



How to find the public Web Pages with broken links to historic public Pages and Files

http://farbe.li.tu-berlin.de/WBM_find_PFs_16.pdf or
http://130.149.60.45/~farbmetrik/WBM_find_PFs_16.pdf

1. Summary

Web pages with former *public Pages and Files (PFs)* sometimes disappear in the internet, for example based on a reorganization of the internet servers with new URLs or other reasons. This paper shows how to use the public *INTERNET ARCHIV: Way Back Machine (WBM)* to find the *PFs*. The URL <http://web.archive.org/web> helps to find many of these *PFs* which may have disappeared within the last 20 years. If a *WBM* request shows the intended *PFs*, then it is usually possible to download the *archive* page or file. If the file has the format zip then the search file is often a folder.

2. Introduction

For the application of standard documents of the Standard Organizations ISO, IEC, CIE, DIN, and others it is often necessary to find the basic documents which are referenced. Often one can not buy these documents and there may be no help by a Standard Organization to get these files. A solution is often a request to the *WBM* which usually takes a few seconds in the internet. According to the top page of the *WBM* more than 279 billion web pages are saved over time and are freely available. The links below lead to files which have been public between the year 1998 and 2017 and may not been *public* any more.

Since the year 2000 the author of this paper was responsible for three different internet servers which are all still *public* on the web and are called Servers 1 to 3 in the following:

<http://www.ps.bam.de>

<http://130.149.60.45/~farbmetrik>

Server 1, archive since 2008, and still active

Server 2, active since around 2008

<http://farbe.li.tu-berlin.de>

Server 3, active since around 2012

The Server 1 is at the *Federal Institute of Materials Research and Testing (BAM)* (about 38 GB, 260.000 items), the Servers 2 and 3 are at the *Berlin University of Technology (TUB)* at different departments by security reasons, and have approximately the same content (about 150 GB, 400.000 items, only parts are crawled up to now).

In addition the author has produced since 2005 contributions to many annual, meeting, and liaison reports on the Servers 4 and 5 of CIE Division 1 *Vision and Colour*.

<http://www.cie.co.at/div1>

Server 4, archive since 2009 and still active

<http://div1.cie.co.at>

Server 5, active since about 2009

http://div1.cie.co.at/?i_ca_id=544

Server 5s, sub page active since about 2009

In addition the author has proposed since 1999 contributions to many annual, meeting, and liaison reports on the Servers 6 and 7 of CIE Division 8 *Image Technology*

<http://www.cie.co.at/div8>

Server 7, archive since 2009 and still active

<http://www.colour.org>

Server 7o, active between 1999 and 2009, closed

<http://div8.cie.co.at>

Server 8, active since about 2009

http://div8.cie.co.at/?i_ca_id=616

Server 8s, sub page active since about 2009

For an easier understanding of the following the **direct internet links are in blue colour** (usually active) and the **WBM internet links are in red colour** (may not be active and/or deleted). In addition there are combinations, for example

http://web.archive.org/web/*/http://broken_link_PLs (test if page is available or not by the WBM).

The character * (star) is usually a crawl date and time. One may write for example only the year 2013. If many files of this name are crawled since the year 1998, then the file crawled nearest to

the date 2013 is downloaded and shown.

For example the red link <http://www.colour.org> is not working because the Server 70 seems to be deleted. However, the combined link is working. This shows a click on the following URL:

http://web.archive.org/web/*/http://www.colour.org/

The *WBM* output shows that the page <http://www.colour.org> has been crawled 88 times between the dates 19990422 and 20110627.

For example the blue link <http://www.ps.bam.de> is working because the Server 1 is an archive (no change since 2008) and this archive server is still active in the internet. Therefore a blue colour is used here. In addition the combined link is working. This shows a click on the following URL:

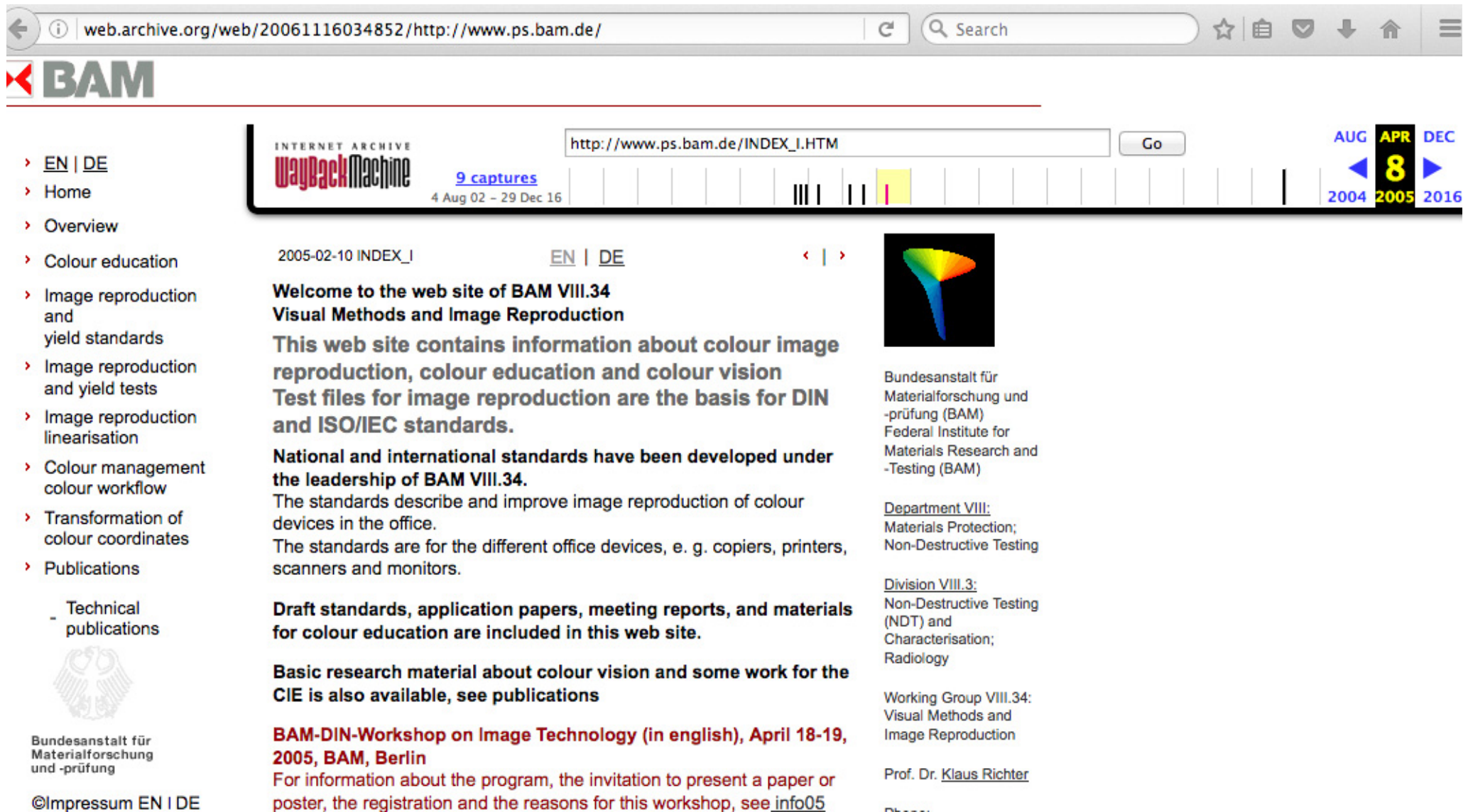
http://web.archive.org/web/*/http://www.ps.bam.de/

The *WBM* output shows that the page <http://www.ps.bam.de> has been crawled 54 times between the dates 20010202 and 20160516.

There is a calendar below the output of the combined links with a * (star) between the two links (see the example above). The calendar allows to choose a crawl date. For example the user may choose the calendar of 2006 and a special available crawl date 20061116. Then the output of a top page of the Server 1 is shown at his date.

Therefore the *WBM* output of the following link with the crawl date 20061116 works

<http://web.archive.org/web/20061116/http://www.ps.bam.de/>



web.archive.org/web/20061116034852/http://www.ps.bam.de/

BAM

INTERNET ARCHIVE
WaybackMachine

9 captures
4 Aug 02 – 29 Dec 16

http://www.ps.bam.de/INDEX_I.HTM

Go

AUG APR DEC
2004 2005 2016

2005-02-10 INDEX_I EN | DE

Welcome to the web site of BAM VIII.34
Visual Methods and Image Reproduction

This web site contains information about colour image reproduction, colour education and colour vision
Test files for image reproduction are the basis for DIN and ISO/IEC standards.

National and international standards have been developed under the leadership of BAM VIII.34.
The standards describe and improve image reproduction of colour devices in the office.
The standards are for the different office devices, e. g. copiers, printers, scanners and monitors.

Draft standards, application papers, meeting reports, and materials for colour education are included in this web site.

