1999

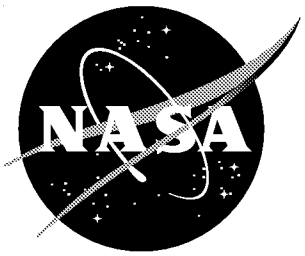
A Digital Library for the National Advisory Committee for Aeronautics

Michael L. Nelson
Old Dominion University

Original Publication Citation

Nelson, M. L. (1999). A digital library for the National Advisory Committee for Aeronautics. *NASA Technical Memorandum: 209127*. Hampton, VA: NASA Langley Research Center

NASA/TM-1999-209127



A Digital Library for the National Advisory Committee for Aeronautics

Michael L. Nelson
Langley Research Center, Hampton, Virginia

April 1999

The NASA STI Program Office ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***

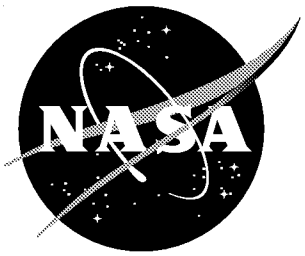
- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/TM-1999-209127



A Digital Library for the National Advisory Committee for Aeronautics

Michael L. Nelson
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

April 1999

Available from:

NASA Center for AeroSpace Information (CASI)
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161-2171
(703) 605-6000

A Digital Library for the National Advisory Committee for Aeronautics

Michael L. Nelson
NASA Langley Research Center
MS 158, Hampton, VA 23681
m.l.nelson@larc.nasa.gov

April 1999

Abstract

We describe the digital library (DL) for the National Advisory Committee for Aeronautics (NACA), the NACA Technical Report Server (NACATRS). The predecessor organization for the National Aeronautics and Space Administration (NASA), NACA existed from 1915 until 1958. The primary manifestation of NACA's research was the NACA report series. We describe the process of converting this collection of reports to digital format and making it available on the World Wide Web (WWW) and is a node in the NASA Technical Report Server (NTRS). We describe the current state of the project, the resulting DL technology developed from the project, and the future plans for NACATRS.

1 Introduction

The National Aeronautics and Space Act of 1958 created the National Aeronautics and Space Administration (NASA) from the National Advisory Committee for Aeronautics (NACA). As NASA's predecessor organization, NACA was chartered in 1915 and was operational from 1917 until 1958. NACA played an integral role in the development of the United States' fledgling aeronautics industry as the main research body for a collection of federal, commercial and university interests.

The main product of NACA's research was its multi-tiered report series. Although the exact number of NACA reports published is unknown, most estimates place this number between 20,000 and 30,000. This collection of work remains in high demand even today, especially in the areas of general aviation and the basic fundamentals of flight [11]. Unfortunately, although significant collections of NACA documents exist at a handful of NASA centers, universities and other government and industrial research laboratories, no single library contains a complete collection. Even what constitutes a complete NACA corpus is subject to debate. Furthermore, because of their age, high circulation, and acid-based paper, many of these reports are in poor condition and will cease being serviceable in the near future. Conversion to digital format is necessary for preservation as well as for wider dissemination.

This paper discusses the ongoing digital conversion of the NACA collection, begun in 1995, and the dissemination of this collection over the World Wide Web. We present the

structure and technology of the NACA Technical Report Server (NACATRS), the digital library (DL) that serves the NACA collection. We discuss the resulting technology from this project and the future work for NASA and NACA DLs.

2 Contents and Access

The NACATRS can be accessed via the WWW at: http://naca.larc.nasa.gov/. The NACATRS currently has over 1800 documents and an accession rate of about 30 documents a week. Figure 1 shows the NACATRS interface.

