1998

Creating a Canonical Scientific and Technical Information Classification System for NCSTRL+

Melissa E. Tiffany

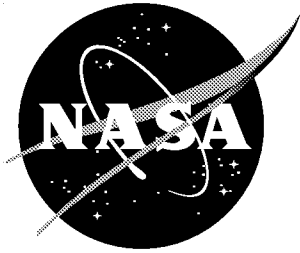Michael L. Nelson
Old Dominion University

Repository Citation

Tiffany, Melissa E. and Nelson, Michael L., "Creating a Canonical Scientific and Technical Information Classification System for NCSTRL+" (1998). *Computer Science Faculty Publications*. 30.
https://digitalcommons.odu.edu/computerscience_fac_pubs/30

Original Publication Citation

Tiffany, M.E., & Nelson, M.L. (1998). Creating a canonical scientific and technical information classification system for NCSTRL+. *NASA Technical Memorandum: 208955*. Hampton VA: NASA Langley Research Center.

NASA/TM-1998-208955

Creating A Canonical Scientific and Technical Information Classification System for NCSTRL+

Melissa E. Tiffany
Computer Sciences Corporation, Hampton, Virginia

Michael L. Nelson
Langley Research Center, Hampton, Virginia

December 1998

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 counter-part 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.

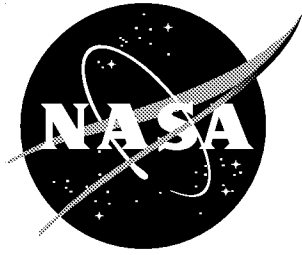- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- 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 help round out 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 Access Help Desk at (301) 621-0134

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

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

NASA/TM-1998-208955

Creating A Canonical Scientific and Technical Information Classification System for NCSTRL+

Melissa E. Tiffany
Computer Sciences Corporation, Hampton, Virginia

Michael L. Nelson
Langley Research Center, Hampton, Virginia

December 1998

Creating a Canonical Scientific and Technical Information Classification System for NCSTRL+

Melissa E. Tiffany
Computer Sciences Corporation
NASA Langley Research Center
MS 157D
Hampton, VA 23681
m.e.tiffany@larc.nasa.gov

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

Abstract

The purpose of this paper is to describe the new subject classification system for the NCSTRL+ project. NCSTRL+ is a canonical digital library (DL) based on the Networked Computer Science Technical Report Library (NCSTRL). The current NCSTRL+ classification system uses the NASA Scientific and Technical (STI) subject classifications, which has a bias towards the aerospace, aeronautics, and engineering disciplines. Examination of other scientific and technical information classification systems showed similar discipline-centric weaknesses. Traditional, library-oriented classification systems represented all disciplines, but were too generalized to serve the needs of an STI oriented digital library. Lack of a suitable existing classification system led to the creation of a lightweight, balanced, general classification system that allows the mapping of more specialized classification schemes into the new framework. We have developed the following classification system to give equal weight to all STI disciplines, while being compact and lightweight.

1   Introduction

Digital libraries (DLs) are quickly gaining acceptance and use in the scientific and research communities. NCSTRL+ is a canonical digital library based on the Networked Computer Science Technical Report Library (NCSTRL). The aim of NCSTRL+ is to provide users with a unified interface for multi-disciplinary/multi-genre searching [13]. One of the problems NCSTRL+ seeks to address is how to facilitate searching for information across diverse collections of specialized scientific and technical information. The two main stumbling blocks for users wishing to search for scientific and technical information are the lack of uniformity among individual DLs and the reliance of the DLs on discipline-specific jargon.

The answer is to create a new canonical classification system. It must be general enough allow more specialized subject categories to be mapped into it, since the purpose is to incorporate specialized classification systems, not replace them. The new system must also be balanced to represent all disciplines equally and avoid over-specialization.

Finally, the new system must also be lightweight, or it will be too cumbersome to work with efficiently.

2   Background

The NCSTRL+ prototype utilized the NASA Scientific and Technical Information (STI) categories [12] (Appendix B). They were chosen because the subjects were already familiar to most users [13], and the structure of the system was relatively close to what was desired (Table 1).

Main Subject Category	Subject Code
Aeronautics	01
Astronautics	12
Chemistry and Materials	23
Engineering	31
Geosciences	42
Life Sciences	51
Mathematical and Computer Sciences	59
Physics	70
Social Sciences	80
Space Sciences	88

Table 1. NASA Scientific and Technical Information Topics

The main problem with the NASA STI classification system is that it has a rather noticeable bias towards aeronautics, astronautics, and engineering topics to the detriment of other subjects. For example, there are 67 main and subcategories under engineering, but only 20 for mathematics and science combined. Social sciences and life sciences exhibit a similar lack of depth in their respective categories.

In order to ensure equal representation within each subject category, it would be necessary to redistribute the number of subcategories allocated to each main subject. It would also be desirable to separate mathematics and computer science into separate categories.

