

## Toward Better Interoperability of the NARCIS Classification

Journal:	<i>Knowledge Organization</i>
Manuscript ID	KO-2018-0081
Manuscript Type:	Special Issue

SCHOLARONE™  
Manuscripts

## Toward Better Interoperability of the NARCIS Classification

**Abstract:** Research information can be useful to science stakeholders for discovering, evaluating and planning research activities. In the Netherlands, the institute tasked with the stewardship of national research information is DANS (Data Archiving and Networked Services). DANS is the home of NARCIS, the national portal for research information, which uses a similarly named national research classification. The NARCIS Classification assigns symbols to represent the knowledge-bases of contributing scholars. A recent research stream in knowledge organization known as comparative classification uses two or more classifications experimentally to generate empirical evidence about coverage of conceptual content, population of the classes, and economy of classification. This paper builds on that research in order to further understand the comparative impact of the NARCIS Classification alongside a classification designed specifically for information resources. Our six cases come from the DANS project Knowledge Organization System Observatory (KOSo), which itself is classified using the Information Coding Classification (ICC) created in 1982 by Ingetraut Dahlberg. ICC is considered to have the merits of universality, faceting, and a top-down approach. Results are exploratory, indicating that both classifications provide fairly precise coverage. The inflexibility of NARCIS makes it difficult to express complex concepts. The meta-ontological epistemic stance of the ICC is apparent in all aspects of this study. Using the two together in the DANS KOS Observatory will provide users with both clarity of scientific positioning and ontological relativity.

### 1.0 Research information and classification

1  
2  
3 Research information serves many purposes. It can be useful to science stakeholders for  
4  
5 discovering, evaluating and planning research activities at both national and regional levels.  
6  
7

8 The principle stakeholders of research information are: researchers, research managers,  
9  
10 policy-makers, research councils, the media and the public (Nabavi, Jeffery and Jamali  
11  
12 2016). In the Netherlands, the institute tasked with the stewardship of national research  
13  
14 information is DANS (Data Archiving and Networked Services). DANS is an institute of the  
15  
16 Royal Netherlands Academy of the Arts and Sciences (Koninklijke Nederlandse Akademie  
17  
18 van Wetenschappen or KNAW) and the Netherlands Organisation for Scientific Research  
19  
20 (Nederlandse Organisatie voor Wetenschappelijk Onderzoek or NWO). DANS is the home of  
21  
22 NARCIS, the national portal for research information. It hosts the Netherlands' wide-ranging  
23  
24 data and research archiving structure (<https://dans.knaw.nl/en>).  
25  
26  
27  
28  
29  
30  
31  
32

33 The Dutch research infrastructure makes use of a national research classification named  
34  
35 after the National Academic Research and Collaborations Information System (NARCIS), a  
36  
37 DANS-maintained national research portal. NARCIS is an information repository of open  
38  
39 access publications and datasets from Dutch scholars, combined with texts of peer reviewed  
40  
41 publications and other research data (<https://dans.knaw.nl/en/about/services/narcis>). The  
42  
43 NARCIS Classification is a framework with which institutes and experts (scientists and  
44  
45 scholars) are identified and clustered symbolically in the infrastructure from which search  
46  
47 and retrieval are facilitated. The classification is designed to provide access to research  
48  
49 information (e.g., researcher names, institutional affiliations, etc.) for scholars who enter  
50  
51 their research into the repository. The NARCIS Classification assigns symbols to represent  
52  
53 the knowledge-bases of contributing scholars, rather than to represent the content of the  
54  
55 publications in its repository.  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 A recent analysis (Smiraglia 2017) demonstrated the strengths of the NARCIS classification  
7  
8 as its disciplinary base and its grounding in the NWO/KNAW milieu. The classification has  
9  
10 eight main classes: D10000 Science and Technology, D20000 Life Sciences, Medicine and  
11  
12 Health Care, D30000 Humanities, D40000 Law and Administration, D50000 Behavioural and  
13  
14 Educational Sciences, D60000 Social Sciences, D70000 Economics and Business  
15  
16 Administration and E10000 Interdisciplinary Sciences. The classes are seen as mutually  
17  
18 exclusive and collectively exhaustive and are derived by the governing bodies from the  
19  
20 institutional participants who supply data to the NARCIS portal.  
21  
22  
23  
24  
25  
26  
27