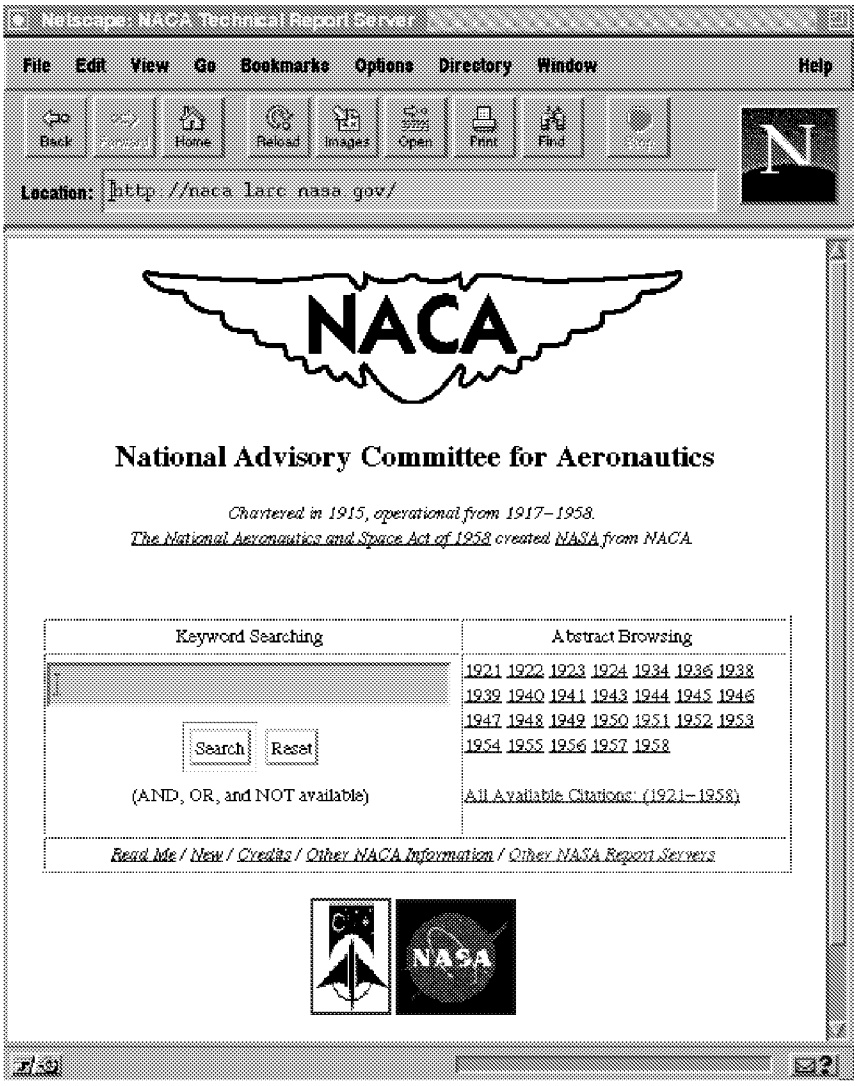


Figure 1: The NACA Technical Report Server

2.1 NACA Contents

NACA published in a variety of internal report series. Currently, the NACATRS holds the following NACA publications series:

- NACA Reports – NACA reports were considered to be the final and complete documentation on a subject or project and they often superceded one or more other NACA publication types. NACA Reports are sometimes (erroneously) referred to as "NACA Technical Reports."

- NACA Technical Notes (TNs) – Technical Notes were the basic unit of the research report series. Some early TNs were translations of foreign works.

- NACA Technical Memorandums (TMs) –TMs are translations of foreign works. The TM series probably replaced translations in the TN series.

- NACA "Wartime Reports" – Reports produced specifically for World War II research, they were declassified after the conflict. Due to their urgent nature, they frequently received little editing when written, and no editing was done after they were declassified. The moniker "Wartime Reports" was added when they were declassified; previously they were issued as Advance Confidential Reports (ACRs), Advance Restricted Reports (ARRs), Restricted Bulletins (RBs) and Confidential Bulletins (CBs).

