

Sayısal (Dijital) Kültürün Korunması: Web Arşivleme

Preservation of the Digital Culture: Archiving the World Wide Web

Ahmet Aldemir¹
Esin Sultan Oğuz²

Öz

Web ortamında yaşanan bilgi artışı, beraberinde bu bilgilerin gelecek nesillere aktarılması amacıyla arşivlenmesi gereğini gündeme getirmiştir. Web'in arşivlenmesi teknik, yasal ve örgütsel boyutları olan çok yönlü bir uygulamadır. Sayısal ortamda üretilmiş bilginin yaşam döngüsündeki her bir aşama, web ortamında yer alan bilgiler için hayati önem taşımaktadır. Dünyada bir çok ülke milli kütüphaneleri öncülüğünde web arşivleme çalışmalarını başlatmış ve bu girişimlerinin yasal bir platforma taşınması için gerekli adımlar atılmıştır. Bu gelişmeler ışığında çalışmamızda, web'in neden arşivlenmesi gerektiğine değinilmiş, web arşivlemede kullanılan belli başlı yaklaşımlar ele alınmış, ulusal ve uluslararası ölçekli web arşivleme çalışmalarına yer verilmiştir.

Anahtar Kelimeler: Sayısal (dijital) koruma, Sayısal (dijital) arşivleme, Web arşivleme, Web arşivleri

Abstract: Information growth in the web medium has required the necessity of archiving these information to transmit them to future generations. Web archiving is a versatile application which covers technical, legal and organizational dimensions. Any stage within the life cycle of digital information is critically important for the information in web environment. All over the world, many countries have started web archiving efforts through the leadership of their national libraries and attempted to carry these initiatives on a legal bases. In the light of these developments, this paper examines the necessity and major techniques in web archiving and it also discuss national and international web archiving projects.

¹ Uzman Kütüphaneci. Sakarya Üniversitesi. e-posta: aaldemir@sakarya.edu.tr

² Arş.Gör. Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü. e-posta: esino@hacettepe.edu.tr

Keywords: *Digital preservation, Digital archiving, Web archiving, Web archives*

Giriş

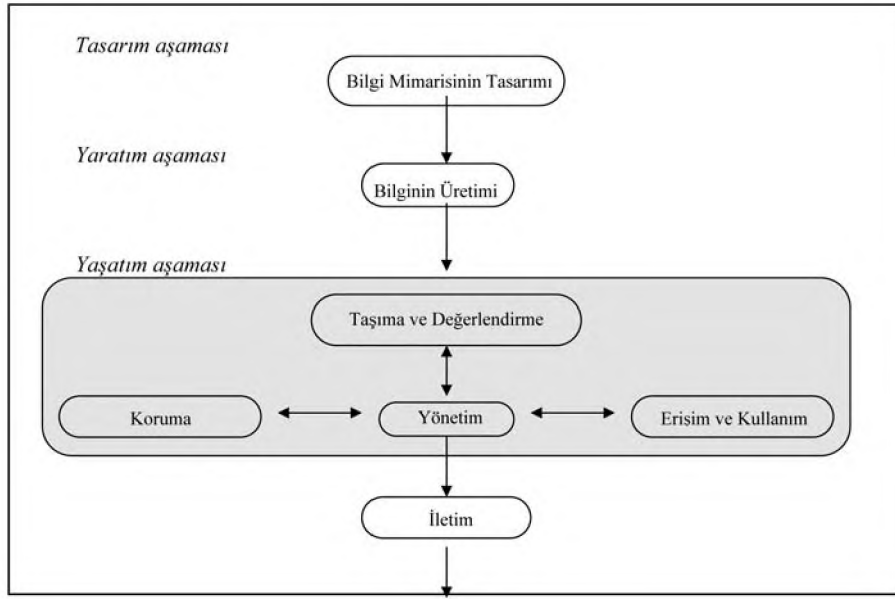
Web'in büyüklüğüne yönelik yapılan çalışmalar, web'in dinamik yapısından dolayı kesin bir yargıda bulunmanın zor olduğuna dikkati çekmekte ve çoğunlukla indekslenebilir olarak nitelenen yüzeysel web'in büyüklüğünü belirlemeye yönelik hesaplamalarda yoğunlaşmaktadır. Bu düşünceden hareketle yapılmış bir araştırma, arama motorları tarafından kapsanan sayfa miktarının Google için 8 milyar sayfa, MSN için 5 milyar sayfa, Yahoo için 4 milyar sayfa ve Ask Jeeves için 2,5 milyar sayfa olduğunu göstermektedir (Sullivan, 2004). Ocak 2005 sonu itibarıyla ise; arama motorları tarafından indekslenebilir web büyüklüğünün 11.5 milyar sayfa olduğu belirtilmiştir (Gulli ve Signorini, 2005). Bununla birlikte Hellerstein (2004), Web'de erişilebilen bilginin, Internet'te yer alan bilgilerin çok küçük bir bölümünü temsil ettiğini; arama motorları tarafından indekslenmiş bilginin, derin Web olarak adlandırılan 7,5 petabyte büyüklüğündeki alandan 550 kez küçük olduğunu belirtmektedir.

2001 yılında üretilmiş tüm yeni bilgilerin, %93 oranında, anadan doğma sayısal (born digital) olduğu kaydedilmiştir (Lyman ve Varian, 2003). Sayısal ortamda üretilmiş bilginin giderek önem kazandığı göz önüne alındığında web'deki bilgi miktarının büyüklüğü kadar bu bilginin kalıcılığı da önem kazanmaktadır. 1998 yılında çevrimiçi olan web sitelerinin %44'üne ertesi yıl erişilemediği belirtilmektedir. Aynı zamanda, yapılan araştırmalar bir Web sitesinin ortalama ömrünün 44 gün olduğunu ortaya koymaktadır (Lyman, 2002). Web'deki bu hızlı büyüme ve değişim, beraberinde bir kaos getirmiştir. Web ortamında yer alan elektronik bilgi kaynaklarının toplanmasına ve korunmasına yönelik bir politika geliştirilmediği sürece akademik, kültürel ve bilimsel kaynaklara gelecek nesillerin erişimi mümkün olmayacak gibi görünmektedir.

Elektronik Bilginin Yaşam Döngüsü

Sayısal ortamda başarılı bir arşivleme çalışması gerçekleştirilebilmek, elektronik bilginin yaşam döngüsündeki tüm aşamaların göz önünde bulundurulmasını

gerektirir. Elektronik bilginin üretimi, yönetimi ve bununla ilişkili olarak sağlama ve elektronik derme geliştirme, kullanılacak sayısal koruma stratejileri, bilgi erişim ve bilginin iletimi, elektronik bilginin yaşam döngüsünü oluşturur. Her birinde farklı görevlerin yerine getirildiği bu aşamalar arasındaki etkileşimin sağlanması, birtakım kurallar ve standartlar sayesinde gerçekleşmektedir (DLM Forum, 1997, s. 17).



Şekil 1: Elektronik bilginin yaşam döngüsü (DLM Forum, 1997, s.17)

Elektronik bilginin üretimi sırasında uygulanan işlemler, bilginin sayısal olarak arşivlenmesini ve korunmasını kolaylaştırıcı etkiler yaratmaktadır. Materyalin arşivlenmesine karar verilmeden önce materyale yönelik formata karar verilmesi, standartlaşmanın sağlanması ve metadata tanımlarının yapılması gibi unsurların göz önünde bulundurulması, koruma ve arşivleme sürecini daha verimli kılmaktadır.

(Hodge, 2005, s. 48). Bu nedenle, elektronik bilgiyi üreten ve yayınlayan kişilerle, bu bilginin korunmasından sorumlu olan kişiler arasındaki iletişimin elektronik bilgi için hayati önemi vardır (Chowdhury ve Chowdhury, 2003, s. 217).

Elektronik bilginin sağlanması, sayısal arşivlemede uygulanan işlemlerin başında yer almaktadır. Sağlama ise; koleksiyon yönetimi politikasının belkemiğini oluşturur. Tipik olarak, kaynak seçimi, bilgi kaynağının formatına göre belirlenmez. Kaynağın koleksiyona sağlanmak üzere seçilmesindeki belirleyici unsur, söz konusu kaynağın koleksiyonun bütünüyle uyumlu olup olmadığıdır. Elektronik kaynakların formatlarında gözlenen hızlı değişim, seçme ve sağlama işlemlerinde yeni ölçütleri beraberinde getirmiştir. Kaynaklar ile koleksiyon arasındaki uyumun, kaynakların formatlarındaki işlevselliğin yanı sıra sürekliliği ile doğru orantılı oluşu, elektronik kaynak sağlama yaklaşımına farklı bir boyut kazandırmıştır. Bir başka deyişle, elektronik kaynaklar söz konusu olduğunda, kaynağın formatı ve bu formata yönelik standartlar, bir kaynağın neden seçildiği sorusuyla doğrudan ilişkilidir (The Cedars Project, 2002, ss.13-14).

Elektronik bilgi kaynakları, yapıları gereği dinamik ve değişkendir. Sayısal ortamda, adres değişikliğine bağlı erişim sorunları ve aynı kaynağın farklı adreslerde ve farklı formatlarda yer alabilme durumu, elektronik bilgi kaynaklarının denetimini ve düzenlenmesini zorlaştıran unsurlar arasındadır (Kaynak, 2001, ss. 193-194). Elektronik bilgi kaynaklarındaki nitel ve nicel artışın yarattığı bilgi erişim sorununa çözüm olarak geliştirilen “metadata” kavramına yönelik en sık kullanılan tanımlamalardan biri Miller (1999) tarafından “veri hakkındaki veri” şeklinde yapılmıştır. Al ve Küçük (2003, s. 180), web’in hızlı gelişimini olumlu bir gelişme olarak nitelendirirken, mevcut olanakların web’in dizinlenmesinde yetersiz kaldığını belirtmekte ve Internet’te kaynak keşfi ve erişiminin güçleştiğine dikkati çekmektedirler. Bu nedenle, metadata gibi, elektronik kaynakların tanımlanmasıyla ilgili standartların geliştirilmesine yönelik yapılacak çalışmaların, kaynak keşfi ve bilgi erişime yönelik sorunlara çözüm getireceği düşünülmektedir.