28 A recent research stream in knowledge organization known as comparative classification  
29  
30 uses two or more classifications experimentally to designate content from a test collection  
31  
32 (Szostak and Smiraglia 2017a; 2017b; 2018). The purpose is to generate empirical evidence  
33  
34 about a) coverage of conceptual content, including precision; b) population of the classes;  
35  
36 and c) economy of classification (in other words, how many expressions are required to  
37  
38 represent the content fully, comparatively?).  
39  
40  
41  
42  
43  
44

45 This paper builds on that research in order to further understand the comparative impact of  
46  
47 the NARCIS Classification in a particular DANS application alongside a classification designed  
48  
49 specifically for information resources. Our sample data come from the DANS project  
50  
51 Knowledge Organization System Observatory (KOSo), which itself is classified using the  
52  
53 Information Coding Classification (ICC) created in 1982 by Ingetraut Dahlberg, the founder  
54  
55 of the International Society for Knowledge Organization and principal proponent of a  
56  
57  
58  
59  
60

1  
2  
3 science of knowledge organization. ICC is considered to have the merits of universality,  
4  
5 faceting, and a top-down approach.  
6  
7  
8  
9

10 Our exploratory study has the following research question:  
11  
12  
13

14  
15 How do NARCIS Classification and Information Coding Classification compare when  
16  
17 applied to Knowledge Organization Systems in the social sciences, humanities, and  
18  
19 life sciences with regard to:  
20  
21

- 22  
23 a) coverage of conceptual content, including precision;  
24  
25 b) population of the classes; and,  
26  
27 c) economy of classification?  
28  
29  
30  
31

32 This research has several implications. First, in making this comparison we can discover  
33  
34 similarities and differences in the classification of a set of resources using the NARCIS  
35  
36 Classification and the ICC. From an academic point of view, the research points to the  
37  
38 potential interoperability of the NARCIS Classification. Practically for DANS, the results can  
39  
40 diagnose existing limitations of the NARCIS Classification and potentially highlight the path  
41  
42 towards improvement, at the same time incorporating the new Observatory more fully into  
43  
44 other DANS data resources.  
45  
46  
47  
48  
49  
50  
51  
52  
53

## 54 **2.0 Methodology**

55  
56  
57  
58  
59  
60

1  
2  
3 Our method is straightforward. All entities in the KOSo are classified using the ICC.  
4  
5  
6 Therefore, we simply also applied NARCIS classification codes  
7  
8 (<https://www.narcis.nl/classification/Language/en>) to each entity. It is important to  
9  
10 understand that the ICC was developed for information environments, which is to say it is a  
11  
12 meta-disciplinary classification (Bates 1999; Dahlberg 2008). The ICC works for the KOSo by  
13  
14 using together codes from Dahlberg's 1999 Classification System for Knowledge  
15  
16 Organization Literature. For example, The Canadian Parks Service Classification is classified  
17  
18 as "048" (classification)-9 (geographic reference). For the purpose of this comparative  
19  
20 analysis we have used only the ICC codes and not those designating types of systems.  
21  
22  
23  
24  
25  
26  
27  
28  
29

### 30 **3.0 Results**

#### 35 **3.1 Population of the two classifications**

36  
37  
38  
39  
40 The KOSs were classified into twenty-eight cases from the life sciences and 299 from the  
41  
42 social sciences and humanities—all those in the KOSo as of August 2018. All but two cases  
43  
44 from the life sciences were assigned existing NARCIS classification codes. Figure 1 shows the  
45  
46 most populous classes from the ICC coding. Most of the KOSs in the sample fell into biology  
47  
48 and agriculture, with fewer in aspects of medicine.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