- NACA Research Memorandums (RMs) – RMs were initially restricted, and represented initial or limited scope results, and thus received less editing and preparation than other report series.

NACATRS currently does not include:

- NACA Annual Reports – Annual Reports were simply the concatenation of a single year's NACA Reports (i.e., excluding TNs, TMs, etc.). Inclusion of the NACA Reports in NACATRS implicitly includes the Annual Reports.

- Aircraft Circulars – Reports published in the 1920s-1930s that reviewed the design and performance of contemporary aircraft (one AC per vehicle).

- Conference or Journal Preprints – We are unaware of how many items this would include. However, their content would likely be covered in a Report or TN, so this exclusion is probably negligible.

- Books – No books by NACA authors are included.

We can provide some generalized observations just from handling the collection. The first NACA Reports were issued in 1917, but TNs and TMs did not appear until 1920. The early publications were often either translations from European aeronautics works or authored by universities or other federal or military research laboratories. This is because NACA was initially truly a committee of aeronautically interested organizations rather than a federal agency in present context. As NACA acquired its own staff and developed its own research facilities, the number of publications authored by non-full-time NACA staff decreased.

2.2 WWW Contents

The NACA publications are scanned, and TIFF images are the output of the process. Optical Character Recognition (OCR) is not being applied for the NACATRS, primarily because the format of the NACA publications are often pages of equations, tables, charts and figures – none of which are well suited for OCR. Instead, the report is converted into a combination of GIF and PDF files for easier WWW dissemination.

NACATRS offers browsing and keyword searching of its holdings. These DL functions are provided using the TRSkit software package [7]. Although NACATRS is a "free-standing" DL, it is also a node in the NASA Technical Report Server (NTRS) [6], a DL that offers keyword searching to over 20 different DLs hosted by various NASA centers, institutes and projects. Although browsing and keyword searching is available, the reports are also accessible via the following naming convention:

```
http://naca.larc.nasa.gov/reports/YEAR/naca-REPORTTYPE-NUMBER
```

So the popular NACA Report 1135 is available at:

```
http://naca.larc.nasa.gov/reports/1953/naca-report-1135/
```

2.3 Document Presentation in NACATRS

One reason previous attempts at NACA archives [1] received limited usage was due to non- or limited-WWW interfaces, so we made every attempt to provide an intuitive, attractive WWW interface to the DL and individual reports. Figure 2 shows the presentation of a report. The thumbnail images are clickable, and will present a large GIF image for easy on-line viewing. Should the user desire to download the entire report for local storage or printing, the entire report in a single PDF file is available.

Currently, 10 thumbnails will be shown at a time with options such as "next", "previous", "first" and "last" being available to paginate through large reports. When viewing single pages (large GIFs), there are also similar pagination commands allowing you to step through the report on-screen one page at a time (Figure 3).

