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
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 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
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

National Aeronautics and Space Administration  
Langley Research Center  
Hampton, Virginia 23681-2199

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).

<table>
<thead>
<tr>
<th>Main Subject Category</th>
<th>Subject Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeronautics</td>
<td>01</td>
</tr>
<tr>
<td>Astronautics</td>
<td>12</td>
</tr>
<tr>
<td>Chemistry and Materials</td>
<td>23</td>
</tr>
<tr>
<td>Engineering</td>
<td>31</td>
</tr>
<tr>
<td>Geosciences</td>
<td>42</td>
</tr>
<tr>
<td>Life Sciences</td>
<td>51</td>
</tr>
<tr>
<td>Mathematical and Computer Sciences</td>
<td>59</td>
</tr>
<tr>
<td>Physics</td>
<td>70</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>80</td>
</tr>
<tr>
<td>Space Sciences</td>
<td>88</td>
</tr>
</tbody>
</table>

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.

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

Table 2. Specialized Classification Schemes Considered for NCSTRL+

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

<table>
<thead>
<tr>
<th>03—Astronomy and Astrophysics</th>
</tr>
</thead>
<tbody>
<tr>
<td>01  Astronomy</td>
</tr>
<tr>
<td>02  Astrophysics</td>
</tr>
<tr>
<td>03  Celestial Mechanics</td>
</tr>
</tbody>
</table>

Table 3. DTIC Astronomy and Astrophysics subcategories

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

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).

<table>
<thead>
<tr>
<th>Name of Generalized Classification Scheme</th>
<th>Reason Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of Congress Classification</td>
<td>Too complex, too detailed</td>
</tr>
<tr>
<td>Dewey Decimal System</td>
<td>Too generalized</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>000</th>
<th>Generalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Philosophy and Psychology</td>
</tr>
<tr>
<td>200</td>
<td>Religion</td>
</tr>
<tr>
<td>300</td>
<td>Social Sciences</td>
</tr>
<tr>
<td>400</td>
<td>Languages</td>
</tr>
<tr>
<td>500</td>
<td>Science</td>
</tr>
<tr>
<td>600</td>
<td>Technology</td>
</tr>
<tr>
<td>700</td>
<td>Arts and Music</td>
</tr>
<tr>
<td>800</td>
<td>Literature</td>
</tr>
<tr>
<td>900</td>
<td>Geography and History</td>
</tr>
</tbody>
</table>

Table 6. Dewey Decimal System Main Classifications

The main advantage to the Dewey Decimal System is that it 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.
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Aeronautics</td>
</tr>
<tr>
<td>02</td>
<td>Aerodynamics</td>
</tr>
<tr>
<td>02-01</td>
<td>Aerodynamics Characteristics</td>
</tr>
<tr>
<td>02-02</td>
<td>Aerodynamics of Bodies</td>
</tr>
<tr>
<td>03</td>
<td>Air Transportation and Safety</td>
</tr>
<tr>
<td>03-01</td>
<td>Commercial and General Aviation</td>
</tr>
<tr>
<td>03-02</td>
<td>Helicopters and Ground Effect Machines</td>
</tr>
<tr>
<td>03-03</td>
<td>STOL/VTOL Aircraft</td>
</tr>
<tr>
<td>03-04</td>
<td>Supersonic Transport</td>
</tr>
<tr>
<td>03-05</td>
<td>Aircraft Noise and Sonic Boom</td>
</tr>
<tr>
<td>03-06</td>
<td>Aircraft Safety and Safety Devices</td>
</tr>
<tr>
<td>03-07</td>
<td>Clear Air Turbulence</td>
</tr>
<tr>
<td>04</td>
<td>Aircraft Communications and Navigations</td>
</tr>
<tr>
<td>05</td>
<td>Aircraft Design, Testing and Performance</td>
</tr>
<tr>
<td>05-01</td>
<td>Hydraulic and Pneumatic Systems</td>
</tr>
<tr>
<td>05-02</td>
<td>Auxiliary Electrical Systems</td>
</tr>
<tr>
<td>06</td>
<td>Aircraft Instrumentation</td>
</tr>
<tr>
<td>07</td>
<td>Aircraft Propulsion and Power</td>
</tr>
<tr>
<td>07-01</td>
<td>Jet Propulsion</td>
</tr>
<tr>
<td>08</td>
<td>Aircraft Stability and Control</td>
</tr>
<tr>
<td>09</td>
<td>Research and Support Facilities (Air)</td>
</tr>
<tr>
<td>09-01</td>
<td>Wind Tunnels</td>
</tr>
</tbody>
</table>

**Table 7. NASA STI Aeronautics main and subcategories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>Aeronautics, General</td>
</tr>
<tr>
<td>000-010</td>
<td>History of Aeronautics</td>
</tr>
<tr>
<td>010</td>
<td>Aerodynamics</td>
</tr>
<tr>
<td>020</td>
<td>Commercial and General Aviation</td>
</tr>
<tr>
<td>030</td>
<td>Aviation Safety</td>
</tr>
<tr>
<td>040</td>
<td>Instrumentation</td>
</tr>
<tr>
<td>050</td>
<td>Communications</td>
</tr>
<tr>
<td>060</td>
<td>Propulsion and Power</td>
</tr>
<tr>
<td>070</td>
<td>Design</td>
</tr>
<tr>
<td>080</td>
<td>Aircraft Control</td>
</tr>
<tr>
<td>090</td>
<td>Research and Support Facilities</td>
</tr>
</tbody>
</table>

**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

We would like to thank Nancy Kaplan, Garland Gouger, and John Ferrainolo of the NASA Langley Research Center Technical Library for their assistance in reviewing and contributing to this classification system.
References

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

   http://www.ams.org/msc

   http://www.acm.org/class/1998/overview.html

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


   http://www.dtic.mil/dtic/subcatguide/#subcats/

   http://www.dlib.org/dlib/january98/dolin/01dolin.html

   http://gcmd.gsfc.nasa.gov/param_search/top.html


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

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


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

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

   http://www.slcc.edu/ir/library/info/dewey.htm

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 Cartography
440 Energy Production
440-010 Energy Resources
450 Environmental Issues
450-010 Pollution
450-020 Global Warning
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  Artificial 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 Physics
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
<table>
<thead>
<tr>
<th>Code</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>Space Sciences</td>
</tr>
<tr>
<td>1010</td>
<td>Astronomy</td>
</tr>
<tr>
<td>1020</td>
<td>Astrophysics</td>
</tr>
<tr>
<td>1030</td>
<td>Solar System</td>
</tr>
<tr>
<td>1030-010</td>
<td>Planetary Exploration</td>
</tr>
<tr>
<td>1040</td>
<td>The Moon</td>
</tr>
<tr>
<td>1050</td>
<td>The Sun</td>
</tr>
<tr>
<td>1050-010</td>
<td>Solar Astronomy</td>
</tr>
<tr>
<td>1050-020</td>
<td>Solar Physics</td>
</tr>
<tr>
<td>1060</td>
<td>Stars</td>
</tr>
<tr>
<td>1070</td>
<td>The Universe</td>
</tr>
<tr>
<td>1070-010</td>
<td>Stellar Systems</td>
</tr>
<tr>
<td>1070-020</td>
<td>Interstellar Medium</td>
</tr>
<tr>
<td>1070-030</td>
<td>Galactic Objects and Systems</td>
</tr>
<tr>
<td>1070-040</td>
<td>Extragalactic Objects and Systems</td>
</tr>
<tr>
<td>1070-050</td>
<td>Space Radiation</td>
</tr>
</tbody>
</table>
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
18-01 SPACECRAFT ATTITUDE CONTROL AND STABILIZATION
18-02 RENDEZVOUS AND DOCKING
18-03 SPACE STATIONS
19 SPACECRAFT INSTRUMENTATION
19-01 SPACECRAFT AND AIRCRAFT INSTRUMENTATION
19-02 SENSORS AND TRANSDUCERS
20 SPACECRAFT PROPULSION AND POWER
20-01 ROCKET ENGINES, NOZZLES AND THRUST CHAMBERS
20-02 AUXILIARY PROPULSION
20-03 ELECTRIC PROPULSION

CHEMISTRY AND MATERIALS
23 CHEMISTRY AND MATERIALS (GENERAL)
23-01 CHEMICAL ANALYSIS
23-02 CHEMICAL PROCESSES AND ENGINEERING
23-03 LUMINESCENCE
23-04 PHOTOCHEMISTRY
24 COMPOSITE MATERIALS
24-01 REINFORCED MATERIALS AND FIBERS
24-02 COMPOSITE MATERIALS
25 INORGANIC AND PHYSICAL CHEMISTRY
25-01 CORROSION
25-02 METAL CRYSTALS
25-03 COatings
25-04 ELECTROCHEMISTRY
26 METALLIC MATERIALS
26-01 ALUMINUM
26-02 BERYLLIUM
26-03 LIQUID METALS
26-04 STEEL
26-05 TITANIUM
26-06 REFRACTORY METALS
26-07 METALLURGY
27 NONMETALLIC MATERIALS
27-01 PLASTICS
27-02 ADHESIVES
27-03 CERAMICS
27-04 ELASTOMERS
27-05 GRAPHITE
27-06 POLYMERS
28 PROPELLANTS AND FUELS
28-01 LIQUID PROPELLANTS
28-02 SOLID PROPELLANTS
29 MATERIALS PROCESSING

ENGINEERING
31 ENGINEERING (GENERAL)
32 COMMUNICATIONS AND RADAR
32-01 COMMUNICATION SATELLITES
32-02 COMMUNICATION EQUIPMENT
32-03 COMMUNICATION SYSTEMS
32-04 TELEMETRY
32-05 RADIO NOISE
32-06 COMMUNICATION THEORY
33 ELECTRONICS AND ELECTRICAL ENGINEERING
39-07 STRESS ANALYSIS
39-08 STRUCTURAL TESTS AND RELIABILITY

GEOSCIENCES
42 GEOSCIENCES (GENERAL)
43 EARTH RESOURCES AND REMOTE SENSING
43-01 EARTH RESOURCES
43-02 GEODESY AND CARTOGRAPHY
44 ENERGY PRODUCTION AND CONVERSION
44-01 ENERGY RESOURCES
44-02 FUEL CELLS AND CHEMICAL BATTERIES
44-03 SOLAR SPACE POWER
44-04 NUCLEAR AUXILIARY POWER
45 ENVIRONMENT POLLUTION
45-01 ENVIRONMENT POLLUTION CONTROL
46 GEOPHYSICS
46-01 UPPER EARTH ATMOSPHERE
46-02 GEOLOGY AND SEISMOLOGY
46-03 GEOMAGNETISM
47 METEOROLOGY AND CLIMATOLOGY
47-01 METEOROLOGICAL SATELLITES
47-02 WEATHER FORECASTING
47-03 MICROMETEOROLOGY
47-04 CLOUD RESEARCH
47-05 METEOROLOGICAL INSTRUMENTS
48 OCEANOGRAPHY
48-01 WATER RESOURCES AND OCEANOGRAPHY

LIFE SCIENCES
51 LIFE SCIENCES (GENERAL)
51-01 BIOLOGY (GENERAL)
51-02 BIOCHEMISTRY
52 AEROSPACE MEDICINE
52-01 AEROSPACE MEDICINE
52-02 CLINICAL MEDICINE
52-03 PHYSIOLOGICAL FACTORS
52-04 BIOLOGICAL RADIATION EFFECTS
53 BEHAVIORAL SCIENCES
53-01 PSYCHOLOGICAL FACTORS
54 MAN/SYSTEMS TECHNOLOGY AND LIFE SUPPORT
54-01 LIFE SUPPORT SYSTEMS
54-02 CREW SAFETY AND PROTECTIVE CLOTHING
54-03 HUMAN ENGINEERING
54-04 MAN-MACHINE SYSTEMS
54-05 BIOINSTRUMENTATION
54-06 ROBOTICS
55 SPACE BIOLOGY
55-01 EXTRATERRESTRIAL LIFE

MATHEMATICAL AND COMPUTER SCIENCES
59 MATHEMATICAL AND COMPUTER SCIENCES (GENERAL)
59-01 APPLIED MATHEMATICS
59-02 DATA PROCESSING
60 COMPUTER OPERATIONS AND HARDWARE
60-01 DIGITAL AND ANALOG COMPUTERS
60-02 AIRBORNE OR SPACEBORNE COMPUTERS
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
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.