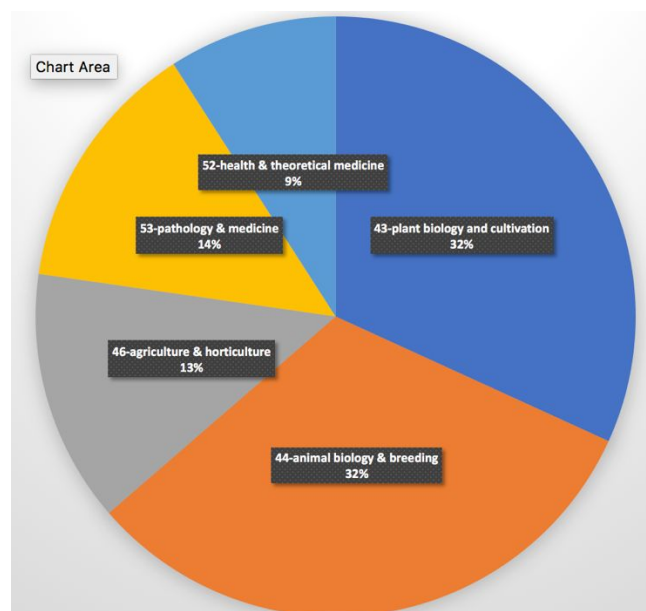


Figure 1. Most populous ICC classes in life sciences KOS sample.

All of the NARCIS classification codes came from the D2xxxx hierarchy; D identifies the “Science and Technology” hierarchy, and D2xxxx identifies “Life Sciences, Medicine and Health Care.” Almost half of the NARCIS codes applied fell in class D22600 “Zoology” within the D22000 “Biology” hierarchy. Smaller clusters fell in D21300 “Biochemistry” under “Life Sciences” and D23100 “Pathology” within “Medicine.” Thus, in the life sciences granularity is not different in either classification, although ICC was more flexible in two cases.

There was greater divergence in the social sciences and humanities, where 127 cases could not be classified with NARCIS classification codes, and 138 did not have ICC coding. Fifteen KOSs required two separate NARCIS classification codes to achieve coverage of the concepts. Figure 2 shows the most populous ICC classes in the social sciences and humanities sample. Most works fall into culture or the fine arts.

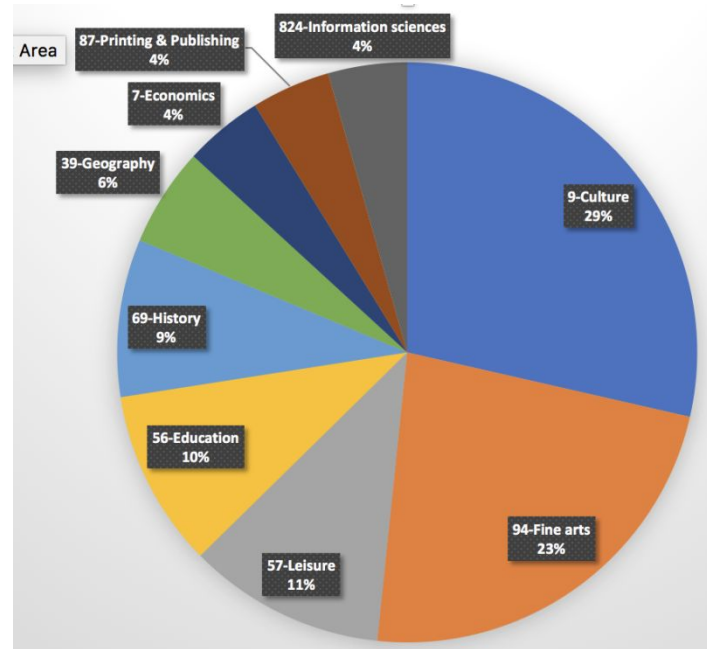


Figure 2. Most populous ICC classes in social sciences and humanities KOS sample.