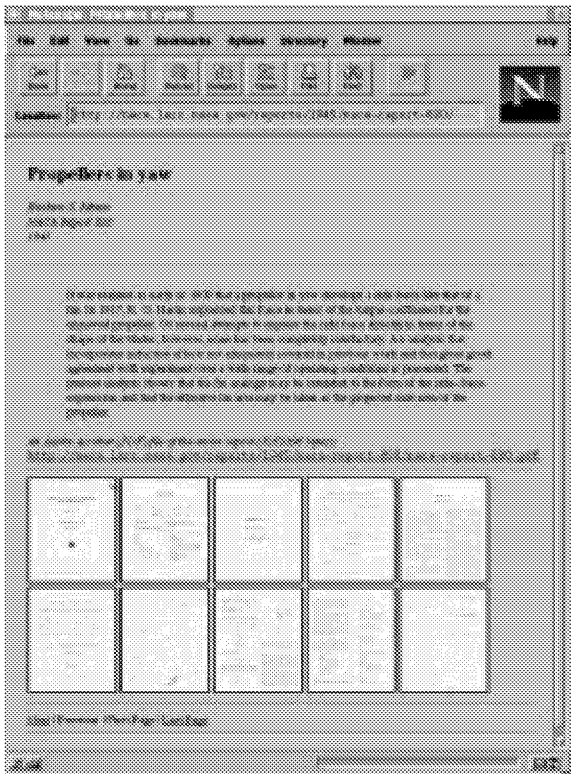


Figure 2: Initial Presentation of a Report

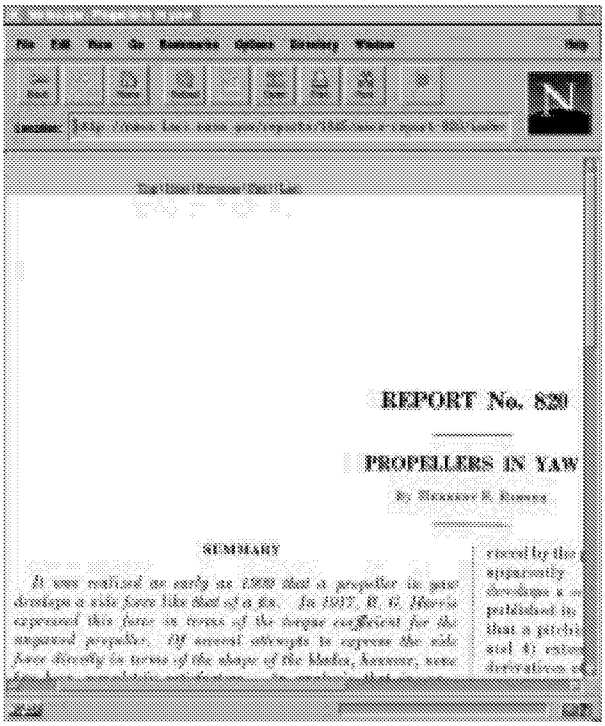


Figure 3: An Expanded View of a Single Page

We feel this method of presentation is superior to serving just PDF files. In NACATRS, PDF files are available, but in addition the pagination of the thumbnail and full size images allows for quick review and access to the document without having to download the entire report (which is often many megabytes).

Additionally, this method allows for the integration of the data files (PDF, GIFs, etc.) with the metadata in a single location. For example, although not explicitly advertised to the user, the structured metadata for NACA Report 1135 is available at:

```
http://naca.larc.nasa.gov/reports/1953/naca-report-1135/naca-report-1135.refer
```

This offers long term advantages in DL maintenance and operation because the metadata and data collocated. This also allows for the reports in NACATRS to be indexed by more than one DL, since the data and structured metadata are publicly available. Given the root of the NACA collection, a WWW robot or gatherer could construct an index to this collection.

3 Preparation

Though many nodes in NTRS are populated by capturing the electronic source files involved in the publishing process, it is obvious given the age of the NACA reports that electronic source is not an option. All the reports must be scanned. Given the poor condition of the some of the reports, finding suitable candidates for scanning is non-trivial. Initially, the NACATRS was fed by reports that were hand scanned at 400 dots per inch (dpi) by volunteers to help populate the prototype. Currently, the bulk of the reports are scanned at 300 dpi under a contract with the Phillips Research Site of the Airforce Research Laboratory. NASA Langley ships Phillips photocopies of the reports to be scanned, and they ship back tapes of TIFF images.

3.1 DTRTWT

Although TIFF is the output of the report scanning process, TIFF is not a widely accepted format for distribution on the WWW. We developed the DTRTWT (Does The Right Thing With TIFF) software package to process the TIFF into a more WWW-friendly presentation. Given the following input:

- 1 report = 1 directory
- 1 page = 1 TIFF file (all files residing in the report's directory)

DTRTWT produces the following output in the report directory:

- 1 PDF file for the entire report
- 1 large GIF image per page
- 1 thumbnail GIF image per page
- 1 Perl 5.0 index.cgi file to negotiate access and presentation to the report directory

DTRTWT expects a metadata file (in "refer" format [5]) to be present to extract values for the title, authors, abstract, etc. into the index.cgi file. We currently generate these metadata files in a semi-automated fashion. This greatly slows down the growth of NACATRS, but allows us to correct many errors and omissions present in current NACA bibliographic databases.