3   Existing Specialized Classification Systems

It would be easiest to replace the NASA STI system with a preexisting scientific or technical classification system. Unfortunately, most scientific and technical classification systems suffer from the same problem as the NASA STI system: the tendency to catalog subjects within the discipline in minute detail, ignoring ancillary subjects or giving them only a cursory categorization. There is a tendency to catalog what you know extremely well, while ignoring the categories that do not directly affect your profession. A summary of the specialized classification systems considered and why

they were ultimately rejected for NSTRL+ can be seen in Table 2. Figure 1 shows a relative placement of both specialized and general classification systems, and how none fall into the desired range. Examining these discipline-specific classification systems underscores the fact that although they do an excellent job of creating classification structures in their subject specialty, they lack the breadth of subject matter required for a general purpose classification system.

Name of Specialized Classification Scheme	Reason Rejected
Center for AeroSpace Information (CASI)	Too large, bias towards aerospace
Defense and Technical Information Center (DTIC)	Heavy emphasis on defense technology
Global Change Master Directory (GMCD)	Earth science specific categories
Physics E-print Archive	Categories are not well-balanced
American Mathematical Society (AMS)	Too many categories
Association for Computer Machinery (ACM)	Categories too discipline-specific
American Institute of Physics (AIP)	Too complex

Table 2. Specialized Classification Schemes Considered for NCSTRL+
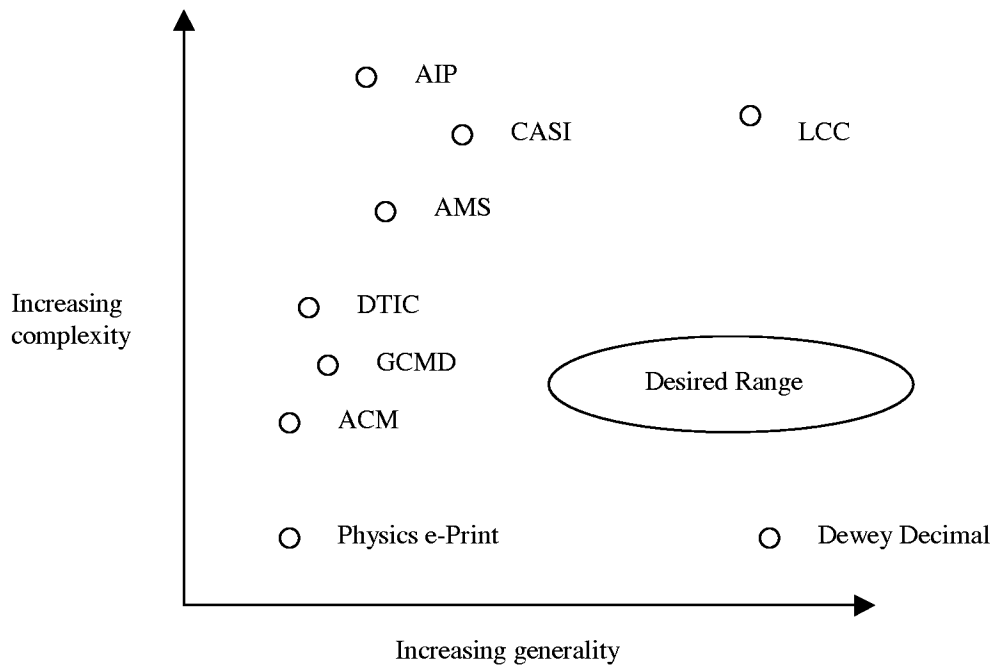

Figure 1. Complexity vs. generality in classification systems considered for NCSTRL+.

3

3.1 Center for AeroSpace Information (CASI)

The Center for AeroSpace Information (CASI) catalogs bibliographic citations for Scientific and Technical Aerospace Reports (STARs). CASI has subject categories and major subject terms [4]. The problem with the subject category is that there are 76 subject categories to choose from—far too many for the NCSRL+ project. In addition, CASI takes its major subject terms from the *NASA Thesaurus* [11], which again, reflects a NASA bias towards aerospace, aeronautics, and engineering. Another level of classification is added by allowing multiple terms to be entered into the secondary subject field, again, from the *NASA Thesaurus*. *Thesaurus* terms are arranged in a hierarchy that is too detailed and complex to easily incorporate into NCSTRL+.

3.2 Defense and Technical Information Center (DTIC)

The Defense Technical Information Center (DTIC) subject categories are also overly specialized, this time in subjects of special interest to the Department of Defense. DTIC has 25 main subject categories and 251 subcategories, with a military emphasis [6]. The main categories are numbered, with subcategories also numerically differentiated. It classifies to three levels deep. For example, the Astronomy and Astrophysics category only has three subheadings (Table 3), while the Guided Missile Technology subject category has nine distinct subcategories (Table 4).