The majority of the NARCIS codes fall into the D3xxxx “Humanities” range. The distribution of the most populous NARCIS classes is shown in Figure 3. The largest classes are “Paleography – library [sic] science” (for information sciences, especially taken together with computer science), economics, and arts and culture.



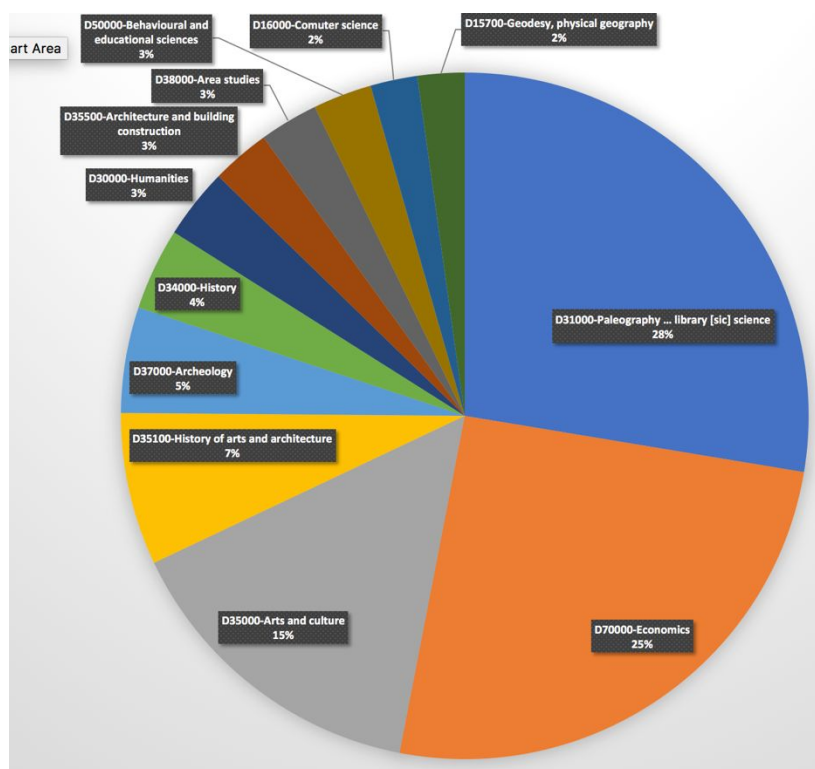


Figure 3. Most populous NARCIS classes in social sciences and humanities KOS sample.

Once again, although the sample is much more diverse, there is little difference between the two classifications in coverage or granularity.

### 3.2 Six case examples

As a kind of simple case study, we present six examples of comparative classification using NARCIS and ICC, three each from the social sciences and humanities and from the life sciences. Our first example (Figure 4) uses the Historical International Standard Classification of Occupations (HISCO).