DTRTWT is a collection of Perl programs and shell scripts. For image processing, it uses both the commercial package "Image Alchemy" [2] and the freeware package "ImageMagick" [3]. DTRTWT runs under Unix, and the Sun Sparcstations and IBM RS/6000s we use process about one page per minute. The processing load can be spread across several machines in parallel.

Figure 4 shows a report being converted using DTRTWT. The TIFF pages are numbered sequentially, collected in a directory. DTRTWT is invoked with arguments for creating the PDF files, converting the TIFFs to GIFs, resize the GIFs to big and thumbnail sizes, and creating the index.cgi to manage the collection. The refer metadata can be added later. The TIFFs can be stored in the directory or removed to conserve storage.

```
% ls naca-tn-3301
naca-tn-3301/0001.tif*  naca-tn-3301/0009.tif*  naca-tn-3301/0017.tif*
naca-tn-3301/0002.tif*  naca-tn-3301/0010.tif*  naca-tn-3301/0018.tif*
naca-tn-3301/0003.tif*  naca-tn-3301/0011.tif*  naca-tn-3301/0019.tif*
naca-tn-3301/0004.tif*  naca-tn-3301/0012.tif*  naca-tn-3301/0020.tif*
naca-tn-3301/0005.tif*  naca-tn-3301/0013.tif*  naca-tn-3301/0021.tif*
naca-tn-3301/0006.tif*  naca-tn-3301/0014.tif*  naca-tn-3301/0022.tif*
naca-tn-3301/0007.tif*  naca-tn-3301/0015.tif*  naca-tn-3301/0023.tif*
naca-tn-3301/0008.tif*  naca-tn-3301/0016.tif*
% dtrtwt -pdf -convert -resize -index naca-tn-3301
% ls naca-tn-3301
0001-t.gif        0007-t.gif        0013-t.gif        0019-t.gif
0001.gif          0007.gif          0013.gif          0019.gif
0001.tif*         0007.tif*         0013.tif*         0019.tif*
0002-t.gif        0008-t.gif        0014-t.gif        0020-t.gif
0002.gif          0008.gif          0014.gif          0020.gif
0002.tif*         0008.tif*         0014.tif*         0020.tif*
0003-t.gif        0009-t.gif        0015-t.gif        0021-t.gif
0003.gif          0009.gif          0015.gif          0021.gif
0003.tif*         0009.tif*         0015.tif*         0021.tif*
0004-t.gif        0010-t.gif        0016-t.gif        0022-t.gif
0004.gif          0010.gif          0016.gif          0022.gif
0004.tif*         0010.tif*         0016.tif*         0022.tif*
0005-t.gif        0011-t.gif        0017-t.gif        0023-t.gif
0005.gif          0011.gif          0017.gif          0023.gif
0005.tif*         0011.tif*         0017.tif*         0023.tif*
```

Figure 4: Report Conversion with DTRTWT

3.2 Storage Requirements

Since the NACA collection is a fixed size, it allows us to make a number of simplifying assumptions in implementation. The Unix file system is used to store the current reports, and we expect that this will be sufficient to store the entire NACA collection. Scanned at 300 dpi, each page consumes roughly 15KB. After processing by DTRTWT, the total of all the resulting pages is about 80KB per page.

If we assume the entire NACA collection is 30,000 reports, and each report has 20 pages, then the NACA collection is equivalent to 600,000 pages. At 80KB per page, this translates into about 48GB of storage required for the entire collection. This could be stored comfortably on 6 9GB disk drives, which have a current retail price of less than $1,000. The files are backed up to the Distributed Mass Storage System (DMSS) [10], so if a drive fails the affected reports can be restored from DMSS.