Due to its heavy emphasis on defense technology and issues, the DTIC classification system was not considered an appropriate candidate to replace the NASA STI subject categories.

03—Astronomy and Astrophysics	
01	Astronomy
02	Astrophysics
03	Celestial Mechanics

Table 3. DTIC Astronomy and Astrophysics subcategories

16—Guided Missile Technology	
01	Guided Missile Launching and Basing Support
02	Guided Missile Trajectories, Accuracy and Ballistics
02/01	Guided Missile Dynamics, Configurations and Control Surfaces
03	Guided Missile Warheads and Fuzes
04	Guided Missiles
04/01	Air- and Space-Launched Guided Missiles
04/02	Surface-Launched Guided Missiles
04/03	Underwater-Launched Guided Missiles
05	Guided Missile Reentry Vehicles

Table 4. DTIC Guided Missile Technology Subcategories

3.3   Global Change Master Directory

The Global Change Master Directory (GCMD) allows users to search by subject for Earth Science data. It has 11 main categories, all relating to specific areas of expertise in the Earth Sciences [8]. The GCMD catalogs data three levels deep, which allows for very specific searches (for example: Cryosphere: Sea Ice: Ice Types). However, GCMD is too limiting to be a general classification system because it categorizes only Earth Science topics.

3.4   Physics E-print Archive

The Physics E-print Archive stores papers primarily written for the physics community, but also has papers on mathematics, nonlinear science, and computer science [15]. They have rudimentary subject classifications that seem to have arisen more out of necessity than intent. Most of the subcategories are under the main "Physics" category. High Energy Physics has 4 main categories (Experiment, Lattice, Phenomenology, and Theory). Mathematics is one major category, with individual disciplines in mathematics listed as subcategories. Simply put, the Physics E-print archive is not structured enough to be useful. The Physics E-print archive classification system does not provide a clear, balanced set of main and subcategories, nor does it list subjects unrelated to physics. This is understandable considering the targeted user group of this server and its evolutionary development.

3.5   American Mathematical Society (AMS)

The American Mathematical Society's Mathematics Subject Classification [2] is geared specifically to classify mathematical papers and information. The Mathematics Subject Classification system has 95 main categories, ranging from "Algebraic Geometry" to "Abstract Harmonic Analysis". While this categorization system does list other disciplines among its main categories, it lists them only if they are in some way related to mathematics. In addition to being a large classification system, it is also quite involved. The instructions deem it "extremely helpful for both readers and classifiers to familiarize themselves with the entire classification system" [2]. A classification system that requires extensive familiarity to implement and search is not suitable for the purposes of NCSTRL+.

3.6   Association for Computer Machinery (ACM)

The Association for Computer Machinery (ACM) Computing Classification System [3] uses the alphabetical letters A-K to denote main categories, separated by a period from numbers to denote subcategories (the exception to this rule is the "Miscellaneous" subcategory at the end of each main category. It is denoted by an "m"). Again, the emphasis on one particular discipline renders this classification system incomplete for NCSTRL+.

3.7 American Institute of Physics (AIP)

Probably the most complex classification system considered was that of the American Institute of Physics (AIP). It is called the Physics and Astronomy Classification System (PACS). Not only was PACS an enormous list (around 150 pages long), but it had a potentially confusing and complicated indexing scheme. According to the description,

> The PACS indexing categories are labeled by six-character
> Codes consisting of four numbers followed by a fifth character
> that can be either an uppercase letter or a plus or minus
> ...[the] sixth character is a lowercase character that serves
> as a check character [1].

PACS would be difficult to implement outside of a physics environment, due to the level of expertise required to catalog information in that scheme. It would also be extremely time consuming to map other classification codes onto PACS. Users unfamiliar with physics terminology would have difficulty finding the correct categories to search in. Last, but not least, it classifies only physics and astronomy categories.

4 Existing Generalized Classification Systems

General classification systems were also considered for use in NCSTRL+. General classification schemes are specifically designed to classify a wide range of subjects in detail. The two most common general classification systems are the Library of Congress Classification System (LCC) and the Dewey Decimal System. It was found, however that the major shortcoming of a generalized classification system was its *generality*—too many subject categories were classified to make it useful for NCSTRL+ (Table 5).

Name of Generalized Classification Scheme	Reason Rejected
Library of Congress Classification	Too complex, too detailed
Dewey Decimal System	Too generalized

Table 5. Generalized Classification Schemes considered for NCSTRL+.