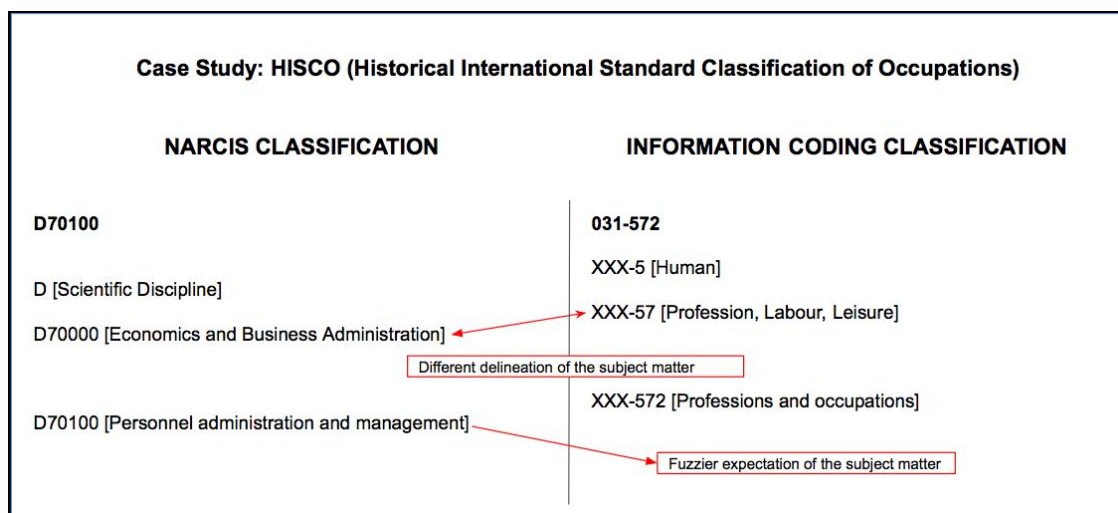


Figure 4. Comparative classification of the Historical International Standard Classification of Occupations.

In this case the ICC has placed all professions within the ontological meta-category of “Human Area,” and has broken the large class including profession, labor and leisure into a smaller class including professions and occupations, which presumably include personnel administration. NARCIS has relied solely on economics as a science and personnel administration as a subdivision of that science. The difference in both coverage and granularity in this case is apparent and arises from the narrow scientific focus of NARCIS over and against the more general aspirations of the ICC.

Our next example (Figure 5) is the Europeana Fashion Vocabulary (EFV).

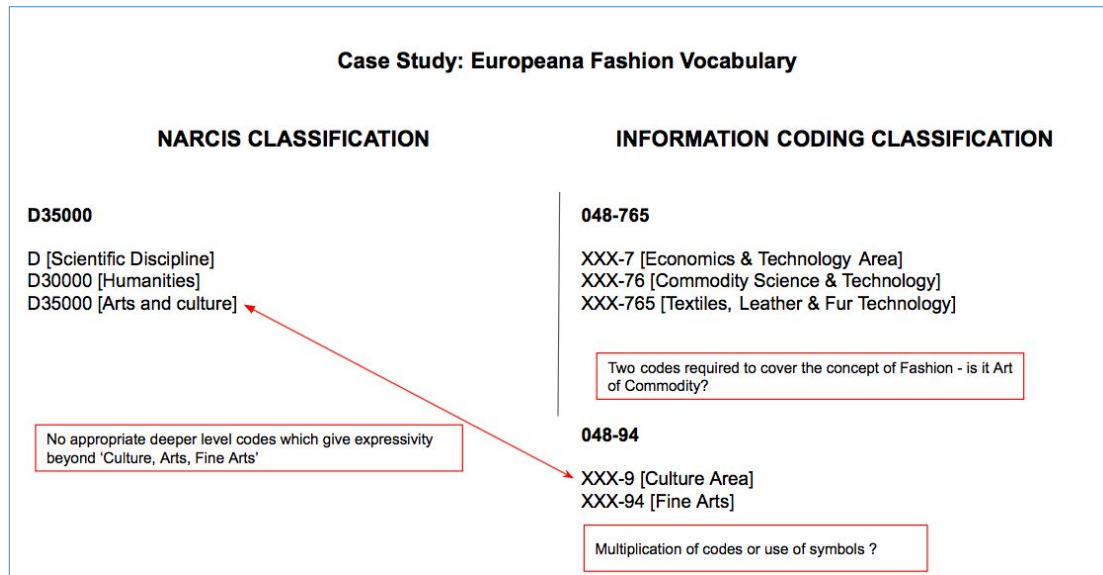


Figure 5. Comparative classification of the Europeana Fashion Vocabulary.