4 Usage

Interest has been high in the NACATRS. On a monthly basis, NACATRS disseminates over 5,000 PDF files. This does not count browsing the GIFs. There are over 3,000 searches per month issued from the NACATRS page, and approximately 10,000 searches a month issued from NTRS. For comparison, NTRS handles over 30,000 searches per month, so as much as 1/3 of the NTRS searches involve NACATRS.

Anecdotally, we frequently receive email from users world wide requesting specific reports to be added and thanking us for making available a collection that to many university and industry users is not ordinarily available.

5 Future Work

The most obvious area for improvement is increasing the number of reports available. At just over 1200, NACATRS is large enough to be interesting, but with as many as 30,000 NACA reports, it is far from a canonical collection. However, there are opportunities to expand the NACATRS holdings beyond simply reports, especially airfoil data.

5.1 NACA Information

We receive many emails requesting access to NACA airfoil ordinates. We currently do not have this information. We also do not have any of the tabular data in non-scanned form. There are programs to generate NACA airfoil data [4], and the NACATRS even has links to Java programs written by non-NASA personnel to generate airfoil data. There are also programs to generate and refine data in handbook-type NACA reports [12]. However, there has been no concentrated effort to generate a database of numerical data to complement the NACA collection.

5.2 Buckets

An anticipated infrastructure upgrade is the switch to *buckets*. Buckets are aggregative, intelligent archival objects developed for DLs developed by the NASA Langley Research Center / Old Dominion University DL Research Group for the NCSTRL+ project [8]. In fact, the inspiration for buckets grew out of our work on the NACATRS, so we frequently refer to the current implementation of reports in NACATRS as "proto-buckets". Buckets allow storage of many different data objects as a single entity within a DL. Buckets also negotiate access and presentation of their contents. This corresponds to the index.cgi in a NACA report directory controlling the presentation of the PDF and the GIFs. Additional features of buckets such as intelligence, object level terms and conditions, and mobility are further described in [9]. Although the user is unlikely to notice specific changes, we intend to upgrade the "proto-buckets" to fully implemented buckets for smoother operation of the NACATRS and to continue refining our bucket concepts and implementation.

6 Conclusions

We have created a digital library, the NACA Technical Report Server, to capture, preserve and disseminate the in-demand, rare and decaying report collection of the National Advisory Committee for Aeronautics. NACATRS currently provides WWW access to over 1800 NACA publications. The publications are available in PDF for local storage and printing, in thumbnail GIFs for quick review, or large GIFs for easy on-screen reading. These formats are generated from the original TIFF images using the DTRTWT software package developed to create NACATRS. The NACATRS is available as a standalone DL, or as a node in NTRS. NACATRS serves approximately 5000 PDF reports per month, and handles over 13,000 monthly keyword searches. Future plans include providing access to the full NACA collection (approximately 30,000 documents), integrating numerical databases and software into the collection and the conversion to "buckets," aggregative and intelligent agents for DLs that were inspired by the original work on the NACATRS.

Acknowledgements

The NACATRS could not have proceeded without the assistance of many people. We would like to thank: Nola Graves (NASA LaRC) and Jim Coon (formerly of NASA LaRC) for metadata entry and population; Michael Nemeth (NASA LaRC) for lending a large collection of hard copy NACA reports to be scanned; Sharon Jenkins (NASA LaRC) , Zylphia Wilkins (NASA LaRC) and Donnell Lassiter (NASA LaRC) for their work in scanning Nemeth's collection; William Kleb (NASA LaRC) for scanning the reports he directly used in his research; Eloise Johnson (NASA LaRC), Henry Woolard (retired aerospace engineer) and Yvonne Kellog (NASA DFRC) for assistance in helping in our understanding the scope, purpose and terminology of the NACA collection; and David Adams (formerly of NASA LaRC), George Roncaglia (NASA LaRC) the NASA Langley Technical Library and the Code R Unilibrary Project for providing scanned reports that form the bulk of NACATRS.