Elektronik bilgiye sürekli erişimin, ancak koruma ve arşivleme ile mümkün olabileceğinin altını çizen Tonta (2005, s. 9), bilginin kayıtlı olduğu fiziksel ortamı korumanın aynı zamanda entelektüel içeriğin korunması anlamına geldiğini ve bu fiziksel ortama zarar gelmediği sürece, içeriğe erişimin mümkün olduğunu ifade etmektedir. Ancak, elektronik ortamdaki bilginin korunması oldukça problemlidir. Diğer geleneksel bilgi taşıyıcılarla (kağıt, mikrofiş gibi) karşılaştırıldığında, elektronik bilginin ömrü çok daha kısadır. Aynı zamanda günümüzde elektronik bilgi kaynaklarına erişimin sağlanması, içeriğin ve o içerikle birlikte gelen teknolojinin bir arada korunmasını gerektirmektedir

(Tonta, 2005, s. 9). Arşive depolama sırasında kullanılacak olan formatın / formatların ne olacağı sayısal arşivlemenin en önemli konuları arasında yer almaktadır. Bu nedenle, sayısal formatların koruma dayanıklılığını (uzun ömürlülüğünü) ölçen yöntemler geliştirilmiştir. Geliştirilmiş çeşitli ölçüm araçları ile, bu formatlar tanımlanarak sayısal bilginin korunması için en uygun formatın belirlenmesi amaçlanmaktadır (Stanescu, 2004). Kaynağın türüne bağlı olarak değişebilecek olan arşivlemede, format belirlemeye yönelik olarak az da olsa standartlaşmanın varlığından söz etmek mümkündür. Birçok sayısal dergi, danışma kaynağı veya kitap için TIFF görüntü dosyaları, PDF ya da HTML kullanılmaktadır. Kağıt üzerindeki bilginin sayısal ortama dönüştürülmesinin gerektiği durumlarda TIFF formatı, tamamen sayısal belgeler için de Adobe PDF (Portable Document Format) en uygun format olarak değerlendirilmektedir. Ancak, uzun vadeli bir koruma söz konusu olduğunda, PDF her ne kadar yaygın bir format türü haline gelmişse de, müseccel (ticari) marka olması nedeniyle yasal bir depolama formatı olarak kabul edilmeyecektir. Bu nedenle Adobe, Bilgi ve İmaj Yönetimi Derneği (Association for Information and Image Management) ve çeşitli organizasyonların bir araya gelmesiyle PDF-A olarak adlandırılan standart geliştirilmiş ve böylece söz konusu formata yasal statü kazandırmak amaçlanmıştır. Bu sayede, en basit PDF özellikleri ve işlemleri için dosya tanımlaması yapmak (specification) mümkün olabilecek ve PDF'in farklı sürümleri arasındaki transferlere (aktarımlar) devam edilebilecektir (Hodge, 2005, ss.49-50).

Sayısal ortamdaki korumayı ya da arşivlemeyi sayısal olarak oluşturulmuş ya da sonradan sayısal ortama aktarılmış elektronik bilgilerin uzun süreli depolanmasını ve erişimini içeren bir süreç olarak tanımlayan Bullock (1999), sayısal korumayla hedeflenen üç amaçtan bahsetmektedir. Dokümanın sabitliğini (stability) sağlamak; fiziksel özelliklerini, içeriğini, dokümanın bütünlüğünü korumak ve fonksiyonlarının ve provenansının korunarak uzun dönemli depolamak şeklinde sıralanan bu amaçları gerçekleştirmede kullanılan iki temel koruma stratejisi vardır: teknoloji göçü (migration) ve öykünüm (emulation). Teknoloji göçü, nesnelerin arşivlenmesi ile yazılım ve donanım teknolojisinde yaşanacak değişimler doğrultusunda transferini içeren işlemler; öykünüm ise, teknoloji göçüne alternatif, eski yazılım ve donanım teknolojisinin yeni yazılım ve donanım teknolojisine uyarlanması süreci olarak tanımlanmaktadır. Bu sayede, eski sistemlerin, yeni sistemler üzerinde, kullanılabilir ve erişilebilir hale getirilmesi ve daha uzun vadeli sayısal koruma hedeflenmektedir (Hodge, 2005, s. 50). Her iki stratejinin de yetersiz yönleri

olduğunu belirten Küçük ve Alır (2003, ss. 353-354), teknoloji göçü stratejisi ile muhtemel veri kayıplarının önüne geçilememiş olmasına; kullanıcı gereksinimleri doğrultusunda korumaya dayalı öykünüm stratejisinin işlerliğine yönelik tartışmalarınsa halen sürdüğüne dikkati çekmektedirler.

Tonta (2001, s. 204) kayıtlı bilgilere erişim sağlama sorununa, sayısal bilginin üretilmesi, toplanması, depolanması ve iletimi için geliştirilmiş ileri bilgi teknolojileriyle çözüm sağlanacağı düşünülmemesi gerektiğini, bilgi erişim sorununa çözüm olarak düşünülmüş teknolojilerin, indeksleme ve sınıflama sistemlerinin dahi bu sorunu çözmekte yetersiz kaldığını belirtmektedir. Metin erişimi ile ilgili sorunların dahi henüz aşılamamış olması, akla çoklu ortamda bilgi erişiminin söz konusu olduğu durumlarda ne yapılması gerektiği sorusunu getirmektedir.

Web Neden Arşivlenmeli? Web’de Neler Arşivlenmeli?

İnsanın ürettiği maddi ve manevi nesnelerin toplamını kültür olarak tanımlayan Kongar (2006), her üretilenin ve her kuralın birer kültür ögesi olduğuna dikkati çekmekte ve fiziksel özellik taşıyanların maddi kültür; zihinsel ve ruhsal nitelikli üretilenlerin de manevi kültür olduğunu belirtmektedir. Basılı ve görsel kültür öğelerinin yanı sıra, sayısal ortamda üretilen kültür öğeleri de özellikle Internet’in hızlı gelişimiyle hayatımızda yoğun şekilde yer almaya başlamıştır. Lyman (2002), çok sayıda geçmiş kültür öğelerinin arşivlenmediğini, güvenliği ve koruması sağlanamadığı için günümüze ulaşmadığını, bu nedenle dinamik bir kültür hazinesi olarak Web’in arşivlenerek gelecek nesillerin hizmetine sunulması gerektiğini savunmaktadır.

Web, www, 3w şeklinde de kullanılan World Wide Web³ (Dünyayı Saran Ağ) kavramı ilk kez 1992’de kullanılmıştır. “Ağ” anlamına gelen web, gazete ve dergi gibi geleneksel yayıncılıktan daha kolay ve daha hızlı büyümektedir. Web’deki materyaller, yüzeysel / açık ve derin / sınırlı erişilebilenler olarak ikiye ayrılmaktadır. Yüzeysel web, web üzerinde bulunan ve arama motorları dizinlerinin eriştiği; içeriği yazı, grafik, ses ve video vb. değişik formatlardaki bilgilerden oluşmaktadır. Herkese açık olan bu siteleri herhangi bir kütüphane serbestçe indirebilir ve arşivleyebilir. Ancak, Internet’te bazı materyallere sınırlı erişimin olduğu göz ardı edilmemelidir. Derin web (deep web) veya görünmeyen web (invisible web) olarak adlandırılan bu alan, web üzerinde

bulunan ve arama motorlarında / dizinlerinde yer almayan belgelerin bulunduğu bölümdür. Derin web'in arşivlenmesi kolay değildir. Bu bölgelerin çoğuna arama motorlarının robotları girememekte ve buradaki dosyalara erişememektedir. Örneğin haber servislerinden CNN, robotların erişemediği önemli bir sitedir. Buradaki materyaller izinsiz indirilememektedir (Arms, Adkins, Ammen ve Hayes, 2001). Derin web'in yüzeysel web'den yaklaşık 550 kat daha büyük olduğunu belirten Lyman (2002), verilerin korunması amacıyla arşivlenmesinin kolay olmayacağını belirtmektedir.

Lyman (2002) içeriği her ay %40 yenilenen / değişen bir Web sitesinde sayfalara yönelik ortalama 15 bağlantı ve ses, görüntü gibi farklı formatlarda yaklaşık 5 kaynak bulunduğunu belirtmektedir. Söz konusu bağlantılar ve materyaller bir kullanıcının sorusuna cevap veriyorsa, o adres referans kaynağı olarak görülüp arşivlenmelidir. Bu bakış açısıyla web, kullanıcıların sorularına cevap bulabilecekleri her türlü bilgi kaynaklarını kapsayan bir referans kütüphanesi olarak da değerlendirilebilir. Başka bir ifadeyle, web arşivleri kişiye istediği adrese ulaşma imkanı vermektedir. Web'de sıklıkla "404 not found", "kırık bağlantı" ve "ölü bağlantı" olarak ta adlandırılan durumla karşılaşmaktadır. Bu sorun, sayısal materyalin dosya adının değişmesi, başka bir dizine taşınması, dosyanın bulunduğu sunucu ve bilgisayarın değişmesi veya dosyanın silinmesi ile ortaya çıkabilmektedir.