Fashion presents an interesting example of the difficulty of close classification of a domain that lies outside of typical academic interests. There is no good place for “fashion” in either classification. *Wikipedia* defines “fashion” as style in clothing and life accessories; and the *EFV* itself intends to represent the unity of all meanings of fashion. Using the NARCIS classification we can assign the code for “arts and culture” within the “humanities” but we cannot express the concept of commodity science. In the ICC we are able to express both “fine arts” as an aspect of “culture” and “textiles, leather & fur technology” as an aspect of commodity, which is closer yet still imprecise. This is a good example of the limits of the NARCIS classification, which is derived from the research infrastructure of Dutch universities and KNAW. The NARCIS database returns over 3,500 results for the term “fashion”, most of which are not classified. The top results among projects are projects concerning fashion in Dutch society—these are classified as D51000 Psychology, D35100 History of arts and architecture and D63000 Cultural anthropology. A quick check of the Universal Decimal

Classification (UDC) found that fashion appears at 391 under Cultural anthropology and 685.34 Fashion and 687.5 Aesthetics, both of which fall within industry.

Our final social science and humanities example is the well-known ICONCLASS iconographic classification frequently used in art history (Figure 6).

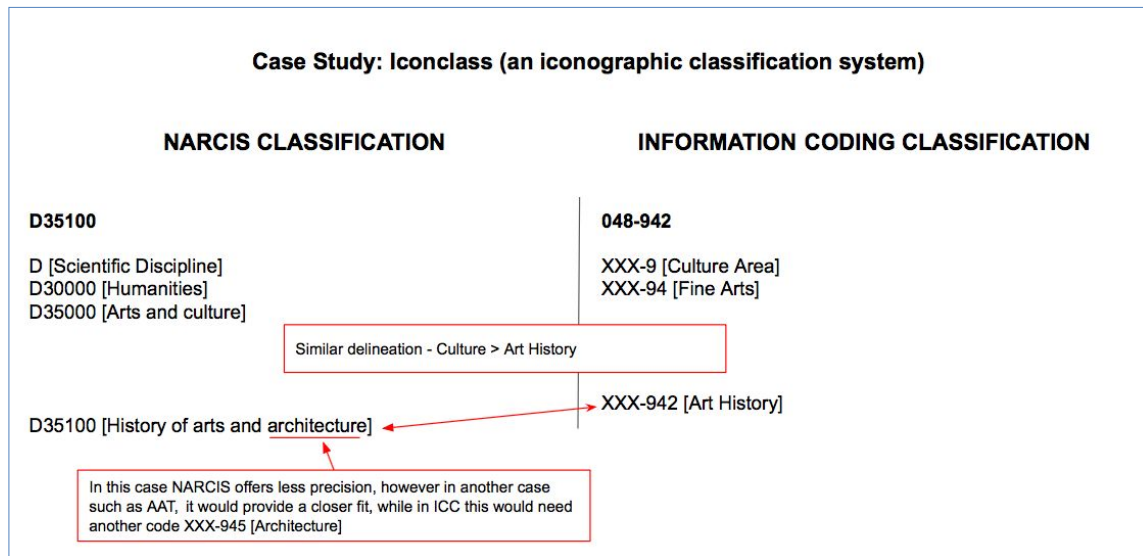


Figure 6. Comparative classification of ICONCLASS.

In this case ICC is able to express “art history” precisely, while NARCIS offers less precision. Interestingly, NARCIS would provide a close fit for the related *Art & Architecture* Thesaurus, because it expresses both art and architecture in one class.

Our first life science example is the International Statistical Classification of Diseases (ICD) (Figure 7).