References

1. M. Y. Ambur, D. L. Adams, P. P. Trinidad, "Desktop Access to Full-Text NACA and NASA Reports: Systems Developed by the NASA Langley Technical Library," AIAA 97-0467, *Proceedings of the 35th Aerospace Sciences Meeting & Exhibit*, Reno NV, January 6-10, 1997.

2. Image Alchemy, http://www.handmadesw.com/his/specs.html

3. ImageMagick, http://www.wizards.dupont.com/cristy/ImageMagick.html

4. C. L. Ladson, C. W. Brooks Jr., A. S. Hill and D. W. Sproles, "Computer Program to Obtain Ordinates for NACA Airfoils," NASA TM-4741, December 1996.

5. M. E. Lesk, "Some Applications of Inverted Indexes on the UNIX System," Bell Laboratories Computing Science Technical Report 69, 1978.

6. M. L. Nelson, G. L. Gottlich, D. J. Bianco, S. S. Paulson, R. L. Binkley, Y. D. Kellogg, C. J. Beaumont, R. B. Schmunk, M. J. Kurtz & A. Accomazzi, "The NASA Technical Report Server," *Internet Research: Electronic Networking Applications and Policy*, 5(2), 1995, pp. 25-36.

7. M. L. Nelson and S. L. Esler, "TRSkit: A Simple Digital Library Toolkit," *Journal of Internet Cataloging*, 1(2), 1997, pp. 41-55.

8. M. L. Nelson, K. Maly, S. N. T. Shen, & M. Zubair, "NCSTRL+: Adding Multi-Discipline and Multi-Genre Support to the Dienst Protocol Using Clusters and Buckets," *Proceedings of Advances in Digital Libraries 98*, Santa Barbara, CA, April 22-24, 1998, pp. 128-136.

9. M. L. Nelson, K. Maly, S. N. T. Shen & M. Zubair, "Buckets: Aggregative, Intelligent Agents for Publishing," *Webnet Journal*, 1(1), 1999, pp. 58-66.

10. J. Z. Pao and D. C. Humes, "NASA Langley Research Center's Distributed Mass Storage System," *Proceedings of the Third Goddard Space Flight Center Conference on Mass Storage Systems and Technologies*, College Park, Maryland, October 19-21 1993, pp. 207-216.

11. R. Smith, "The NACA Collection: When Old is New Again – A Status Report on the NACA Documents Project," *Sci-Tech News*, 46(2), April 1992, pp. 15-20.

12. D. W. Witte and K. E. Tatum, "Computer Code for Determination of Thermally Perfect Gas Properties," NASA TP-3447, September 1994.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1999	Technical Memorandum

4. TITLE AND SUBTITLE
A Digital Library for the National Advisory Committee for Aeronautics

5. FUNDING NUMBERS

6. AUTHOR(S)
Michael L. Nelson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L-17844

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA/TM-1999-209127

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 82 Distribution: Standard
Availability: NASA CASI (301) 621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
We describe the digital library (DL) for the National Advisory Committee for Aeronautics (NACA), the NACA Technical Report Server (NACATRS). The predecessor organization for the National Aeronautics and Space Administration (NASA), NACA existed from 1915 until 1958. The primary manifestation of NACA's research was the NACA report series. We describe the process of converting this collection of reports to digital format and making it available on the World Wide Web (WWW) and is a node in the NASA Technical Report Server (NTRS). We describe the current state of the project, the resulting DL technology developed from the project, and the future plans for NACATRS.

14. SUBJECT TERMS
Digital Libraries, NACA, WWW, Scanning

15. NUMBER OF PAGES
15

16. PRICE CODE
A03

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL