیچیک به حفاظت از وب عمومی یا وب «نمایان» توجهی نشان نداده است. از این رو، ضروری است نهادهای ملی متولی حفاظت یا نهادهای علاقمند به میراث معنوی هر چه سریعتر حفاظت از منابع وبی را در دستور کار قرار دهند. نتایج به دست آمده در پژوهش حاضر می تواند الهام بخش فعالیت های حفاظت وب در ایران باشد. خطوط کلی برنامه های مورد بررسی را می توان به طور خلاصه چنین به تصویر کشید:

۱. گستره حفاظت: دامنه ملی، در عین حال توجه به زیرمجموعه های خاص
۲. گستره موضوعات: (تقریباً) همه موضوعات
۳. نوع منابع: تلفیقی از انواع منابع
۴. الگوی دسترسی: دسترسی آزاد
۵. جامعه کاربری: جهانی

گستره ی حفاظت به شدت متأثر از اهداف برنامه و سازمان مادر و نیز توانایی های مالی، تخصصی و انسانی آن است. با آن که برنامه های حفاظت منابع وبی، انواع مختلفی از گستره ها شامل بین المللی، ملی، منطقه ای، سازمانی یا موضوعی را پوشش می دهند، اما پوشش بیشتر برنامه ها در سطح ملی است زیرا در بیشتر برنامه ها کتابخانه ملی مسئولیت راه اندازی این برنامه ها را برعهده دارد. از این رو، در برنامه حفاظت ملی وب ایران می توان به حفاظت کل دامنه ملی کشور پرداخت. این روش، که می تواند به صورت خودکار صورت گیرد مستلزم صرف وقت، هزینه و نیروی انسانی کمتری است، در عین حال برنامه را از گزینش انواع منابعو انتخاب موضوعات برای حفاظت و دشواری های مربوط به آن مصون می دارد. ضمن این که این روش راه را بر گزینشگری نیز نمی بندد. چنانکه مشاهده شد بسیاری از برنامه ها، پس از گردآوری محتوای کل دامنه ملی خود، به استخراج مجموعه های موضوعی خاص از درون آن می پردازند که بهره برداری از منابع درو شده را تسهیل کرده و در نتیجه بهره وری آنها را افزایش می دهد. همچنین، با هدایت کردن کاربران به طرف مجموعه های خاص از کارآیی و امنیت سامانه نیز مراقبت می شود. در عین حال، با انتخاب دامنه سایبری ملی، نباید اهمیت منابعی را که درباره ایران در سراسر جهان در وب منتشر می شود از نظر دور داشت. حفاظت از مجموعه های خاص با موضوع ایران و ایران شناسی از ضروریات کشور است.

تدوین خط مشی حفاظت از چنین آثاری نیز نیاز به تحقیقات گسترده به ویژه به لحاظ استراتژی‌های استخراج موضوع و تحلیل محتوای وب سایت‌ها دارد.

سیاست کلی برنامه‌ها گرایش به ارائه الگوی دسترسی آزاد است. این از آن رو است که بیشتر این برنامه‌ها، وب «نمایان» را هدف قرار داده‌اند و کمتر به وب «پنهان» پرداخته‌اند. برنامه ملی حفاظت در ایران نیز می‌تواند تا جایی که به حفاظت وب «نمایان» مربوط می‌شود دسترسی آزاد عموم کاربران در سراسر جهان را به رسمیت بشناسد. دسترسی باید تنها به نسخه آرشیو شده (و نه نسخه زنده وب سایت‌ها) فراهم شود تا به منافع مالی مالکان وب سایت‌ها نیز خللی وارد نشود.

در صورتی که برنامه حفاظت ملی وب ایران، بخش «پنهان» وب را نیز هدف قرار دهد، تسری دادن قانون واسپاری ملی به منابع وبی نیز از ضروریات این برنامه خواهد بود. این کار این مزیت را خواهد داشت که مجموعه وب پنهان با ارسال منابع از سوی ناشران و مالکان به روز خواهد شد و توسعه خواهد یافت. در عین حال، می‌توان مجموعه وب پنهان را مجزای از مجموعه وب نمایان حفاظت کرده و تحت سازوکارهای امنیتی خاص تنها در اختیار کاربران مجاز قرار داد. لازم الاجرا کردن قوانین حق مولف و مالکیت معنوی و همچنین سازگار ساختن آن با مقتضیات فضای سایبری نیز از دیگر ملزومات حفاظت از وب پنهان در ایران است. نکته قابل توجه آن است که سالهای بسیاری از انتشار منابع در فضای سایبری ایران می‌گذرد. هر گونه تلاشی که از این پس آغاز گردد نمی‌تواند منابع گذشته را به خوبی و به طور کامل حفاظت کند. از این رو، همکاری با نهادهای حفاظت دیگر کشورها به ویژه برنامه‌هایی که دامنه جهانی را هدف حفاظت قرار داده‌اند جهت دستیابی به محتوای فضای سایبری ایران در سالهای گذشته ضروری است.