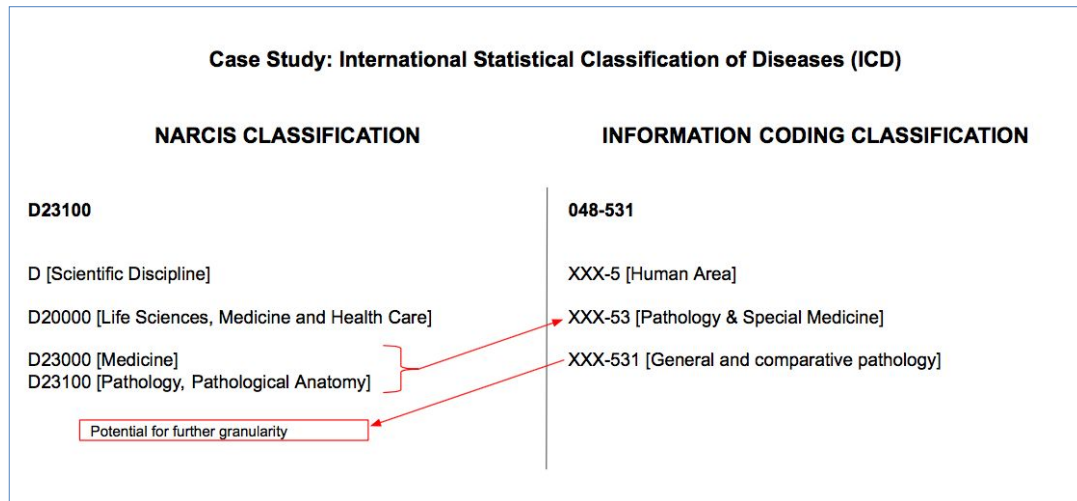


Figure 7. Comparative classification of the International Statistical Classification of Diseases.

Disease falls within pathology, which falls within medicine in both classifications. The ICC has a broader upper ontological level, however, of “Human Area,” incorporating many aspects of human life beyond the life sciences. The obverse is also apparent, that NARCIS is more closely focused directly on the sciences involved. ICC offers some potential for greater granularity than does NARCIS.

Our next example, the Catalogue of Life (CoL), is a species taxonomy that aims to index the world's known species of animals, plants, fungi and micro-organisms. Neither classification reaches the precision to describe the CoL without using two codes, the alternative is leaving the KOS classified under the rather broad bucket of ‘Biology/Bio-Area’

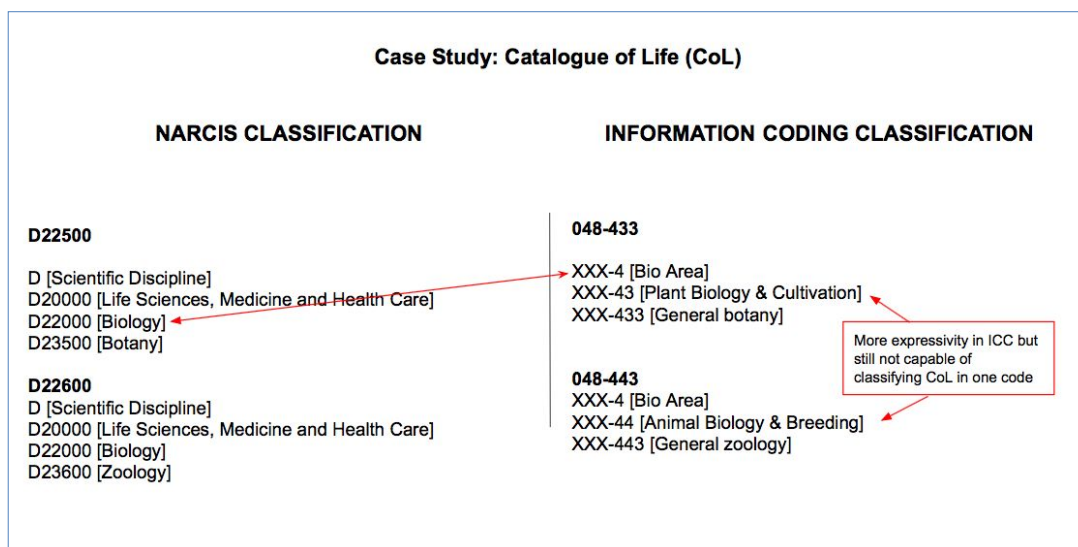


Figure 8. Comparative classification of the Catalogue of Life (CoL).

Both classifications require multiple class assignments to cover both botany and zoology.

The enumerative style of the NARCIS classification generates each domain in its own class under biology. There is slightly more expressivity in the ICC.

Our last example is the Chemical Entities of Biological Interest (ChEBI) ontology of molecular entities (Figure 9).