Basic research material about colour vision and some work for the CIE is also available, see publications

BAM-DIN-Workshop on Image Technology (in english), April 18-19, 2005, BAM, Berlin
For information about the program, the invitation to present a paper or poster, the registration and the reasons for this workshop, see [info05](#)

Bundesanstalt für Materialforschung und -prüfung (BAM)
Federal Institute for Materials Research and -Testing (BAM)

Department VIII:
Materials Protection;
Non-Destructive Testing

Division VIII.3:
Non-Destructive Testing (NDT) and
Characterisation;
Radiology

Working Group VIII.34:
Visual Methods and
Image Reproduction

Prof. Dr. [Klaus Richter](#)

©Impressum EN | DE

Figure 1: WBM crawl output of the server 1 at the date 20061116

Figure 2 shows the output of the server 1 (<http://www.ps.bam.de>) stored by the WBM at the date

20061116. This output is identical, if we input the URL (compare top line in Fig. 2):

<http://web.archive.org/web/20061116034852/http://www.ps.bam.de/>

However, usually we do not know the data of the date and the time of the crawl. An alternate input with only the year 2006

<http://web.archive.org/web/2006/http://www.ps.bam.de/>

leads to the date 20070101 which is nearby. This *WBM* output seems identical to Figure 1.

We realize that a link with a * (star) between the two links produces the amount of crawls during a time period, for example between 2001 and 2016. If we use instead of the * (star) a year or a date (20161116) or a date and time (20161116034852), then the *WBM* output shows the design of the web page for this crawl date and time or for the nearest crawl date and time.

Now we look more carefully at the *WBM* output, see Fig. 1, for the URL

<http://web.archive.org/web/20061116034852/http://www.ps.bam.de/>

The output of the *WBM* allows to navigate within the output. This navigation is usually identical to the navigation of the Server 1 <http://www.ps.bam.de> at the date 20061116. The red colour is used here to indicate that the content at 20061116 is shown by the *WBM*. The blue colour of <http://www.ps.bam.de> will produce the content of today, please click the blue URL for this output. For example the logo of the BAM is different now.

The output of the *WBM* allows to navigate between all the links on the Server 1 at any crawl date. The Server 1 has 60.000 links (*PFs*) which have not changed since the archiving date 2007. As *summary* we have seen here that the *WBM* preserves the internet knowledge for future generations.

The links to the 10 web pages of the Servers 1 to 8 on the pages 1 and 2 of this paper are useful to produce for example one or more special CIE pages which allow to download all former **public**

1. CIE Meeting Minutes
2. CIE Annual Reports, and
3. CIE Reportership Reports

of the CIE Divisions 1 and 8 between 1999 and today. At present this information is distributed over many of the 10 listed web pages at the beginning.

The present web pages with this information of CIE Division 1

http://div1.cie.co.at/?i_ca_id=544

and of Division 8

http://div8.cie.co.at/?i_ca_id=616

include only a limited information back for about the last 5 years compared to the *WBM* historic material of about 20 years.

The development of a CIE Technical Report and/or Standard usually takes more than 5 years. Therefore the origin and the reason for the creation of a CIE Technical Committee is often not any more visible at the present CIE web pages, for example the creation of the present CIE TC1-76 *Unique Hue Data* within the Minutes of the CIE Division meeting in Budapest 2009, see

<http://web.archive.org/web/20130512060855/http://files.cie.co.at/523.pdf>

According to these minutes the reason for this TC1-76 is based on the presentation and publication of the Reportership Report CIE R1-47 *Hue Angles of Elementary Colours*, see

<http://web.archive.org/web/20130426090142/http://files.cie.co.at/526.pdf>

Often an author likes to reference again a former public paper by an URL.

As examples two URLs of two **public** papers in the **standard reference format MLA** are given in the following:

Thorstein Seim, Reportership Report CIE R1-47, Hue angles of elementary colours, 2009, see (35 pages),

[<http://files.cie.co.at/526.pdf>]._Internet Archive_.[<http://web.archive.org/web/20160304130704/http://files.cie.co.at/526.pdf>].

Thorstein Seim, Reportership Report CIE R1-57, Border between Blackish and Luminous Colours, 2013, see (23 pages),

[http://files.cie.co.at/716_CIE%20R1-57%20Report%20Jul-13%20v.2.pdf]._Internet Archive_.[http://web.archive.org/web/20150413002133/http://files.cie.co.at/716_CIE%20R1-57%20Report%20Jul-13%20v.2.pdf].

The author is responsible that at least the second URL (in brackets) is available at the date of the presentation or publication of his paper.

The first URL output may be for example newer compared to the second one on the *WBM* and/or the first URL may be broken. Usually there are many second URLs if the paper has been public for many years which is true for the two examples. The first one was public since 2009 and the second one since 2013. The link to both was deleted at the present public CIE web site during September 2016.

Author: *Klaus Richter*, Berlin University of Technology (TUB), 2016

internet:<http://130.149.60.45/~farbmetrik>

<http://farbe.li.tu-berlin.de>