تعارض منافع

هیچ‌گونه تعارض منافع از سوی نویسندگان گزارش نشده است.

References

- Ashley, K., Davis, R., Guy, M., Kelly, B., Pinsent, Ed, Farrell, S. (2010). "A Guide to Web Preservation". Retrieved October 20, 2011, from: <http://jiscpowr.jiscinvolve.org/wp/files/2010/06/Guide-2010-final.pdf>.
- Barateiro, J., Antunes, G., Borbinha, J. (2009). Addressing Digital Preservation: Proposals for New Perspectives. In: First International Workshop on Innovation in Digital Preservation, June 19, 2009, Austin, Texas, USA. Retrieved October 20, 2011, from: <http://cs.harding.edu/indp/papers/barateiro7.pdf>.
- Beagrie, N. (2003). National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity. Available at: <http://www.cilir.org/pubs/reports/pub116/sec1.html>.
- Charlesworth, A. (2006). Digital Curation, Copyright, and Academic Research. The International Journal of Digital Curation, 1(1): 17-32.
- Connertz, Th. (2003), Long-term archiving of digital documents: what efforts are being made in Germany?. Learned Publishing. 16(3): 207- 211.
- Day, M. (2006). Long_Term Preservation of Web Content. Available at: www.ukoln.ac.uk/preservation/publications/2006/web.../md-final-draft.pdf
- Emmanuelle, B. and Gildas, I. (2009). Metrics and Strategies for Web Heritage Management and Preservation. In: 75TH IFLA General Conference and Council (23-27 August 2009, Milan, Italy).
- Gomes, D., Freitas, S., Silva, Mário J. (2006). Design and Selection Criteria for a National Web Archive. Retrieved October 20, 2011, from: <http://xldb.fc.ul.pt/daniel/docs/papers/gomes06tomba.pdf>
- Hodge, G. and Frangakis, E. (2004). Digital Preservation Access to Scientific Information: The State of the Practice. Retrieved October 20, 2011, from: <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA423497>
- Kelly, B., Ashley, K., Guy, M., Pinsent, E., Davies, R. and Hatcher, J., (2008). Preservation of Web Resources: The JISC PoWR Project. In: iPress 2008 Conference, 29-30th September 2008, British Library, London, UK.
- Library of Congress (2005). Minerva Web archiving project. Retrieved October 20, 2011, from: <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html>
- Murray, K. R. and Hsieh, I. (2007). Archiving Web-published materials: A needs assessment of librarians, researchers, and content providers. Government Information Quarterly, 52: 66-89.
- Ntoulas, A., Cho, J., Olston, C. (2004). What's new on the web? The evolution of the web from a search engine perspective. In: 13th international conference on World Wide Web, p. 1-12. Retrieved October 20, 2011, from: oak.cs.ucla.edu/~cho/papers/cho-new.pdf.

- Rauber, A. and Hunter, J (2007). Introduction to the Special Issue on Web Archiving. *New Review of Hypermedia and Multimedia*, 13(1).
- Shadanpour, Farzaneh (2006). The digital heritage. *National studies on librarianship and information organization*, 72:93-107.
- Strodl, S., Becker, C., Rauber, A. (2009). Digital Preservation. Retrieved October 20, 2011, from: http://www.google.com/#hl=fa&q=Digital+Preservation+Stephan+Strodl&bav=on.2,or.r_gc.r_pw.&fp=f3307e7a3d9ecbc4&biw=1600&bih=695.
- Verheul, I. (2006). Networking for digital preservation: current practice in 15 national libraries. Retrieved October 20, 2011, from: <http://archive.ifla.org/V/pr/saur119.htm>.
- Waller, M. and Sharpe, R. (2006). Mind the gap: Assessing digital preservation needs in the UK. Retrieved October 20, 2011, from: www.dpconline.org/.../340-mind-the-gap-assessing-digital-preservation-needs-in-the-uk.html

COPYRIGHTS



© 2019 by the authors. Licensee SCU, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)