Milli ve derleme kütüphanelerinin birinci rolü şimdiki ve gelecek nesillere erişim imkanı vermek şartı ile her türlü bilgi materyalini toplamak, tanımlamak ve korumak olduğu bilinmektedir. Geleneksel kütüphane koleksiyonları; kitaplar, dergiler, gazeteler, haritalar, posterler, müzik eserleri ve broşürler gibi materyalleri kapsamaktadır. Günümüzde ise, kütüphane koleksiyonları çeşitlenmiş; mikrofilmleri, değişik film kaydedicileri, ses kasetlerini, videoları, bilgisayar disketlerini, CD ROM'ları ve DVD'leri içine alarak genişlemiştir. Ama bu materyallerin erişimi için özel araçların gerekli olması, bilgilerin saklandığı formatların modasının geçmesi ve bazen bilgilerin son derece hassas malzemeler üzerine kayıtlı oluşları, bu tür materyallerin kütüphanede bulunmasına yönelik tartışmaları beraberinde getirmiştir. Milli ve derleme kütüphaneleri mevcut koleksiyonlarını, 1993 yılından sonra Internet'in gelişmesiyle elektronik ortamda yayınlanan, her yerden kolayca erişilebilen ve diğerlerine göre pahalı olmayan yayın ve dağıtım bilgisini kapsayacak şekilde genişletmektedir (Phillips, 2005, ss. 57-58). Çoğu kütüphanenin, 5-6 yıl içinde daha iyi şekillenecek sayısal arşivleme alanında ciddi yatırımlar ve stratejik

planlamalar yaptıklarını ve bir çok ülkenin sayısal yayınları toplamak için eleman istihdamı yaptıklarını vurgulayan Phillips (2005, ss. 58-59), sayısal yayıncılığın geleneksel yayıncılıktan farklı bir paradigma olarak ortaya çıktığını, bunun sonucu olarak da milli kütüphanelerin 1990'lı yılların ortasında ulusal ve kültürel sayısal materyalleri arşivleme çalışmalarına başladıklarını belirtmektedir.

Sayısal yayıncılığın hızla büyüdüğü bir ortamda web arşivleme çalışmalarında nelerin arşivleneceği sorusuna cevap bulmak kolay olmayacaktır. Etraflarımızı saran bu yeni yayıncılık şekli, toplama, tanımlama, yükleme, saklama ve erişim gibi girişimleri zorunlu kılmıştır. Web ortamında yer alan yayınlardaki çeşitlilik ve web'in çok sayıda dosya ve karışık yazılımları içermesi, arşivleme stratejilerini zorlamaktadır.

Bu aynı zamanda, herkese kendi yayıncısı olma imkanı da sunmaktadır (Phillips, 2005, s. 58). Tonta (2001, s. 204), bilgi erişimi sorununu ele aldığı çalışmada, insanların ürettiği tüm ürünlerin depolanmasının "makro koruma" olarak tanımlandığını belirterek, üretilen her şeyin depolanmasının tartışılabileceğini, neyin depolanacağına karar vermenin de sorunu tam anlamıyla çözmeyeceğini vurgulamaktadır. Bu durumda web arşivleme stratejisinin koleksiyon hacmi ve seçimiyle ilişkili olduğu söylenebilir. Web arşivleme kapsamını belirlerken kişilerin özel yaşamlarına ait bilgilerin, özellikle alışveriş listelerinin, arşivlenmesi ve bunun erişime açılmasının tartışılmasına neden olacağı beklenebilir. Lyman (2002), Amazon.com'daki gibi, kişisel bilgilerin izin alınmadan arşivlenmemesi ve erişime açılmaması gerektiğini savunmaktadır. Bu durum sadece firma ile müşteri arasında yaşanacak bir sorun değil, aynı zamanda Internet'in gizliliği ve özel hayata müdahale anlamına da gelebileceği ileri sürülebilir.

Web Arşivleme Yaklaşımları

Web sitelerinin diğer sayısal kaynaklardan ayrılan en belirgin özelliği kısa süreli oluşlarıdır. Birleşik Krallık Web Arşivleme Konsorsiyumu 2004 yılında hedefini ortaya koyarken, "ortalama ömrü 44 gün olan web sitelerinin daha uzun ömürlü olabilmesini sağlamak" ifadesini kullanarak web arşivleme girişimlerinde uzun vadeli koruma hedeflerine dikkati çekmişlerdir (Press Release, 2004). Web arşivleme amacıyla geliştirilen yaklaşımların hepsinde, belirlenen hedefe yönelik, en iyi teknikle en çok web sitesinin arşivlenmesi

amacı yatmaktadır. Bu amaçla geliştirilen arşivleme yaklaşımları arasında *harmanlama (harvesting)*, *seçimli yaklaşım*, *tematik yaklaşım*, *derleme yaklaşımı* ve adı geçen yaklaşımların bir arada kullanıldığı *karma yaklaşım* bulunmaktadır.

Harmanlama (harvesting) terimi, İnternet arama motorlarının örümcekleri (crawler) aracılığı ile web sayfalarını yakalayıp indekslemelerini ifade etmek amacıyla veya web sayfaları üzerinden elektronik posta adreslerinin toplanarak reklam vb. nedenlerle yığın iletilerin (spam mail) gönderilmesi amacıyla kullanılmaktadır (Harvesting, 2005). Bu yaklaşımda, web örümcek (crawler) programları aracılığı ile daha önceden belirlenmiş koleksiyon kuralları dahilinde web siteleri izlenir ve içerikleri ağ üzerinden yüklenir. Web arşivleme girişimleri içinde *Internet Archive* ve İsveç Kraliyet Kütüphanesi'nin *Kulturarw3* projeleri, harmanlama yaklaşımını ilk kullanan çalışmalar arasında yer almaktadır. Örneğin, İsveç'in *Kulturarw3* örümceği, sadece İsveç alan adlı (.se) siteleri toplamak üzere programlanmıştır. İçeriği İsveç hakkında olup, farklı alan adlarına sahip olan web siteleri ise seçilerek koleksiyona dahil edilmektedir (Day, 2003 August). Web arşivlemede harmanlama yöntemini kullanan diğer ülkeler arasında Finlandiya, İzlanda, Norveç ve Avusturya milli kütüphaneleri yer almaktadır. Harmanlama yönteminin en belirgin avantajı, web sayfalarının belirli zaman aralıklarıyla otomatik olarak toplandığı sırada insan gücünün en az oranda kullanılması ve dolayısıyla elde edilen her bir parça için maliyetin az olmasıdır (Phillips, 2005, s. 61).

Seçimli yaklaşım, Web alanında tanımlanmış bölgelerin ya da belirli kaynakların önceden belirlenen ölçütler doğrultusunda arşivlenmesi amacını taşımaktadır. Söz konusu olan seçim, kaynakların önemi, kalitesi, konuları ya da ilgili web siteleri ile olan bağlantısı göz önünde bulundurularak yapılır (Day, 2003). Seçimli yaklaşımı kullanan web arşivleme girişimleri arasında Avustralya Milli Kütüphanesi'nin projesi olan PANDORA başı çekmektedir. İngiliz Milli Kütüphanesi, Kanada ve Japon milli kütüphaneleri de web arşivleme projelerinde seçimli yaklaşımı kullanan ülkeler arasında yer alır. Seçimli web arşivleme yaklaşımında, harmanlama yaklaşımında benimsenen belirli bir alan adına ait tüm web sayfalarını tamamen arşivleme mantığının aksine, belirli bir alanda (domain) yer alan web siteleri içinde kütüphanecinin entelektüel katkısının devreye girdiği bir seçim işlemi söz konusudur (Cunnea, 2005). Seçimli yaklaşımın avantajları arasında arşivdeki her bir parçanın kalitesinin onaylanması ve mevcut teknik kapasite dahilinde tüm içeriğin

kullanılabilmesinin yanı sıra, arşivdeki her bir nesnenin kataloglanmasına olanak sağlaması ve böylece ulusal bibliyografyanın bir parçası olabilmesi yer almaktadır. Aynı zamanda, web sitelerinin toplama planlarında göz önüne alınan güncelleme sıklığı vb. unsurların belirlenen her bir konuda içerik toplama işlemini tam olarak gerçekleşmesine olanak tanınması, seçimli yaklaşımın bir diğer avantajıdır (Phillips, 2005 June).

Web arşivlemede tematik yaklaşım, seçilmiş, belirli bir konuyla ilgili olduğu düşünülen URL adreslerinin belirlenmesi ve toplanması ilkesine dayanır. Tematik arşivlemede dikkatli bir şekilde seçilmiş bir dizi URL adresi, koleksiyon oluşturma işleminin çekirdeğini oluşturur. Bu URL adresleri, ilgili konudaki web sitelerinin ya da sayfalarının, belirli aralıklarla ve açıkca belirlenmiş kurallara göre örümcekler tarafından izlendiğini göstermektedir. Bu özelliği nedeniyle tematik arşivleme ile harmanlama yaklaşımı arasında benzerlik olduğu düşünülse de, harmanlama yaklaşımı, web örümcekleri aracılığıyla saptanan web adresleri arşivlenirken birbirleri arasında içerik ilişkisi aranıyor olmasıyla farkını ortaya koymaktadır. Tematik arşivlemede, Web örümceği programı, arşive dahil edilecek nesnelere, çekirdek koleksiyondaki bağlantıları izleyerek toplar. Tematik arşivlemeki bu mekanikleştirilmiş koleksiyon geliştirme politikasının sürekliliği çekirdek koleksiyonun Web'deki sürekliliğine bağlıdır (Schneider, Foot, Kimpton ve Jones, 2003, ss. 1-2).

Derleme yaklaşımında, web arşivleme çalışmalarının kanunlarla desteklendiği uygulamalar, özellikle, sayısal formda üretilmiş kaynakların denetimi hedeflenmektedir. Son yıllarda Avrupa ülkelerinde sayısal kaynakların derlenmesine yönelik çıkarılan kanunlar, sayısal kaynakların giderek artan önemini göstermektedir (Pennock ve Kelly, 2006). Söz konusu derleme kanunları halihazırda İsveç, Almanya, Hollanda gibi ülkelerin milli kütüphanelerince uygulanmaktadır. Bununla birlikte Birleşik Krallık Web Arşivi Konsorsiyumu (United Kingdom Web Archiving Consortium) da web arşivleme çalışmalarını derleme yaklaşımı çerçevesinde gerçekleştirmektedir. Web arşivlemede derleme yaklaşımı, tıpkı geleneksel derleme yaklaşımında olduğu gibi, web sitesi sahiplerinin / yöneticilerinin veya sayısal bilgi üreten yayıncıların, web sitelerinde yer alan içeriğin kopyalarını veya görüntülerini, belirlenen kanunun öngördüğü şekilde, arşive göndermesi ilkesine dayanır (Day, 2003).