4.1 Library of Congress Classification (LCC)

The LCC system is well known to anyone who has visited an academic library. It consists of 21 main categories [10], with subcategories defined first by letters and then numbers. The LCC is a very large classification system intended for large collections. It

provides enough breadth and depth to classify almost any collection. The fact that the LCC is such a large, complete classification system is precisely why it is unsuitable for use in NSCTRL+: it provides too much detail. Finding a copy of the LCC on the web is also a challenge, not to mention adapting it for use in a digital library environment, as the Pharos team discovered [14]. Aside from the implementation problems that LCC provides, properly mapping another DL's subject headers into the Library of Congress Classification system would take a fair amount of skill and time, negating the whole idea of adopting a simple, yet complete classification system.

4.2   Dewey Decimal System

The Dewey Decimal System is used primarily by public libraries. It is, like the LCC, a general purpose classification system. It is much easier to use than the LCC, limiting itself to 10 major subjects, each with 10 secondary subjects [16]. Specificity is obtained by adding numbers after the decimal point. The 10 major areas are shown in Table 6.

000	Generalities
100	Philosophy and Psychology
200	Religion
300	Social Sciences
400	Languages
500	Science
600	Technology
700	Arts and Music
800	Literature
900	Geography and History

Table 6. Dewey Decimal System Main Classifications

The main advantage to the Dewey Decimal System is that is well known by most users. It is also reasonably compact, and easy to work with. The reason it was not chosen as the classifying system for NCSRL+ is that the subject headings are too general for a specialized library. While Dewey is appropriate for public libraries, it is simply not adequate for STI applications.

Generalized library classifications schemes have the breadth of subject matter to be used by NCSTRL+, but lack the depth required by a scientific and technical library. They can also be bulky and difficult to implement in a digital library environment, and may require additional expertise to effectively catalog information and map other library classifications into them.

5   Creating a New Canonical Classification System

To create a new canonical classification system for NCSTRL+, a structure of 11 major subject headings (similar to the Dewey Decimal System [16]) with 11 subclasses per subject heading was chosen.

Once the number of main and subcategories was decided upon, the next phase was deciding what main/subcategories should be used. For the most part, the original NASA STI topics remained. The mathematics and computer science topic was divided into separate categories, and some subclasses were incorporated into newly created generalized subclasses or removed altogether. To see an example of the reshuffling and pruning, refer to the NASA STI Aeronautics subject classification (Table 7) and compare it to the NCSTRL+ Aeronautics subject classification (Table 8).

In order to create the subcategory headers, sources that had previously been dismissed as too specialized to be used as a stand alone classification system were consulted to decide what constituted a "general" subcategory. For Chemistry and Materials, ChemDex Plus [5] was used.

The Geosciences subject was renamed Earth Sciences, to make it consistent with NASA's Earth Science Enterprise. To rework the subclasses, the dictionary was used, as well as the author's experience working with Earth Science data.

PACS [1] was useful in helping to solidify the subclasses for Physics and Space Sciences. PACS was a good detailed framework to check NCSTRL+'s general subclasses against (PACS categories were able to map to NCSTRL+ categories).

The Computer Science category was developed with the help of the ACM Computing Classification System [3]. What was to be listed was already known, and the ACM classification system helped to identify which items were subcategories and which were sub-subcategories.

Members of the NASA Langley Research Center's Technical Library staff with experience in cataloging reviewed the initial NCSTRL+ classification system. They suggested additions and clarifications, especially to the Aeronautics, Astronautics, Engineering, and Social Sciences categories. After the requisite changes were made, they gave their approval for its use as a classification system. The finished Canonical Classification System for NCSTRL+ can be seen in Appendix A.

01	Aeronautics
02	Aerodynamics
02-01	Aerodynamics Characteristics
02-02	Aerodynamics of Bodies
03	Air Transportation and Safety
03-01	Commercial and General Aviation
03-02	Helicopters and Ground Effect Machines
03-03	STOL/VTOL Aircraft
03-04	Supersonic Transport
03-05	Aircraft Noise and Sonic Boom
03-06	Aircraft Safety and Safety Devices
03-07	Clear Air Turbulence
04	Aircraft Communications and Navigations
05	Aircraft Design, Testing and Performance
05-01	Hydraulic and Pneumatic Systems
05-02	Auxiliary Electrical Systems
06	Aircraft Instrumentation
07	Aircraft Propulsion and Power
07-01	Jet Propulsion
08	Aircraft Stability and Control
09	Research and Support Facilities (Air)
09-01	Wind Tunnels

Table 7. NASA STI Aeronautics main and subcategories

000	Aeronautics, General
000-010	History of Aeronautics
010	Aerodynamics
020	Commercial and General Aviation
030	Aviation Safety
040	Instrumentation
050	Communications
060	Propulsion and Power
070	Design
080	Aircraft Control
090	Research and Support Facilities

Table 8. NCSTRL+ Aeronautics main and subcategories

6 Related Projects

Perhaps the most closely related project to the Canonical Classification System for NCSTRL+ is Pharos [14], an offshoot of the Alexandria Digital Library Project at University of California, Santa Barbara [7]. Pharos mapped newsgroups to the Library of Congress Classification subjects. It allowed users to type in keywords, and it returned the newsgroups that were most likely to contain the information the user was looking for. The Pharos authors detailed the difficulty they had in making LCC suitable for automated classification [7]. In particular, some of the cataloging conventions were redundant or inconsistent. Pharos was begun in 1997; however, it does not seem to have progressed past the demonstration stage. It is viable, but at present, it does not appear to be under further development.

