Disciplinary, Asynthetic, Domain-Dependent: NARCIS a National Research Classification in Isolation

Richard P. Smiraglia
Data Archiving and Networked Services, Royal Netherlands Academy of the Arts and Sciences
The Hague, The Netherlands
School of Information Studies, University of Wisconsin-Milwaukee
smiragli@uwm.edu

ABSTRACT
NARCIS, the National Academic Research and Collaborations Information System, is the national research portal of the Netherlands. NARCIS is governed by a knowledge organization system—a classification—by the same name. For a variety of reasons—a disciplinary base, a lack of synthesis, and domain-dependency—the NARCIS classification is highly compartmentalized and therefore inhospitable for interoperability. In addition, the classification has been revised repeatedly leading to the problems of scheme-versioning and subject ontogeny.

Keywords
NARCIS, classification, synthesis, disciplinary, domain-dependent.

NARCIS CLASSIFICATION

NARCIS is the national research portal of Netherlands' wide-ranging data and research archiving structure for the twenty-first century. The acronym stands for National Academic Research and Collaborations Information System (NARCIS). According to various websites of the Dutch research community, NARCIS is a repository that combines open access publications and datasets from Dutch scholars with texts of peer reviewed publications and other research data. NARCIS is governed by a knowledge organization system—a classification—by the same name. For a variety of reasons addressed in this concept paper—a disciplinary base, a lack of synthesis, and domain-dependency—the NARCIS classification is "siloed" or highly compartmentalized and therefore inhospitable for interoperability. In addition, the classification has been completely revised at least once, leading to the problems of scheme-versioning if not also subject ontogeny.

Advances in Classification Research, 2017, October 27, 2017, Washington, DC, USA.

Domain dependence

The NARCIS Classification is domain-dependent, meaning it is entirely and only designed for the contents of the NARCIS data portal, which is the current output of Dutch scholarship. The NARCIS Classification is designed to provide access to scientific information from Dutch scholars who enter their research into its repository. NARCIS Classification symbols are assigned to represent the knowledge-bases of contributing scholars, rather than to represent the content of the publications in the NARCIS repository. The NARCIS project (DANS a) began in 2004 as a cooperative project of Dutch research institutes resulting in the opening of its original portal in 2007. Since 2011 it has been housed at DANS (Data Archiving and Networked Services, A Division of the Royal Netherlands Academy of the Arts and Sciences). The current classification dates from 2015, although it is not clear from public documents who exactly is responsible for its intellectual management. The classification is made up of two classes (DANS b)—D for the sciences broadly, and E for interdisciplinary areas. Altogether there are 223 classes, divisions and subdivisions.

The classification is overwhelmingly for the sciences: The general outline shows the two classes and the seven divisions of class D, each with its respective number of divisions and subdivisions:

D10000 Science and technology	89
D20000 Life sciences, medicine and health care	63
D30000 Humanities	28
D40000 Law and public administration	11
D50000 Behavioural and educational sciences	4
D60000 Social sciences	9
D70000 Economics and business administration	1
E10000 Interdisciplinary sciences	8