Web arşivleme çalışması yapan kimi ülkeler, halihazırda kullanılmakta olan web

arşivleme yaklaşımlarından hiçbirinin web ortamındaki sayısal kültürün korunmasında yeterli olamayacağını düşünmektedirler. Bu nedenle, yeni yaklaşımlar geliştirmek yerine, bazı ülkeler tarafından mevcut yaklaşımların bir arada kullanılması fikri benimsenmiştir. Literatürde karma yaklaşımlar olarak anılan bu yaklaşım, web arşivlerinde mümkün olabilecek en kapsamlı koleksiyonu oluşturmak amacıyla harmanlama, seçimli, tematik ve derleme yaklaşımlarının en az ikisinin bir arada kullanılması şeklinde tanımlanmaktadır. Web arşivleme çalışmalarında bu yaklaşımı benimseyen ülkeler içinde Fransa, milli kütüphanesi ile öncü ülkedir (Day, 2003).

Web Arşivlerinin Korunması

Temel amacı, bilginin hızlı üretimi karşısında kaybolan bilgileri arşivleyerek koruma altına almak olan web arşivlerinde; depolama, yedekleme, teknoloji göçü, öykünüm, erişim stratejileri gibi arşiv güvenliği öne çıkmaktadır. Sayısal materyallere uzun süreli erişimi sağlamanın; yazılım ve donanımın iyi şekilde yönetimi ile metadata provenansı olarak adlandırılan sayısal içerikleri ve formatları depolama tekniğinin kullanılmasıyla sağlanabileceği bilinmektedir. Örneğin, Avustralya sayısal yayın arşivi PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)'da birinci görev, sayısal yayınlara uzun süreli erişim sağlamak olarak belirlenmiştir. Bilgisayar teknolojisi çok hızlı değişmekte, ara yüzler, teknik standartlar ve dosya yapılarının gelecek 20 yılda bugünden çok farklı olacağı düşünüldüğünde, sayısal yayınlara erişim için uygun teknolojik değişikliklerin yapılması gerekmektedir (Koerbin, 2004). Bu nedenle, web arşivleme çalışmalarında sayısal korumaya önem verilmektedir. Örneğin, PANDORA, OCLC (Online Computer Cataloging Center) merkezli metadata provenansını koruma çalışmalarına katılmaktadır (Arms ve diğerleri, 2001).

Lyman (2002), arşivlenecek nesnelerin korunması kapsamında alınacak ilk kararın nesnelerin metadata bilgilerinin kaydedilmesine yönelik olması gerektiğine dikkati çekmektedir. Web arşivlemede, web sitesindeki metadata alanları dikkate alınmaktadır. Amerika'da 2002 seçimi ve 11 Eylül olayları ile ilgili arşivlenen 2500 web sitesinin metadata alanları; yaratıcı veya yayıncı adı, başlık, öz, toplama tarihi, türü, fiziksel tanımlamaları, dil, konu başlığı gibi bilgileri kapsayacak şekilde kaydedilmiştir. Arşivleme sürecinde LCSH (Library of Congress subject Headings) kontrollü sözlüğü kullanılarak, arşivlenen materyallerin indeksleri bir makine yardımıyla kontrol edilmekte ve kayıt içinde

elle (manuel) ekleme ve çıkarma yapılabilmektedir. Web arşivlemede konu başlıkları kontrollü şekilde yapılmaktadır. PANDORA Web ana sayfasında konu başlıkları listelenmiştir. Seçimli yaklaşımı tercih eden PANDORA, arşivindeki tüm başlıkların MARC kayıtlarını, Avustralya Ulusal Bibliyografik Veri tabanı ve Milli Kütüphane çevrimiçi kataloglama kuralları çerçevesinde kataloglanmaktadır (Arms ve diğerleri, 2001).

Web arşivlemede karşılaşılabilecek sorunlara değinen Lyman (2005), nesnelerin yaşam döngüsünün tarihsel önemi üzerinde durulmasını *kültürel problemler*; yazılım, donanım, dosya formatı ve boyutu gibi arşivleme esnasında karşılaşılan sorunları *teknik problemler*; telif hakları ve derleme yasası gibi kanuni düzenlemeleri *yasal problemler*; teknolojik altyapı, ar-ge ve personel giderleri gibi sorunları ise, *ekonomik problemler* olarak göstermektedir. Web arşivlemede dokümanın bütünlüğünün ve provenansının korunarak uzun süreli depolamak ve etkili erişim öne çıkarken, depolama maliyetinde ucuzlama olduğu söylenebilir. Tonta (2001, s. 204), son yıllarda depolama kapasitelerinin artmasıyla beraber fiyatlardaki ucuzlamaya dikkat çekmektedir. Lyle (2004), 1 gigabytelik veri depolama maliyetinin 1992 yılında \$1000 (Dolar) seviyesindeyken, 2010 yılında bu rakamın 2 cent'e düşeceğini tahmin etmektedir. Arms (2001, s. 24), depolama maliyetlerinin iki kez kopyalamanın önüne geçilmesi ile düşeceğini, bir web sitesinin ortalama 60 Megabyte, 30.000 (otuzbin) web sayfasının 21.6 Terabyte, çift olanların ayıklanması ile 5 Terabytelik depolama alanının yeteceğini ve yıllık depolama maliyetinin 500.000\$ olacağını belirtmektedir.

Web Sayfalarının Arşivlenmesinde Yasal Düzenlemeler

Sayısal yayıncılık kütüphaneler için olduğu kadar yayıncı ve araştırma merkezleri için de gerçek bir paradigma kaymasına neden olmuştur. Bu çerçevede değişimi söz konusu kurumları yeni politikalar, yeni işletme modelleri, yeni altyapı ve teknoloji ürünleri geliştirmeye zorlamaktadır. Yaşadığımız dünya sayısal bir durum almış, yeni yazılım türleri, bilgisayar uygulamaları ve dosya türleri bize göz dağı verirken, özellikle uzun süreli erişim imkanı veren arşivleme çalışmaları öne çıkmıştır.

Sayısal bilgilerin en çok bulunduğu ortam olan web'i koruma yollarından biri toplamak / arşivini oluşturmaktır. Fakat toplamak bazen tek başına çözüm olmamakta, özellikle derin web, Intranet, elektronik postalar, haber servisleri gibi ulaşılması kolay olmayan alanlar bulunmaktadır. Sayısal yayınların, basılı

ortamdaki bilgi kaynaklarından farkı; içerik, basıldığı ortam, toplama, koruma ve erişim gibi unsurlardan kaynaklanmaktadır. Bunların yanı sıra geleneksel kaynaklardan ayrılan bir diğer özelliği ise uzun süreli erişime çözüm olarak görülen arşivlemeden kimin veya hangi kuruluşun sorumlu olacağına yönelik yapılan tartışmalardır. Yayıncıların sayısal yayınların arşivlenmesine yönelik tutumlarına dikkati çeken Tonta (2000, s.108), bazı yayıncıların kendi yayınlarını sayısal ortamda uzun süreli arşivlemekten kaçındıklarını belirtmekte, bu sorumluluğu yayıncıların üstü ve merkezdeki bir kuruluş olarak kütüphanelerin taşıması gerektiğini savunmaktadır. Sayısal yayınların arşivlenmesinden kimin sorumlu olması gerektiğine yönelik sunulan görüşler üç yaklaşımın egemen olduğunu göstermektedir. Bu yaklaşımlardan ilki Sayısal Bilginin Arşivlenmesi Çalışma Kolu (Task Force on Archiving of Digital Information) tarafından sunulan raporda yer aldığı üzere (1996, s.21) yayıncıların kendi yayınlarını korumaktan sorumlu oldukları yaklaşımdır. İkinci yaklaşım, İsveç, Avustralya ve Kanada gibi ülkelerin başlattığı ve ulusal çapta yürütülen çalışmalardır. Milli kütüphanelerin öncülüğünde yürütülen çalışmalar yasalarca desteklenerek belli bir politika dahilinde yürütülür. Bu yaklaşıma göre, derleme yasaları kapsamında ele alınan sayısal bilgilerin koruma görevi geleneksel yayınlar için bu misyonu taşıyan milli kütüphanelere düşmektedir. Üçüncü yaklaşım ise tıpkı The Internet Archive, IA (Internet Arşivi) örneğinde olduğu gibi kar amacı gütmeyen bir yapıda uluslararası çapta web'i arşivleme hedefine yönelik çalışmalar yapmak, çözümler üretmek üzerine kuruludur (Mannerheim, 1998).

Milli kütüphaneler geleneksel misyonlarını yerine getirmek, ülkelerinde yayımlanan yayınları toplamak için derleme kanununa gereksinim duymaktadır. Derleme yasalarında amaç ulusal kültür mirasını gelecek için korumaktır. Milli kütüphanelere derleme kanunu kapsamında basılı yayınların birer nüshası gönderilmektedir. Fakat e-yayınlar bu kapsama henüz dahil edilmemiştir. Kütüphanenin dijital bilgilerin korunması sorumluluğunu başarması için -ki bunların çoğu telif hakları ile ilgili olanlardır- yasal çalışmaların belirsiz olmaması gerekmektedir. Sayısal materyallerin telif hakları üzerindeki eleştirilere rağmen gelecek bilim kuşaklarına sağlayacağı fayda nedeniyle sayısal materyallerin arşivlenmesi tarihi bir görev olarak algılanmalıdır (Arms ve diğerleri, 2001). Ülkemizde yürürlükte olan Basma Yazı ve Resimleri Derleme Kanunu'nda "...Memleketin kültür hayatıyla alakası olduğu Maarif Vekilliğince tespit ve ilan olunan eserler de bu mecburiyete tabi tutulabilir" (Basma yazı...madde 2, 1934) ifadesi ile derleme kanununun genel çerçevesi belirlenmiştir. Yasada çerçeve geniş tutulmuş olmasına rağmen günümüzün

teknolojik gelişmeleri ve yayın formatlarındaki değişimler karşısında yeterli olduğunu söylemek olanaksızdır. Çoğu ülkenin derleme yasası bulunmakla beraber, derleme kanunlarının çoğu uzun yıllar önce sayısal yayınlar yokken hazırlanmıştır. Bu nedenle ülkeler, sayısal yayınları kapsayacak şekilde derleme kanunlarını gözden geçirmeye başlamışlardır.