Larson [9] has also done research with LCC categories and automated classification. After conducting experiments with differing methods of automatic classification, he concluded "fully automatic classification may not be possible" using the LCC, but conceded that "semiautomatic classification...appears to be effective" [9]. This bolsters the contention that the LCC (in its present form) is simply too large and too complex to be used for automatic classification.

The Scorpion research project used the Dewey Decimal System as the basis for its automatic classification system, and reported favorable results [17]. Dewey's class integrity (how well subject classifications are differentiated) and hierarchical structure were cited as the reasons for its success. The authors concluded "results indicate that Dewey is a very good knowledge base for automatic subject assignment tools" [17].

7 Future Work

Although the initial work of creating the main and subcategories for NCSTRL+ has been completed, work on the project continues. The current catalog of NCSTRL+ will need to be mapped to the new classification codes. As NCSTRL+ grows and incorporates the holdings of other DLs, those collections will also need to be mapped to the appropriate categories.

It is possible that the current list may be incomplete or inadequate to handle certain specialized classification schemes. To test this new classification scheme, it will need to be implemented. Feedback from users should be encouraged, and the system will probably need to be adjusted to better serve the users.

Another area that can be explored is whether or not an additional level of subcategorizing is useful (or necessary). It may turn out that two levels of classification are not enough. Again, only a real world test will give the necessary data to decide the relative merit of this classification system.

8 Conclusions

Most scientific and technical classification schemes are too narrow in their focus to adequately fill the demands of NCSTRL+. They catalog within their areas of expertise in great detail, but only give cursory, if any, attention to fields outside of their specialties. In addition, the plethora of specialized, highly technical subclasses often found in

scientific and technical classification systems can be confusing for a user unfamiliar with that particular subject.

On the other hand, traditional library cataloging systems offer general classification subjects that are familiar to a majority of users. The drawback is that these systems were created to catalog large, diverse collections in minute detail. Not only is this level of classification not necessary, it is not wanted. Also, the general subject categories of traditional library cataloging systems are not completely relevant to NCSTRL+.

Since existing classification systems were either too complex or too generalized to be used to catalog NCSTRL+, a canonical classification system was created to fill the need for a lightweight, general-purpose classification system. The goal is to provide a balanced classification system that will be familiar enough to allow novice users to find the information they are looking for, even if they lack specific keywords or terms.

The new classification system presents a set number of main categories, each with a set number of subcategories. All disciplines relevant to the STI holdings are given equal weight in the listing. Specific topics can be placed appropriately under each subcategory. Existing, specialized categorization schemes can also be mapped at the subcategory level to allow users to search across diverse DLs. For viability, the NCSTRL+ classification system has been reviewed and approved by members of NASA Langley Research Center's technical library cataloging staff.

Acknowledgements

References

1. American Institute of Physics, "1998 Physics and Astronomy Classification System (PACS)."
   http://www.aip.org/pacs/pacs.html

2. American Mathematical Society, "Mathematics Subject Classification (1991)."
   http://www.ams.org/msc

3. Association for Computing Machinery, "ACM Computing Classification System (1998)."
   http://www.acm.org/class/1998/overview.html

4. Center for Aerospace Information Technical Report Server,
   http://www.sti.nasa.gov/RECONselect.html

5. ChemWeb ChemDex Plus, "Subject Headings."
   http://www.chemweb.com/databases/chemdex/chemdex.exe?action=browse

6. Defense Technical Information Center, "Subject Category Coverage."
   http://www.dtic.mil/dtic/subcatguide/#subcats/

7. R. Dolin, D. Agrawal, A. El Abbadi, & J.Pearlman, "Using Automated Classification for Summarizing and Selecting Heterogeneous Information Sources," *D-Lib Magazine, the Magazine for Digital Library Research,* January, 1998.
   http://www.dlib.org/dlib/january98/dolin/01dolin.html

8. Global Change Master Directory, "Science Keyword Interface."
   http://gcmd.gsfc.nasa.gov/param_search/top.html

9. R. R. Larson, "Experiments in Automatic Library of Congress Classification," *Journal of the American Society for Information Science,* 43(2), 1992, pp. 130-148.

10. Library of Congress Classification System
    http://geography.miningco.com/library/congress/bllc.html

11. NASA Thesaurus
    http://www.sti.nasa.gov/98Thesaurus/98thes.htm

12. NASA Scientific and Technical Information Program, "NASA STI Topics."
    ftp://ftp.sti.nasa.gov/pub/scan/SCAN-TOPICS

13. M. L. Nelson, K. Maly, S. N. T. Shen, & M. Zubair, "NSTRL+: 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.

14. Pharos
http://pharos.alexandria.ucsb.edu

15. Physics E-print Archive,
http://xxx.lanl.gov

16. Salt Lake Community College, "Dewey Decimal System."
http://www.slcc.edu/lr/library/info/dewey.htm

17. R. Thompson, K. Shafer, D. Vizine-Goetz, "Evaluating Dewey Concepts as a Knowledge Base for Automatic Subject Assignment," *Proceedings of ACM Digital Libraries '97*, Philadelphia, PA, July 23-26, 1997, pp. 37-46.

Appendix A

A Canonical STI Classification System for NCSTRL+

Aeronautics
000	Aeronautics, General & History
010	Aerodynamics
020	Commercial and General Aviation
030	Aviation Safety
040	Instrumentation
050	Communications
060	Propulsion and Power
070	Design
080	Aircraft Control
090	Research and Support Facilities

Astronautics
100	Astronautics, General & History
110	Astrodynamics
120	Space Vehicles and Space Stations
130	Safety
140	Instrumentation
150	Communications
160	Propulsion and Power
170	Design
180	Navigation and Guidance Systems
190	Research and Support Facilities

Chemistry and Materials
200	Chemistry and Materials, General
210	Electrochemistry
220	Chemical Processes
230	Chemical Analysis
240	Organic Chemistry
250	Inorganic Chemistry
260	Physical Chemistry
270	Materials
270-010	Metallic
270-020	Non-metallic
270-030	Composite
280	Propellants and Fuels
290	Processing

Engineering and Applied Technology
300	Engineering, General
310	Electrical Engineering
320	Communications
330	Electronics
340	Lasers and Masers
350	Fluid Mechanics and Heat Transfer
360	Mechanical Engineering
370	Instrumentation and Measurement
380	Structural Mechanics

| 390 | Quality Assurance |
| 395 | Photography |

Earth Sciences

400	Earth Sciences, General
410	Geophysics
410-010	Geology
410-020	Seismology
410-030	Geomagnetism
420	Oceanography
430	Geography
430-010	Cartograpy
440	Energy Production
440-010	Energy Resources
450	Environmental Issues
450-010	Pollution
450-020	Global Warming
460	Atmospheric Science
460-010	Meteorology
460-020	Climatology
460-030	Climatological Phenomena
460-030	Upper Atmosphere
460-040	Satellites
470	Hydrology

Life Sciences

500	Life Sciences, General
510	Biology
520	Biochemistry
530	Medicine
530-010	Aerospace Medicine
530-020	Clinical Medicine
530-030	Physiological Factors
540	Life Sciences Technology
540-010	Life Support Systems
550	Space Biology
550-010	Extraterrestrial Life
560	Biological Physics
570	Pharmacology
580	Psychology
580-010	Cognition
590	Botany

Mathematics

600	Mathematics, General
610	Applied Mathematics
620	Theoretical Mathematics
630	Statistics
640	Numerical Analysis
650	Geometry
660	Topology
670	Probability
680	Logic
690	Mathematical Physics

Computer Science
700	Computer Science, General
710	Computer Networks
710-010	Internet
720	Hardware
730	Software
730-010	Software Engineering
730-020	Programming Languages
740	Information Systems
740-010	Information Management
740-020	Database
740-030	Information Retrieval
750	Data
750-010	Data Storage
750-020	Data Encryption
750-030	Data Structures
760	Artifical Intelligence
770	Robotics
780	Artificial Intelligence
790	Human-Computer Interaction

Physics
800	Physics, General
805	Elementary Particles and Fields
805-010	Relativity
805-020	Unified Field Theories and Models
810	Statistical Physics
815	High Energy Physics
820	Thermodynamics
825	Quantum Physics
830	Solid-State Pysics
840	Gases, Plasmas, and Electrical Discharges
850	Optics
860	Nuclear Physics
870	Atomic and Molecular Physics
880	Condensed Matter
890	Acoustics

Social Sciences
900	Social Sciences, General
910	Law
920	Political Science
925	Government and Military Science
930	Economics
940	Business
940-010	Administration and Management
950	Communications and Media
960	Transportation
970	Technology Transfer
970	Sociology
970-020	Social Psychology
980	Education
985	Library and Information Science
990	History
995	Biography

Space Sciences
1000   Space Sciences
1010   Astronomy
1020   Astrophysics
1030   Solar System
1030-010   Planetary Exploration
1040   The Moon
1050   The Sun
1050-010   Solar Astronomy
1050-020   Solar Physics
1060   Stars
1070   The Universe
1070-010   Stellar Systems
1070-020   Interstellar Medium
1070-030   Galactic Objects and Systems
1070-040   Extragalactic Objects and Systems
1070-050   Space Radiation

Appendix B

NASA STI SCAN Topics

AERONAUTICS
    01 AERONAUTICS (GENERAL)
    02 AERODYNAMICS
    02-01 AERODYNAMICS CHARACTERISTICS
    02-02 AERODYNAMICS OF BODIES
    02-03 AIRFOIL AND WING AERODYNAMICS
    03 AIR TRANSPORTATION AND SAFETY
    03-01 COMMERCIAL AND GENERAL AVIATION
    03-02 HELICOPTERS AND GROUND EFFECT MACHINES
    03-03 STOL/VTOL AIRCRAFT
    03-04 SUPERSONIC TRANSPORT
    03-05 AIRCRAFT NOISE AND SONIC BOOM
    03-06 AIRCRAFT SAFETY AND SAFETY DEVICES
    03-07 CLEAR AIR TURBULENCE
    04 AIRCRAFT COMMUNICATIONS AND NAVIGATIONS
    05 AIRCRAFT DESIGN, TESTING AND PERFORMANCE
    05-01 HYDRAULIC AND PNEUMATIC SYSTEMS
    05-02 AUXILIARY ELECTRICAL SYSTEMS
    06 AIRCRAFT INSTRUMENTATION
    07 AIRCRAFT PROPULSION AND POWER
    07-01 JET PROPULSION
    08 AIRCRAFT STABILITY AND CONTROL
    09 RESEARCH AND SUPPORT FACILITIES (AIR)
    09-01 WIND TUNNELS

ASTRONAUTICS
    12 ASTRONAUTICS (GENERAL)
    13 ASTRODYNAMICS
    13-01 CELESTIAL MECHANICS AND ORBITAL CALCULATIONS
    14 GROUND SUPPORT SYSTEMS AND FACILITIES (SPACE)
    14-01 SPACECRAFT GROUND SUPPORT
    14-02 TEST FACILITIES
    14-03 SIMULATORS AND SIMULATION
    14-04 STERILIZATION
    15 LAUNCH VEHICLES AND SPACE VEHICLES
    15-01 LAUNCH VEHICLES
    15-02 SOUNDING ROCKETS
    15-03 SPACE PROBES
    15-04 SCIENTIFIC SATELLITES
    15-05 REENTRY VEHICLES
    15-06 U.S.S.R SPACECRAFT
    16 SPACE TRANSPORTATION
    16-01 SPACE TRANSPORTATION AND MANNED SPACECRAFT
    17 SPACE COMMUNICATIONS, SPACECRAFT COMMUNICATIONS, COMMAND AND TRACKING
    17-01 SPACE COMMUNICATIONS
    17-02 NAVIGATION SYSTEMS
    17-03 GUIDANCE SYSTEMS
    17-04 TRACKING
    18 SPACECRAFT DESIGN, TESTING AND PERFORMANCE

33-01 RADAR EQUIPMENT
33-02 SEMICONDUCTORS AND TRANSISTORS
33-03 ANTENNAS
33-04 ELECTRONIC COMPONENTS
33-05 CIRCUITRY
33-06 ELECTRICAL EQUIPMENT
33-07 AMPLIFIERS
33-08 FEEDBACK AND CONTROL THEORY
33-09 ELECTROMAGNETIC RADIATION
33-10 MICROELECTRONICS
33-11 MICROWAVE AND SUBMILLIMETER WAVE TECHNOLOGY
33-12 MAGNETISM
34 FLUID MECHANICS AND HEAT TRANSFER
34-01 BOUNDARY LAYER TECHNOLOGY
34-02 GAS DYNAMICS
34-03 FLUIDICS
34-04 FLUID FLOW
34-05 COMBUSTION PHYSICS
34-06 HEAT TRANSFER, BASIC
34-07 REENTRY HEAT TRANSFER
34-08 THERMAL PROTECTION
34-09 ABLATION
34-10 CRYOGENICS
35 INSTRUMENTATION AND PHOTOGRAPHY
35-01 PHOTOGRAPHY
35-02 INFRARED TECHNOLOGY
35-03 INSTRUMENT STANDARDS AND CALIBRATION TECHNIQUES
35-04 TEMPERATURE MEASUREMENT
35-05 PRESSURE MEASUREMENT
35-06 DISPLAY SYSTEMS
35-07 DATA RECORDING
35-08 GAS FLOW MEASUREMENT
36 LASERS AND MASERS
36-01 LASERS AND MASERS
36-02 LASER APPLICATIONS
37 MECHANICAL ENGINEERING
37-01 BEARINGS AND GEARS
37-02 LUBRICATION AND LUBRICANTS
37-03 MACHINING
37-04 FRICTION AND WEAR
37-05 SEALS
37-06 WELDING
37-07 METAL FORMING
37-08 PUMPS
37-09 VACUUM TECHNOLOGY
37-10 NONDESTRUCTIVE TESTING
37-11 TURBOMACHINERY
38 QUALITY ASSURANCE AND RELIABILITY
38-01 QUALITY CONTROL AND RELIABILITY
39 STRUCTURAL MECHANICS
39-01 SHELLS
39-02 STRESSES AND LOADS
39-03 STRUCTURE VIBRATION AND DAMPING
39-04 IMPACT PHENOMENA
39-05 STRUCTURAL FATIGUE
39-06 SANDWICH CONSTRUCTION

61 COMPUTER PROGRAMMING AND SOFTWARE
61-01 COMPUTER SOFTWARE
61-02 CAD/CAM
62 COMPUTER SYSTEMS
63 CYBERNETICS
63-01 CYBERNETICS AND BIONICS
63-02 ARTIFICIAL INTELLIGENCE
64 NUMERICAL ANALYSIS
64-01 NUMERICAL ANALYSIS
65 STATISTICS AND PROBABILITY
65-01 PROBABILITY AND STATISTICS
66 SYSTEMS ANALYSIS
67 THEORETICAL MATHEMATICS

PHYSICS
70 PHYSICS (GENERAL)
71 ACOUSTICS
71-01 ACOUSTICS
71-02 ULTRASONICS
72 ATOMIC AND MOLECULAR PHYSICS
72-01 ATOMIC PHYSICS
72-02 MOLECULAR PHYSICS
73 NUCLEAR AND HIGH-ENERGY PHYSICS
73-01 NUCLEAR PHYSICS
73-02 RADIOACTIVITY
74 OPTICS
74-01 OPTICS
74-02 LIGHT
75 PLASMA PHYSICS
75-01 PLASMA APPLICATIONS
75-02 PLASMA DYNAMICS
75-03 MAGNETOHYDRODYNAMICS
76 SOLID-STATE PHYSICS
76-01 SOLID STATE DEVICES
76-02 SUPERCONDUCTIVITY
76-03 DIELECTRICS
76-04 EPITAXIAL DEPOSITION
77 THERMODYNAMICS AND STATISTICAL PHYSICS

SOCIAL SCIENCES
80 SOCIAL SCIENCES (GENERAL)
81 ADMINISTRATION AND MANAGEMENT
81-01 AEROSPACE MANAGEMENT
82 DOCUMENTATION AND INFORMATION SCIENCE
82-01 INFORMATION TECHNOLOGY
83 ECONOMICS AND COST ANALYSIS
84 LAW, POLITICAL SCIENCE AND SPACE POLICY
84-01 WORLD SPACE PROGRAMS AND AEROSPACE LAW
84-02 SPACE COMMERCIALIZATION
85 URBAN TECHNOLOGY AND TRANSPORTATION
85-01 URBAN TECHNOLOGY AND TRANSPORTATION

SPACE SCIENCES
88 SPACE SCIENCES (GENERAL)

89 ASTRONOMY
89-01 SOLAR ASTRONOMY
89-02 STELLAR ASTRONOMY AND COSMOLOGY
89-03 METEORS AND METEORITES
90 ASTROPHYSICS
90-01 GRAVITATION
90-02 ASTROPHYSICAL PLASMAS
91 LUNAR AND PLANETARY EXPLORATION
91-01 THE MOON
91-02 PLANETARY SCIENCES AND EXPLORATION
92 SOLAR PHYSICS
93 SPACE RADIATION
93-01 COSMIC RADIATION
93-02 SOLAR RADIATION AND ACTIVITY
93-03 RADIATION BELTS

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 1998	3. REPORT TYPE AND DATES COVERED Technical Memorandum

4. TITLE AND SUBTITLE Creating a Canonical Scientific and Technical Information Classification System for NCSTRL+	5. FUNDING NUMBERS
6. AUTHOR(S) Melissa E. Tiffany 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-17797

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-1998-208955

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)

The purpose of this paper is to describe the new subject classification system for the NCSTRL+ project. NCSTRL+ is a canonical digital library (DL) based on the Networked Computer Science Technical Report Library (NCSTRL). The current NCSTRL+ classification system uses the NASA Scientific and Technical (STI) subject classifications, which has a bias towards the aerospace, aeronautics, and engineering disciplines. Examination of other scientific and technical information classification systems showed similar discipline-centric weaknesses. Traditional, library-oriented classification systems represented all disciplines, but were too generalized to serve the needs of a scientific and technically oriented digital library. Lack of a suitable existing classification system led to the creation of a lightweight, balanced, general classification system that allows the mapping of more specialized classification schemes into the new framework. We have developed the following classification system to give equal weight to all STI disciplines, while being compact and lightweight

14. SUBJECT TERMS Digital Libraries, Subject Categories, Classification	15. NUMBER OF PAGES 28
	16. PRICE CODE AO3

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