Table 1. NARCIS Classification Class "D" Sciences

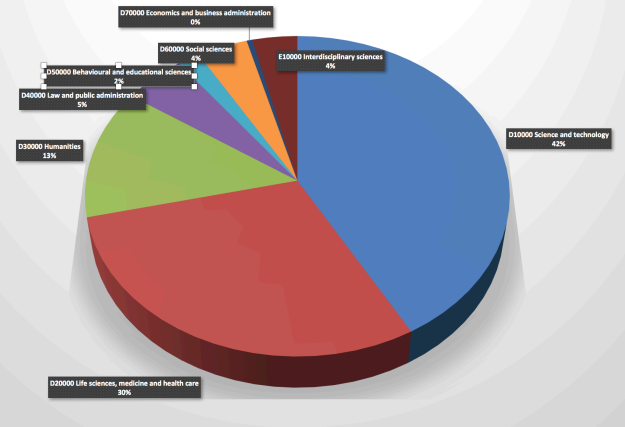
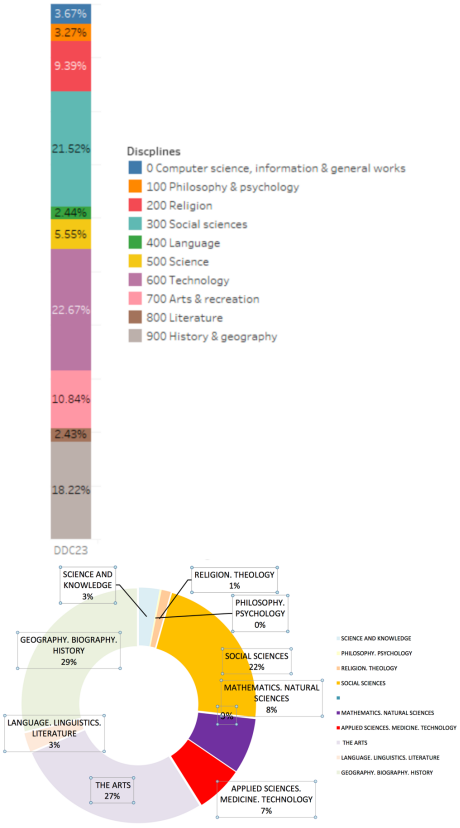


Figure 1. Distribution of Disciplines in NARCIS Classification.

The "sciences" occupy 76%. We can compare this visualization with those from the *Dewey Decimal Classification* 23 (Choi 2017, 8), Wikipedia categories 2008, and the Universal Decimal Classification 2008 (data from the Knowledge Space Lab):



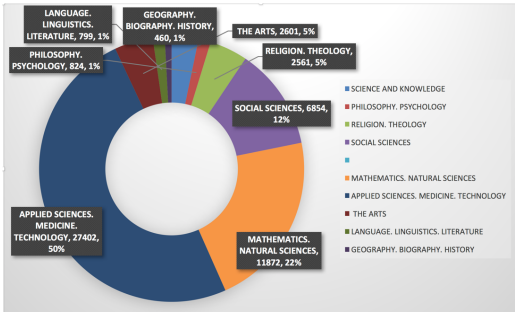


Figure 2. Distribution of Disciplines in DDC, Wikipedia and UDC.

In *DDC23* the sciences occupy 28.1% (600 22.6%, 500 5.5%), in Wikipedia 18%, in UDC 72%. Although comparisons obviously are not exact, it is interesting to see how the disciplinary focus of NARCIS aligns with the literary warrant-based UDC, as well as how different it is proportionately from the Wikipedia categories.

Anonmalies include the fact that Economics occupies its own division with business administration at the end of the D class. Humanities occupies one division in total. The breakdown of the Humanities is as follows:

D30100 Digital humanities	1
D31000 Paleography, bibliology, bibliography, library science	1
D32000 Philosophy	5
D33000 Theology and religious studies	1
D34000 History	3
D35000 Arts and culture	5
D36000 Language and literature studies	6
D37000 Archaeology	1
D38000 Area Studies	1

Table 2. Humanities in NARCIS Classification

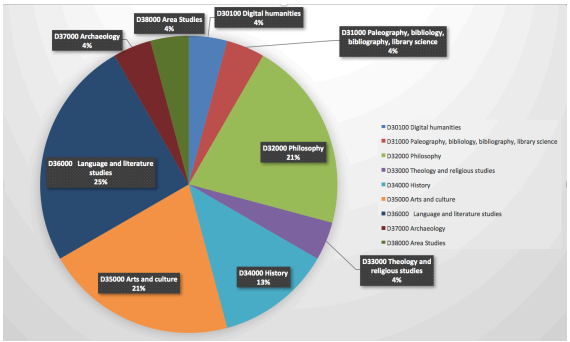


Figure 3. Placement of Humanities in NARCIS Classification.

A further anomaly occurs with information science, which is not present in the NARCIS classification. "Library science" occurs as a division of bibliography, which is a

methodology of history under humanities. “Computer science,” however, occupies a division with 8 subdivisions including information systems, artificial intelligence.

The question for classification research becomes the degree to which the classification is influenced politically by its domain rather than empirically representing scholarship. Smiraglia (2014) suggests the political disciplinarity is a result of social epistemological forces. But if the goal of the repository is to properly represent scholarship a more empirical basis for the structure of the classification would be appropriate. The absence of information science and the misnaming and misplacing of librarianship suggests political cultural pervasiveness (Smiraglia 2015) as a form of unseen objective disciplinary violence (Tennis 2013).

ASYNTHESIS

The NARCIS classification has no evidence of synthesis. Divisions and subdivisions may be independently represented but not combined in knowledge representation or either pre- or post-coordinate searching. The NARCIS classification website (DANS b) describes each individually named “category” as a “facet.” But there is no evidence of facet analytical theory in the construction of implementation of the NARCIS classification.

INTERDISCIPLINARITY

Interdisciplinarity is poorly served in the NARCIS classification. A separate class E is set aside for interdisciplinary sciences. It includes:

E11000 Biotechnology
E12000 Technology in medicine and health care
E13000 Development studies
E14000 Migration, ethnic relations and multiculturalism
E15000 Environmental studies
E16000 Nanotechnology
E17000 Greenhouse gas mitigation
E18000 Biobased economy

Table 3. Interdisciplinarity in NARCIS Classification

The isolation of these areas of transdisciplinarity from the rest of the knowledge base is another example of how the NARCIS classification silos by discipline—a distinctly anti-interdisciplinary approach. According to Szostak, Gnoli and López-Huertas (2016), interdisciplinarity requires the ability to search together by phenomenon in order to avoid the obstacles imposed by disciplinary boundaries. No specific phenomena are identified in the NARCIS classification, nor are any scope notes available to assist in choice of classification for knowledge representation.

SCHEME CHANGE

NARCIS Classification was completely revised in 2015 when the database migrated from an earlier repository to its current home at DANS. According to the website (DANS b), the classification was changed radically in 2015:

> The previous NARCIS classification code consisted of two main categories. The "A" code gave an overview of areas of interest, and the "D" code classified scientific disciplines. In addition, the classification included a "C" code for interdisciplinary research areas. This classification consisted of 94 "A" codes, eight "C" codes, and 182 "D" codes.

The new classification is includes 223 codes (and terms): 214 "D" codes (disciplines) and 8 "E" codes (interdisciplinary sciences).

The changes [we]re:

> -Removal of the "A" codes (areas of interest): All "A" codes have expired and where possible, have been modified or merged with a "D" code.
>
> -Change in the "D" codes (disciplines): due to the addition of new fields of science, the number of "D" codes has been expanded. There are 41 new disciplines.
>
> -“C" codes have become "E" codes (interdisciplinary studies), and are included in a category "interdisciplinary sciences".

There is no indication of whether the data in the repository were amended at this time to reflect the scheme change. The repository consists of links to institutional records. That is, authors “deposit” texts in NARCIS by first making them available in their university or institute-based online repositories and then linking to the NARCIS portal. This makes any such shift in knowledge representation unlikely. According to the well-known work by Tennis (2006; 2007) this raises two situations for records classified using NARCIS. First the problem of subject ontogeny (Tennis 2002; 2012); there likely are many classified terms (areas of interest, for example, or older “fields of science,” or former C codes that now have become E codes) for which representation has shifted from the earlier version of the classification. Second, there is no way to connect records represented by either version of the classification together to support collocation or precise retrieval.

SUMMARY AND CONCEPTS FOR SIG/CR

The NARCIS classification, criticisms above notwithstanding, supports a vital research portal that, in turn, supports a nationally-coordinated research effort designed to provide better inter-institutional communication of scholarly productivity. In many ways the NARCIS classification is typical of domain-dependent institutional knowledge organization systems. Unlike general bibliographic systems, these classifications are designed to meet specific domain requirements over and above either

user needs or general knowledge discovery priorities. Studying the NARCIS classification from the points offered in this paper is useful for SIG/CR.

The main points, to reiterate, are:

-Domain-dependence: the classification is derived by and for the research institutes of The Netherlands and therefore reflects the cultural imperatives of the Netherlands' research community, but at the expense of empirical knowledge representation.

-Asynthesis: the classification has no synthetic features, defeating any attempt at the use of facet analytical theory, which also risks obscuring knowledge representation of specific phenomena within its discipline-based silos.

-Interdisciplinarity: inter-, trans- and multi-disciplinarity are high priorities for global knowledge discovery; the classification isolates interdisciplinary communities, and obscures the phenomena of interest to interdisciplinary research.

-Scheme change: the classification has been overhauled once, likely creating the problems of subject ontogeny

The 2017 SIG/CR call for papers asked for "conceptual and technical issues of creating a relationship among ontologies." The four points raised in this paper serve as starting points for such a gathering of conceptual aspects of interoperability, as well as (one hopes) useful criticisms of a working domain-dependent classification.

ACKNOWLEDGMENTS

This work is part of the Visiting Professor Program of the Royal Netherlands Academy of the Arts and Sciences. The author is grateful to the entire NARCIS team of the Data Archiving and Networked Services Division (DANS) of the academy.

REFERENCES

Choi, Inkyung. (2017). Visualizations of Cross-cultural Bibligraphic Classification: Comparative Studies of the Korean Decimal Classification and the Dewey Decimal Classification. *Proceedings of the North American Symposium on Knowledge Organization 6*, 39-55. http://www.iskocus.org/NASKO2017papers/NASKO2017_paper_21.pdf

DANS a. Data Arching and Networked Services (DANS), Royal Netherlands Academy of the Arts and Sciences. "About NARCIS." https://www.narcis.nl/about/Language/en

DANS b. "NARCIS Classification." https://www.narcis.nl/classification/Language/en

Scharnhorst, Andrea, Richard P. Smiraglia, Christophe Guéret and Alkim Almila Akdag Salah. (2016). "Knowledge Maps of the UDC: Uses and Use Cases." *Knowledge Organization 43*:641-54.

Smiraglia, Richard P. (2014). *Cultural Synergy in Information Institutions.* New York: Springer, 2014.

Smiraglia, Richard P. (2015). Cultural Pervasiveness or Objective Violence?: Three Questions about KOS as Cultural Arbiters. In Melissa Adler (Ed.), *26th ASIS[T] SIG/CR Classification Workshop*, (pp. 9-13). DOI: http://dx.doi.org/10.7152/acro.v26i1.14981

Szostak, Rick, Claudio Gnoli & María López-Huertas. (2016). *Interdisciplinary Knowledge Organization.* Switzerland: Springer.

Tennis, Joseph T. (2002). Subject Ontogeny: Subject Access through Time and the Dimensionality of Classification. In Maria J. López-Huertas and Francisco J. Munoz-Férnandez (Eds.), *Challenges in Knowledge Representation and Organization for the 21st Century, Integration of Knowledge across Boundaries: Proceedings of the Seventh International ISKO Conference, 10-13 July 2002. Granada, Spain*, Advances in Knowledge Organization 8, (pp. 54-59). Würzburg: Ergon Verlag.

Tennis, Joseph T. (2006). "Versioning Concept Schemes for Persistent Retrieval." *Bulletin of the American Society of Information Science and Technology 32*(5), 13-16.

Tennis, Joseph T. (2007). Diachronic and Synchronic Indexing: Modeling Conceptual Change in Indexing Languages. In C. Arsenault & K. Dalkir (Eds.), *Information Sharing in a Fragmented World, Crossing Boundaries. Proceedings of the 35th Annual Meeting of the Canadian Association for Information Science/L'Association Canadienne Des Sciences De L'information, Montreal* [pp. 1-13]. Montreal: Canadian Association for Information Science.

Tennis, Joseph T. (2012). The Strange Case of Eugenics: A Subject's Ontogeny in a Long-Lived Classification Scheme and the Question of Collocative Integrity. *Journal of the American Society for Information Science and Technology 63*(7), 1350-1359. doi:10.1002/asi.22686

Tennis, Joseph T. (2013). Ethos and Ideology of Knowledge Organization: Toward Precepts for an Engaged Knowledge Organization. *Knowledge Organization 40*, 42-49.