Web arşivlemeye yönelik yasal düzenlemeler derleme kanunlarının yanı sıra telif hakları kapsamında da ele alınmaktadır. Sayısal kaynakların derlenmesinin, web arşivlemeyi ulusal düzeyde gerçekleştiren kurumların yasal hakkı olup olmadığı tartışılmaktadır. Gross (2003), Çin’de web arşiv çalışması başladığında telif haklarını büyük bir sorun olarak gördüklerini belirtmektedir. Çek derleme kanunu sayısal yayınları kapsamaması nedeniyle Çek Milli Kütüphanesi ile yayıncılar arasında olası bir telif hakkı ihlaline yönelik görüşmeler yapılmıştır. Bu görüşmeler sonucunda yayıncılar Dublin Core ve Metadata kayıtları konusunda işbirliğine ikna edilerek, metadata bilgilerini web sitelerine eklemişlerdir (Zabicka, 2003). Uluslararası düzeyde arşivleme yapan The Internet Archive, IA (İnternet Arşivi) sadece telif hakları açısından yaklaştığı web arşivleme çalışmalarında kişilerin entelektüel mülkiyet haklarına saygılı olduklarını, dürüst kullanım gereği kullanıcıların verdikleri hizmetten faydalanabileceklerini ifade etmektedir (Terms of Use, 2001).

Avustralya’da web arşivleme projesi 1996 yılında başlamasına rağmen, Avustralya Milli Kütüphanesi Derleme Kanunu hükümlerindeki gerekli değişiklikler için çalışmalar halen devam etmektedir (Legal Deposit, 2005). Amerika’da ise Kongre Kütüphanesi (Library of Congress) web sayfalarının toplanması ve korunmasına yönelik gerçekleştireceği yasal düzenlemeleri üç açıdan ele almıştır. Öncelikle, Kongre Kütüphanesi web sayfalarının indirilmesi ve korunmasında Telif Hakları Ofisi, Derleme Kanunu’ndan farklı olarak geniş kapsamlı bir yasa çıkarmayı düşünmüştür. Aynı zamanda web siteleri toplandıktan sonra erişim alanlarıyla ilgili düzenlenecek politikalara yönelik alternatifler üretilmiştir. Sonuç olarak, Kongre Kütüphanesi çalışanlarının yapmış olduğu çalışmalar sonucunda Telif Hakları Yasası Bölüm 407’de, web sayfalarının korunmasında ve erişiminde karşılaşılabilecek duruma göre küçük düzeltmeler yapma hakları gizli tutularak, İnternet üzerinden açık erişilen web sitelerini arşivleme izni verilmiştir (Arms ve diğerleri, 2001).

Ulusal Web Arşivleme Çalışmaları

Internet World Stats şirketinin yaptığı bir araştırmaya göre İnternet'te en çok kullanılan dillerin; İngilizce %30, Çince %13, Japonca %8,5, İspanyolca %7,9, Almanca %5,6, Fransızca %4, Korece %3,3, Portekizce %3,2, İtalyanca %2,8, Rusça %2,3 ve diğer diller %19 olduğu belirtilmiştir (Internet User by Language, 2006). Bu dağılıma göre İnternet'te İngilizce dışındaki dillerin %70 civarında yaygın olduğu görülmektedir. Her ülke web materyallerini ana dilinde yayınlamak İnternet'teki dil oranını yükseltmeyi hedeflemektedir. Söz konusu hedefin prestij ve kaygı eksensli olduğu söylenebilir. Ülkelerin ana dillerindeki web belgelerini toplama ve koruma çalışmalarının bu prestij ve kaygıların bir sonucu olduğunu söylemek mümkündür. Ulusal ölçekte yapılan çalışmalarda genelde dil, ulusal alan adları ve yaratıcıların kimlikleri temel alınmaktadır. Başlatılan projelerin ana dil merkezli ve milli kütüphane öncülüğünde yapıldığı dikkate değerdir. Avustralya, Avusturya, Fransa, İngiltere, Amerika, Danimarka, İtalya, İsveç, Kanada, Japonya, Litvanya, Çin, Finlandiya, İzlanda, Çek Cumhuriyeti, Yunanistan, Slovenya ve Yeni Zelanda ulusal ölçekte web arşivleme çalışmaları yapan ülkeler arasında yer almaktadırlar.

Avrupa ülkelerindeki milli kütüphanelere Ocak 2002 — Ocak 2003 tarihleri arasında web arşivi ile ilgili bir anket uygulanmıştır. 39 Avrupa ülkesine yollanan ankete 25 ülkenin milli kütüphanesi katılmıştır. Bu anketle, Avrupa ülkelerinden %65'inin web arşivleme çalışması başlattığı belirlenmiştir ve buna göre, 15 ülke bir web arşivleme çalışmasını aktif veya ciddi olarak test etmekte, 5 tanesi ise (Danimarka, İzlanda, Litvanya, Norveç, İsveç) derleme kanunlarında bu konuya yer vermektedir (Hallgrimsson ve Bang, 2003).

İsveç Kraliyet Kütüphanesi, Kulturarw³ web arşivleme projesiyle bu işe öncülük etmiştir. Kulturarw³ projesine 1996 yılında 3 milyon İsveç Kronu bütçe ayrılmış ve İsveççe “.se”, “.com”, “.net”, “.org” ve “.nu” uzantılı adreslerin toplanmasıyla işe başlanmıştır. İlk aşamada 25.000 “.se” uzantılı ve 6.000 değişik uzantılı olmak üzere İsveç dilinde oluşturulmuş toplam 31.000 web sitesinde yer alan 9,7 milyon dosyanın 25.000'i arşivlenmiştir. Harmanlama (Harvesting) yaklaşımının benimsendiği Kulturarw³ projesiyle arşivde yaklaşık 500 GB veri depolanmıştır. İsveççe web arşiv kapsamında yaklaşık yüz değişik formatta farklı dosya bulunmakla beraber bunlardan %95'inin HTML, JPEG,

³ Bu çalışmada *Web* kullanımı tercih edilecektir.

GIF ve metin tabanlı oldukları görülmektedir. Ayrıca, İnternet’te ulusal bibliyografya hazırlamak amacıyla web kaynaklarını kataloglamak ve arama hizmeti veren Svesök adlı proje başlatılmıştır. Svesök, verileri Kulturarw³’dan almaktadır. Alınan veriler Dublin Core formatında kataloglanmaktadır. Bununla beraber, İsveç Kraliyet Kütüphanesi, aralarında Belize, Kosta Rika, El Salvador, Guatemala, Honduras, Nikaragua ve Panama’nın bulunduğu yedi Orta Amerika ülkesinin web sayfalarını da toplamaktadır (Mannerheim, 1998).

Avustralya Milli Kütüphanesi, Avustralya ve Avustralyalılarla ilgili kültürel ve tarihi dokümanları hangi biçimde olursa olsun kütüphane koleksiyonuna katmak amacıyla 1996’da Preserving and Accessing Networked Documentary Resources of Australia (PANDORA) adlı çalışmayı başlatmıştır. PANDORA, sosyo-kültürel konularda ve politika, din, bilim ve ekonomi alanlarında sayısal ortamdaki bilgileri seçimli yaklaşımla toplayarak uzun süreli erişim sağlamak amacıyla kurulmuş Avustralya web arşivi olarak tanımlanmaktadır. Arşiv kapsamına, e-dergi, hükümet yayınları ve web siteleri girmektedir. PANDORA kapsamında seçilmiş sayısal yayınlar konu başlıkları altında arşivlenmekte, seçim, politikalar, uzun süreli erişim sağlama, sayısal yayınları koruma, teknolojik göç, risk yönetimi, konularında araştırmalar yapılmaktadır (Koerbin, 2004).

Tablo 1: PANDORA Arşiv Büyüklüğü ve Aylık Büyüme 26 Mart 2006 tarihli veri girişi⁴

	Şubat-2006	Mart-2006	1 Aylık Fark
Toplam Arşiv Başlığı	11,059	11,260	201
Toplam Arşiv Sayısı	21962	22461	499
Toplam Dosya sayısı	29,804,222	30,231,368	427,146
Toplam Gigabyte	1090.39	1116.4	26

Çek Cumhuriyeti Milli Kütüphanesi, web ortamındaki kültürel belgeleri toplamak ve korumak amacıyla 2000 yılında web arşivleme proje çalışmalarına

⁴ Statistics (2006). 1 Mayıs 2006 tarihinde <http://pandora.nla.gov.au/statistics.html> adresinden erişildi.

başlamıştır. Pilot çalışma, ülke alan uzantısı “.cz” olanların toplanmasıyla 2001 yılında tamamlanmış ancak yazılım sorunu nedeniyle 2 ay sonra ara verilen çalışma süresince 130 GB büyüklüğünde belge toplanmıştır. 2001 sonunda proje, sorunlar giderilmiş olarak tamamlanmış ve “WebArchive” adını almıştır. Web arşivleme bağlamında harmanlama (harvesting) yaklaşımının benimsendiği Web Archive’da 240 GB kapasiteli 10 milyon doküman bulunmaktadır ve bunların %54,2 si HTML, %24,4’ü JPG, %14,2’si GIF ve %3,2’si diğer biçim dosyalardan oluşmaktadır. Çek web arşiv çalışmalarını tam anlamıyla oturtmak için ülkedeki diğer bibliyografik çalışmalarla işbirliği yapılmıştır. Bu süreçte Çek ISSN ajansı ile beraber sayısal süreli yayımların bibliyografik veritabanı oluşturulmuş, kayıt numarası alan yayın aynı zamanda web arşive kaydedilmiştir (Zabicka, 2003).

Tablo 2: Web Arşivleme Projesi Olan Ülkeler

ÜLKE	Proje Adı	Strateji/ Model	Başlangıç Tarihi	URL Adresi
Almanya	Archivserver	Derleme	1998	http://deposit.ddb.de/
A.B.D.	Minerva	Tematik- Seçimli	2001	http://www.loc.gov/minerva
Avusturya	Pandora	Seçimli	1996	http://pandora.nla.gov.au
Avusturya	AOLA (The Austrian Online Archive)	Harmanlama	2001	http://www.ifs.tuwien.ac.at/~aola/
Çek Cumhuriyeti	Webarchiv	Harmanlama	2003	http://Webarchiv.nkp.cz
Çin	DACHS – Digital Archive For Chinese Studies	Harmanlama	2001	http://www.sino.uni-heidelberg.de/dachs/
Danimarka	Netarchive.dk	Harmanlama	2003	http://www.netarchive.dk/
Finlandiya	Eva	Harmanlama	1997	http://www.lib.helsinki.fi/finelib/english/index.html
Fransa	Gallica Digital Library	Harmanlama	1997	http://gallica.bnf.fr
Hollanda	Archipol	Derleme	2000	http://www.archipol.nl
İngiltere	UK Web Archiving Consortium	Seçimli	2004	http://www.Webarchive.org.uk/
İsveç	Kulturarw3	Harmanlama	1996	http://www.kb.se/kw3/
İzlanda	The National Web Archive	Harmanlama	1996	http://www.kb.dk/kb/proj/nwa/
Japonya	Web Archiving Project	Seçimli	2002	http://warp.ndl.go.jp/
Kanada	The Electronic Collection Of National Library Of Canada	Seçimli	1996	http://www.bibliat.gouv.qc.ca
Litvanya	Lithuanian Archive Of Electronic Resources	Harmanlama	2001	http://www.lnb.lt
Norveç	The Paradigma Project	Harmanlama	2001	http://www.nb.no/paradigma/
Slovenya	Slovenian Web Archiving	Harmanlama	2002	http://www.nuk.uni-lj.si/
Yeni Zelanda	Digital Collections	Seçimli	2002	http://www.natlib.govt.nz/en/digital/index.html

Uluslararası Web Arşivleme Çalışmaları

Uluslararası ölçekli web arşivleme çalışmaları, birkaç ülkenin ve/veya özel/tüzel kuruluşların bir araya gelmesiyle oluşturulan işbirlikleri aracılığı ile yürütülmektedir. Halihazırda yürütülmekte olan uluslararası projeler arasında The International Internet Preservation Consortium, Networked European Deposit Library Project ve The Internet Archive'ın yanı sıra Ülkeleri Web Arşivi (The Nordic Web Archive), ve Birleşik Krallık Web Arşivi Konsorsiyumu (UK Web Archiving Consortium) bulunmaktadır. Kuzey Ülkeleri Web Arşivi, Danimarka, Finlandiya, İzlanda, Norveç ve İsveç milli kütüphanelerinin aralarında bulunduğu bir işbirliği projesidir. Bu proje ile ülkeler web arşivleme konusundaki deneyimlerini paylaşmayı amaçlamaktadırlar. Birleşik Krallık Web Arşivi Konsorsiyumu, bilimsel, kültürel ve akademik kaynakların gelecek nesillere aktarılmasını sağlamak amacıyla oluşturulmuş bir konsorsiyumdur. Hak sahiplerinin izinleri ile seçimli web arşivleme çalışmalarını yürüten bu konsorsiyum, İngiliz Milli Kütüphanesi öncülüğünde yürütülmektedir. İnternet Arşivi (The Internet Archive, IA) diğer örneklerden farklı olarak kar amacı gütmeyen ticari bir ortaklığın sonucunda ortaya çıkmıştır. Adı geçen uluslararası çalışmalar arasındaki ortak noktayı, ülkelerin ulusal çapta yürüttükleri web arşivleme çalışmalarını uluslararası platformda geliştirme gayretleri şeklinde tanımlamak mümkündür. IA, bir yönüyle diğer uluslararası çalışmalardan ayrılmaktadır. IA'nın hareket noktasında web'i dünya çapında arşivlemek hedefi yatmaktadır. Bu çalışmada adı geçen uluslararası web arşivleme çalışmaları içinden İnternet Arşivi, Uluslararası İnternet Koruma Konsorsiyumu (International Internet Preservation Consortium, IIPC) ve Avrupa Derleme Kütüphaneleri Ağı Projesi (Networked European Deposit Library Project, NEDLIB) kısaca tanıtılmaktadır. Web üzerindeki bilginin ileriye dönük kullanımını sağlamak amacıyla başlatılan ilk sayısal koruma çalışması olma özelliğini taşıyan İnternet Arşivi, 1996 yılında hayata geçmiştir. Kar amacı gütmeyen bir girişim olan IA, 2001 yılında sahip olduğu koleksiyonu *The Wayback Machine* adındaki arayüz aracılığı ile serbest kullanıma açmıştır (Day, 2003). Web sayfalarının arşivlenmesinde harmanlama (harvesting) yaklaşımını benimseyen IA, 55 milyar sayfalık bir koleksiyona

sahiptir. Her geçen ay 20 terabyte büyümenin gerçekleştiği IA koleksiyonu, 2 petabyte büyüklüğündeki veriye erişim olanağı sağlamaktadır (Frequently Asked Questions, 2006). IA sadece sayısal ortamda oluşturulmuş bilginin değil, aynı zamanda sayısallaştırılmış basılı materyallerin korunması işlemini de üstlenmiştir. Bu amaçla aralarında kütüphanelerin de yer aldığı çeşitli organizasyonlarla işbirliğine gidilmiş, bibliyografik kaynakların sayısallaştırılması ve sayısal ortamda dağıtımını sağlamaya yönelik projeler üretilmiştir. Bu projeler arasında iki tanesi dikkati çekmektedir. Birincisi, Amerikan Kongre Kütüphanesi'nin desteği ile International Children's Digital Library adındaki sayısallaştırılmış çocuk kütüphanesidir. 225 adet çocuk kitabının sayısallaştırılmasıyla başlayan bu proje, 100'den fazla ülkeye ait 10.000'den fazla kitabın sayısallaştırıldığı bir projeye dönüşmüştür (Collection Development Policy, 2005). İkincisi ise, IA'nın İskenderiye Kütüphanesi ile yaptığı işbirliğidir. IA, içinde 2.000 saatlik Mısır ve Amerikan televizyon yayınlarının da yer aldığı 100 terabyte büyüklüğündeki verilerinin birer kopyasını İskenderiye Kütüphanesi'ne bağışlamıştır. Bunun yanı sıra Arapça kaynakların sayısallaştırılmasında kullanılan teknik donanımını da kütüphaneler arası ödünç verme yöntemi ile İskenderiye Kütüphanesi'nin kullanımına açmıştır (Internet Archive, 2002). Kullandığı ileri teknolojiye rağmen IA'nın yetersiz kaldığı bazı noktalar bulunmaktadır. Web'in 1992 yılından itibaren Internet'in bir parçası haline gelmesi nedeniyle IA, Internet'in ilk yıllarına yönelik toplama işlemi gerçekleştirememiştir. Buna rağmen az da olsa Usenet ve FTP arşivlerinden Web dışı materyaller toplanabilmiştir. Ayrıca, şifre korumalı veya IP tanımlı sitelerdeki erişimi kısıtlayıcı unsurlar bazı web sayfalarına erişimi, dolayısıyla bu sayfaların arşivlenmesini engellemektedir. Sahipsiz web siteleri de arşivlemede sorun olan diğer unsurlar arasında yer almaktadır. Alexa/IA tarafından geliştirilen teknolojiyle bir web sitesinden diğerine verilen linkler izlenmekte ve bir başka siteye link vermemiş olan siteler web'in geri kalanı tarafından erişilemez kabul edilerek arşive dahil edilmemektedir. Telif hakları ise, halen çözüm bekleyen konuların başında yer almaktadır. Birçok site sahibi, web sitelerinin arşivlenmesini engellemek amacıyla koydukları filtrelerle kendilerine ait bilgilerin arşivlenmesi ve halka açılmasıyla ilgili inisiyatifleri ellerinde bulundurmaktadır (Panos, 2003, s. 346).

IA'nın, web'de arşivleme çalışması yapan kütüphane kaynaklı diğer projelerden ayrılan belirgin özellikleri vardır. Bu özelliklerinden biri koleksiyonuna yönelik olarak herhangi bir indeksleme ya da kataloglama çalışması yapmamasıdır. IA'da sınıflandırma, web sayfalarının URL adreslerine göredir. Koleksiyona

erişimde kullanılabilir tek yol URL adresleridir. Aynı zamanda IA'da koleksiyondaki materyallere yönelik bibliyografik tanımlama yer almadığı gibi koleksiyona anahtar sözcüklerle erişim sağlayabilecek bir arayüz de bulunmamaktadır. IA'nın kütüphane destekli projelerden ayıran özelliklerinden bir diğeri de, koleksiyon geliştirme politikasının olmayışıdır. Hedefi, sahip olduğu teknoloji elverdiğince web'i arşivlemek olan IA'nın koleksiyonuna dahil ettiği içeriğe yönelik herhangi bir uygunluk, kalite, doğruluk denetimi olmadığı gibi, uygunsuz içeriği nedeniyle sansür uygulanmış web sitelerine yönelik herhangi bir denetim söz konusu değildir (Edwards, 2004, ss. 5-7).

Uluslararası İnternet Koruma Konsorsiyumu (The International Internet Preservation Consortium, IIPC), 24 Temmuz 2003 tarihinde, içinde Avustralya, Kanada, Danimarka, Finlandiya, Fransa, İzlanda, İtalya, Norveç, İsveç gibi ülkelerin milli kütüphanelerinin yanı sıra İngiliz Milli Kütüphanesi'nin, Amerikan Kongre Kütüphanesi'nin ve kar amacı gütmeyen bir oluşum olan IA'nın biraraya geldiği 12 katılımcı kuruluşla oluşturulmuş bir konsorsiyumdur (Halgrimsson, 2005). IIPC, üç aşamada ele aldığı hedeflerini İnternet ortamında yer alan dünya çapındaki bilgilerin arşivlenmesi, korunması ve zaman içinde erişilmesini sağlamak; uluslararası arşivler oluşturma hedefine yönelik ortak araçlar, teknikler ve standartlar geliştirmek ve milli kütüphanelerin web'de arşivleme yapmalarını özendirici çalışmalarda bulunmak şeklinde sıralamıştır (About the Consortium, 2006).

Fransa Milli Kütüphanesi başkanlığında yürütülen IIPC'de uygulamalar, Web arşivlemede politikalar, araçlar, standartlar ve yöntemler geliştirmek amacıyla oluşturulmuş çalışma grupları tarafından yürütülmektedir. Konsorsiyumda altı çalışma grubu bulunmaktadır. Her bir üye ülkenin çalışma gruplarından en az birinde yer alma zorunluluğu vardır. Bu çalışma grupları, ortak teknik yapı, araştırmacı gereksinimleri, erişim araçları, ölçümler, web örümceklerinin değerlendirilmesi, derin web ve içerik yönetimi başlıkları altında oluşturulmuşlardır (Henriksen, 2005). Web arşivlemede kullanılan erişim araçları ve teknik altyapılara yönelik mevcut durumun ortaya konmasının yanı sıra IIPC'ye özgü oluşturulmuş arayüzler, metadata setleri, derin web'in keşfinde kullanılabilir alternatif yöntemler geliştirilmiştir (Henriksen, 2005). IIPC'nin kısa süreli hedefleri arasında web arşivleme çalışmalarında kullanmak üzere, web arşivlemedeki temel uygulamaları kapsayan bir araç geliştirmek yer almaktadır. Geliştirilecek bu aracın özellikleri arasında seçime odaklı ve sağlama sürecini destekleyen, harmanlama (harvesting) veya derleme

yaklaşımlarından birini kullanacak, koleksiyon depolama ve yaşatım (maintenance) fonksiyonuna sahip, son kullanıcıya erişim olanakları sunabilmesi hedeflenmektedir. Bununla birlikte, deneme amaçlı geliştirilmiş ve IIPC standartlarının oluşturulmasına imkan sağlayan araçlar da bulunmaktadır. Bunlardan biri, IA ve İskandinav ülkelerinin kütüphanelerince 2003 yılının başlarında geliştirilmiş, geniş kapsamlı harmanlama yapabilen, *Heritrix* isimli araçtır. Ayrıca Web`de uzun süreli koruma ve erişim için arşivlenecek materyalin dermeye aktarılmadan önce farklı bir formata dönüşmesini sağlamak amacıyla Fransa Milli Kütüphanesi`nce *DeepArc* isminde bir araç geliştirilmiştir (Lupovici, 2005).

1 Ocak 1998 tarihinde düzenlenen Avrupa Milli Kütüphaneler Konferansı ile hayata geçen Avrupa Derleme Kütüphanesi Ağı Projesi (Networked European Deposit Library Project, NEDLIB) projesi, aynı yıl Avrupa Komisyonu Telematik Uygulamalar programınca tahsis edilen bir fonla desteklenmiş ve 2000 yılında tamamlanmıştır. Hollanda Milli Kütüphanesi başkanlığında yönetilen NEDLIB, Fransa, Norveç, Finlandiya, Almanya, Portekiz, İsviçre ve İtalya'nın dahil olduğu bir üye profiline sahiptir. Sekiz ülkenin milli kütüphanelerinin yanı sıra, NEDLIB projesinin katılımcıları arasında, bir milli arşiv, iki bilgi iletişim teknolojileri organizasyonu ve Kluwer Academic, Elsevier Science ve Springer-Verlag gibi yayımcılar da bulunmaktadır (Factsheet, 2001).

NEDLIB projesinin başlatılmasındaki temel amaç, Avrupa milli kütüphanelerinin sayısal yayınların yönetiminde ve yaşatımında ortaklaşa hareket etmeleridir. NEDLIB projesi ile, sayısal yayınların derlenmesine yönelik sistem geliştirmeyi sağlayacak ortak bir teknik altyapının oluşturulması ve temel araçların geliştirilmesi hedeflenmektedir. Bu noktada milli kütüphanelerin önemi aynı zamanda derleme kütüphanesi olmalarından ileri gelmektedir. Projenin çalışma alanı kapsamında milli kütüphanelerin sayısal belgelere yönelik derlemelerini geliştirirken karşılaştıkları belli başlı teknik ve yasal konular yer almaktadır (Werf-Davelaar, 1999).

NEDLIB projesi ile hedeflenen, aralarındaki derleme politikası farklılıklarına ve farklı coğrafi konumlarına rağmen, derleme kütüphaneleri arasında ortak bir yaklaşımın benimsenebileceğini ve derleme kütüphanelerindeki sayısal yayınlara yönelik iyi bir sistemin geliştirilebileceğini göstermektedir. Bu amaçla sayısal yayınlara yönelik ortak bir model ve terminoloji geliştirmeyi hedefleyen

NEDLIB, *Reference Model for Open Archival Information System*'i temel alarak ve International Standard Organisation (ISO), standartları doğrultusunda *Process Model for Deposit System (PMDS)* adında bir sistem geliştirmiştir. NEDLIB tarafından geliştirilen bu model, sayısal yayımlara uygulanan tüm aşamaları (derleme, depolama, koruma ve erişim) ayrıntılı bir şekilde ele almaktadır (Steenbakkers, 2000, s. 3). PMDS'in yanı sıra NEDLIB projesi kapsamında Web kaynaklarının harmanlanmasında ve arşivlenmesinde kullanılmak üzere *NEDLIB Harvester* adında bir yazılım geliştirilmiştir. Helsinki Üniversite Kütüphanesi ve *Center for Scientific Computing* bu uygulamanın yaşatımını birlikte sağlamaktadır. Aynı zamanda *MMB System for Multimedia Access* adında çoklu ortamda erişimi sağlamaya yardımcı bir sistem geliştirilmiştir. Bu sistem, sayısal kaynakların iş akışında kullanılacak bütünleşik bir istemci/sunucu ortamıdır (Beagrie, 2003, s. 42). NEDLIB projesi ile ortaya konan en önemli sonuç, derleme sisteminin, sayısal kütüphane ortamı içinde, kendine has özellikleriyle tasarlanması gerektiğinin anlaşılmasını sağlamak olmuştur. Derleme kütüphanelerinin sayısal kaynakların depolanmasında ve korunmasında başlı başına farklı tasarlanmış bir sisteme sahip olmasının gerekçelerinin başında, diğer kütüphane sistemlerince uygulanan sağlama, kataloglama, tarama ve erişim faaliyetlerinin derleme sisteminin bir parçası olmamasıdır (Steenbakkers, 2000, s.3).

Sonuç

Web arşivleme, sayısal bilgilerin korunması ve erişime açılması için tek başına çözüm değildir. Dinamik bir büyüme karşısında atılmış büyük bir hamledir. Özellikle bilgilerin uzun süreli erişimini sağlamak için, depolarken veri kaybını önlemek, teknolojik göçe uygun yazılım geliştirmek, içerik teknoloji bütünleşmesi sorununa çözüm üretmek gerekecektir. Web arşivleri, sayısal kütüphane olarak gelecekte en çok tercih edilecek bilgi erişim noktaları olacaktır. Verilere erişim için sadece URL adresleriyle değil, anahtar kelime, konu başlıkları gibi metadata alanlarından bir kaçında arama yapma imkanı veren yazılım ve arayüzlere gereksinim duyulacaktır. Belirli web sitelerinin derleme kapsamına alınması, yayıncılarla görüşülerek metadata bilgileri ve e-derleme çalışması stratejileri belirleme çalışmalarına önem verilmelidir. İnternet'in tüm dünyayı birleştiren, sınırları kaldıran özelliği bu süreci işbirliği noktasına taşımakta, sorunlar herkesi ilgilendirdiğinden çözümleri de beraber aranmaktadır. Ülkemizin de bu tür işbirliklerinde yer alması, özellikle Milli Kütüphane'nin bu konudaki sorumluluğunu yerine getirmesi gerekmektedir. Bu

süreçte bilgi kaynaklarının düzenlenmesi, korunması ve kullanıma sunulmasını temel misyon edinen kütüphanecilerin bu konuya eğilmeleri, yayınlar, çalıştaylar yaparak konuyu gündemde tutmaları ve proje kapsamında öneriler geliştirmeleri beklenebilir.

Kaynakça

- About IA* [Internet Arşivi Hakkında]. (2006). 25 Mart 2006 tarihinde <http://www.archive.org/about/about.php> adresinden erişildi.
- About the Consortium* [Konsorsiyum Hakkında]. (2006, Ocak). 28 Şubat 2006 tarihinde <http://netpreserve.org/about/index.php> adresinden erişildi.
- Al, U. ve Küçük, M. E. (2003). Üst veri standartları ve uygulamaları. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 20 (1), 167-185.
- Arms, W.Y. (2001). Web preservation project interim report: A report to the library of congress. 2 Şubat 2005 tarihinde <http://www.cs.cornell.edu/wya/LCweb/interim.doc> adresinden erişildi.
- Arms, W.Y., Adkins R., Ammen C. ve Hayes A. (2001). Collecting and preserving the web: the minerva prototype. *RLG Diginews*, 5 (2). 15 Mart 2006 tarihinde <http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1> adresinden erişildi.
- Basma Yazı ve Resimleri Derleme Kanunu (1934). 02 Nisan 2006 tarihinde <http://www.kultur.gov.tr/teftis/BelgeGoster.aspx?F6E10F8892433CFF03077CA1048A18348F72AE2C955C7C52> adresinden erişildi.
- Beagrie, N. (2003). National digital preservation initiatives: An overview of development in Australia, France, the Netherlands, and the United Kingdom, and of Related International Activity. 13 Nisan 2006 tarihinde <http://www.clir.org/PUBS/reports/pub116/pub116.pdf> adresinden erişildi.

- Bullock, A. (1999). Preservation of digital information: issues and current status. 16 Ocak 2006 tarihinde <http://www.collectionscanada.ca/9/1/p1-259-e.html> adresinden erişildi.
- Chowdhury, G.G. ve Chowdhury, S. (2003). *Introduction to digital libraries*. London: Facet Publishing.
- Collection Development Policy*. [Koleksiyon Geliştirme Politikası]. (2005, 7 Mart). 28 Mart 2006 tarihinde ICDL web sitesinden erişildi: <http://www.childrenslibrary.org/about/policies/collection.html>
- Cunnea, P.(2005, Spring). Selective web archiving in the UK : a perspective of the National Library of Scotland within UK Web Archiving Consortium (UKWAC). *Sconul Focus*, 34. 1 Mart 2006 tarihinde http://www.sconul.ac.uk/pubs_stats/newsletter/34/11.pdf adresinden erişildi.
- Day, M. (2003, August). *Preserving the fabric of our lives: a survey of Web preservation Initiatives*. Paper presented at the ECDL 2003: 7th European Conference on Research and Advanced Technology for Digital Libraries, Trondheim, Norway. 11 Mart 2006 tarihinde <http://www.ukoln.ac.uk/metadata/presentations/ecdl2003-day/day-paper.pdf> adresinden erişildi.
- DLM Forum. (1997). Guidelines on best practices for using electronic information. 10 Mart 2006 tarihinde <http://europa.eu.int/ISPO/dlm/documents/gdlines.pdf> adresinden erişildi.
- Edwards, E. (2004). Ephemeral to Enduring: The Internet Archive and Its Role in Preserving Digital Media. *Information Technology and Libraries*, 23(1), 3-8.
- Factsheet*. [Gerçek bilgiler]. (2001,11 Mart). 1 Nisan 2006 tarihinde NEDLIB web sitesinden erişildi: <http://nedlib.kb.nl/index.html>

- Frequently Asked Questions* [Sıkça Sorulan Sorular]. (2006). 10 Nisan 2006 tarihinde Internet Archive web sitesinden erişildi:
<http://www.archive.org/about/faqs.php> adresinden erişildi.
- Gross, J. (2003). *Learning by doing: the digital archive for chinese studies (DACHS)*. Paper presented at the 3rd. ECDL Workshop on Web Archives, Trondheim, Norway. 20 Nisan 2006 tarihinde
<http://bibnum.bnf.fr/ecdl/2003/index.html> adresinden erişildi.
- Gulli, A. ve Signorini, A. (2005). The indexable web is more than 11,5 billion pages. 5 Nisan 2006 tarihinde <http://www.cs.uiowa.edu/~asignori/web-size/> adresinden erişildi.
- Hallgrímsson, T. ve Bang, S. (2003, August). *Nordic web Archive*. Paper presented at the 3rd ECDL Workshop on Web Archives, Trondheim, Norway. 20 Nisan 2006 tarihinde
<http://bibnum.bnf.fr/ecdl/2003/index.html> adresinden erişildi.
- Halgrímsson, T. (2005). Special presentation: The International internet preservation consortium (IIPC). 1 Nisan 2006 tarihinde
<http://consorcio.bn.br/cdnl/2005/HTML/Presentation%20Thorsteinn%20Halgrímsson.htm> adresinden erişildi.
- Harvesting* [Harmanlama]. (2005). 15 Şubat 2006 tarihinde CORDIS web sitesinden erişildi: http://lu.com/odlis/odlis_h.cfm
- Hellerstein, J. (2004). Data on the deep web: queries, trawls, policies and countermeasures. 30 Mart 2006 tarihinde
http://www.citrisuc.org/research/projects/data_on_the_deep_web_queries_trawls_policies_and_countermeasures adresinden erişildi.
- Henriksen, B.N. (2005). The International internet preservation consortium (IIPC). 5 Nisan 2006 tarihinde
<http://consorcio.bn.br/cdnl/2005/PRESENTATIONS/NL%20Iceland.rtf> adresinden erişildi.
- Hodge, G. (2005). Preservation of and permanent access to electronic information resources: A system perspective. *Information Services and Use*, 25, 47-57.

Internet Archive. (2002). The bibliotheca alexandriana: A truly digital library for 21st century/ Internet archive: towards universal access to all human knowledge. 28 Şubat 2006 tarihinde Internet Archive web sitesinden erişildi: http://www.archive.org/about/bibalex/bibalex_pr.pdf

Internet User by Language [Dillere Göre Internet Kullanıcıları]. (2006, 31 Mart). 21 Nisan 2006 tarihinde Internet World Stats web sitesinden erişildi: <http://www.Internetworldstats.com/stats7.htm>

Kaynak, E. A. (2001). Elektronik kaynakların bibliyografik kontrolü ve Metadata *Bilgi Dünyası*, 2(2), 188-205.

Koerbin, P. (2004, September). *The Pandora digital srchiving system (PANDAS) managing web archiving In Australia: A case study*. Paper presented at the 4 th International Web Archiving Workshop, Bath, UK. 13 Nisan 2006 tarihinde <http://www.nla.gov.au/nla/staffpaper/2004/koerbin2.html> adresinden erişildi.

Kongar, E. (1996). kültür ve bilim insanının sorumluluğu. 12 Nisan 2006 tarihinde http://www.kongar.org/makaleler/mak_kulb.php adresinden erişildi.

Küçük, M. E. ve Alır, G. (2003). Dijital koruma (arşivleme) stratejileri ve bazı uygulama örnekleri. *Türk Kütüphaneciliği*, 17(4), 340-356.

Lyle, J.A. (2004, September) Sampling the Umich.edu Domain.” Paper presented at the 4th International Web Archiving Workshop, Bath, UK. 02 Nisan 2006 tarihinde <http://www.iwaw.net/04/proceedings/Lyle.pdf> adresinden erişildi.

Legal Deposit [Yasal Depolama]. (2005, 11 Kasım). 12 Mart 2006 tarihinde PANDORA web sitesinden erişildi: <http://pandora.nla.gov.au/legaldeposit.html>

Lupovici, C. (2005). *Web archives long term access and interoperability: the International Internet Preservation Consortium Activity*. Paper presented at the World Libraries and Information Congress: 71th IFLA General Conference and Council, Oslo, Norway. 16 Mart 2006 tarihinde http://www.ifla.org/IV/ifla71/papers/194e_Lupovici.pdf adresinden erişildi.

- Lyman, P. (2002 April). Archiving the world wide web. *Building a national strategy for preservation: Issues in digital media archiving* içinde. Washington D.C.: Council on library and information resources. 5 nisan 2006 tarihinde <http://www.clir.org/pubs/reports/pub106/web.html> adresinden erişildi.
- Lyman, P. ve Varian, H. (2003). How much information? 12 Nisan 2006 tarihinde <http://www.sims.berkeley.edu:8000/research/projects/how-much-info-2003/> adresinden erişildi.
- Mannerheim, J. (1998, October). *Problems and opportunities of web archiving, towards the background of experiences from the Kulturarw3 Project*. Paper presented at the Nordic Conference on Preservation and Access, Stockholm. 12 Mart 2006 tarihinde <http://affisch.kb.se/ENG/Mannheim.htm> adresinden erişildi.
- Miller, P. (1999). Metadata for the Masses. *Ariadne*. 20 Şubat 2006 tarihinde <http://www.ariadne.ac.uk/issue5/metadata-masses/> adresinden erişildi.
- Panos, P. (2003). The Internet Archive: an end to the digital dark age. *Journal of social Work Education*, 39(2), 343:-347.
- Pennock, M. ve Kelly, B. (2006). Archiving web site resources: A records management view. 8 Nisan 2006 tarihinde <http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/WWWpp100-pennock.pdf> adresinden erişildi.
- Phillips, M. (2005). What should we preserve? The question for heritage libraries in a digital World. *Library Trends*, 54(1), 57- 71.
- Phillips, M. (2005, June). Selective archiving of web resources: A Study of acquisition costs at the national library of Australia. *RGL Digi News*. 22 Şubat 2006 tarihinde http://www.rlg.org/en/page.php?Page_ID=20666#article0 adresinden erişildi.

Press Release [Basın Açıklaması]. (2004, 21 Haziran). 10 Mart 2006 tarihinde UK Web archiving consortium web sitesinden erişildi:
<http://info.webarchive.org.uk/pressrelease21-06-04.html>

Schneider, S., Foot, K., Kimpton, M., Jones, G. (2003, August). *Building thematic web collections: challenges and experiences from the september 11 web archive and the election 2002 web archive*. Paper presented at the ECDL 2003: 7 th European Conference on Research and Advanced Technology for Digital Libraries, Trondheim, Norway. 11 Mart 2006 tarihinde
<http://bibnum.bnf.fr/ECDL/2003/proceedings.php?f=schneider> adresinden erişildi.

Stanescu, A. (2004, November). Assessing the durability of formats in a digital preservation: the Inform methodology environment, *Dlib Magazine*, 10 (11). 10 Mart 2006 tarihinde
<http://www.dlib.org/dlib/november04/stanescu/11stanescu.html> adresinden erişildi.

Steenbakkens, J. (2000). The NEDLIB Guidelines. 10 Nisan 2006 tarihinde
<http://nedlib.kb.nl/results/NEDLIBguidelines.pdf> adresinden erişildi.

Sullivan, D. (2004). Search engine size wars V erupts. 12 Nisan 2006 tarihinde
<http://blog.searchenginewatch.com/blog/041111-084221> adresinden erişildi.

Terms of Use [Kullanım Kuralları]. (2001, 10 Mart). 28 Ocak 2005 tarihinde Internet Archive web sitesinden erişildi:
<http://www.archive.org/about/terms.php>

The Cedars Project (2002). “Cedars guide to digital collection management” [March 2002] Elektronik adres:
<http://www.leeds.ac.uk/cedars/guideto/collmanagement/guidetocolman.pdf>
[7 Nisan 2006]

Tonta, Y. (2000). Elektronik yayıncılıkta son gelişmeler. *Bilgi Dünyası*, 1(1), 89-132.

Tonta, Y. (2001). Bilgi erişim sorunu. T. Fenerci ve O. Gürdal (Yay. Haz.), “21. Yüzyıla Girerken Enformasyon Olgusu”: *Ulusal Sempozyum Bildirileri, 19-20 Nisan 2001, Hatay* içinde (ss. 198-206). Ankara: Türk Kütüphaneciler Derneği.

Tonta, Y. (2005). Internet and electronic information management. *Information Services and Use*, 25, 3-12.

Werf-Davelaar, T. (1999). Long term preservation of electronic publications, *D-Lib Magazine*, 5(9). 1 Nisan 2006 tarihinde <http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html> adresinden erişildi.

Zabicka, P. (2003, August). *Archiving the czech web: issues and challenges*. Paper presented at the 3 rd ECDL Workshop on Web Archives Trondheim, Norway. 12 Mart 2006 tarihinde <http://bibnum.bnf.fr/ECDL/2003/> adresinden erişildi.