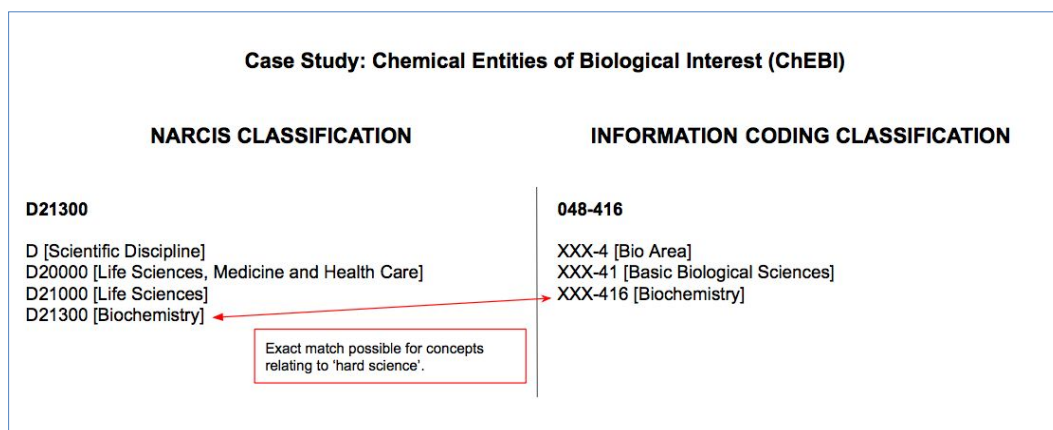


Figure 9. Comparative classification of Chemical Entities of Biological Interest (ChEBI).

1  
2  
3  
4  
5  
6 In this case both classifications offer a precise location; concepts relating to “hard science”  
7  
8 are explicitly expressed in both systems.  
9

#### 10 11 12 13 **4.0 Discussion** 14

15  
16  
17 We wanted to compare two ontological structures for the subjects of research. The  
18  
19 Dahlberg Information Coding Classification was designed specifically for research, and the  
20  
21 NARCIS Classification was designed specifically for the research in the NARCIS database.  
22  
23 Thus, the comparison should tell us whether either or both provide appropriate granularity  
24  
25 and coextensively—that is, can concepts in research be represented directly and fully in one  
26  
27 system or the other, or are the two comparable. We also wanted to see how using the  
28  
29 NARCIS Classification alongside ICC might enhance the information infrastructure at DANS.  
30  
31 At DANS the ICC is used as the primary classification for the KOSo project and NARCIS  
32  
33 Classification is used for everything else. Would using the two together provide  
34  
35 complementarity or redundancy?  
36  
37  
38  
39  
40  
41  
42  
43  
44

45 In the social science and humanities cases, two—occupations and art history—were  
46  
47 provided appropriate granularity in ICC. Fashion was not provided a hospitable location in  
48  
49 either. In the life sciences cases, one case—diseases—proved hospitable only at a meta level  
50  
51 but not precisely. Botany and zoology required multiple class assignments from both  
52  
53 systems to express the relation of the two in one context. In the last—biochemistry—both  
54  
55 systems offer precision.  
56  
57  
58  
59  
60

## 5.0 Conclusions

This study was intended to be exploratory, with the goal of creating better empirical understanding of how these two classifications might contribute to the enrichment of resources in DANS, in particular the KOSs described in the Observatory. Results, therefore, are themselves exploratory. We see that both classifications provide fairly precise coverage. In only a few cases was it difficult to find a precise code in either classification. We see also that the inflexibility of NARCIS makes it difficult to express complex concepts, such as those expressed more easily in ICC and other general classifications. The meta-ontological epistemic stance of the ICC is apparent in all aspects of this study. That is, NARCIS provides more clarity and granularity in the representation of sciences, but ICC provides better ontological structuring of sciences within other realms of human knowledge. This result is tautological in part, because of course, those are divergent goals of the two classifications. It does suggest that using the two together in the DANS KOS Observatory will provide users with both clarity of scientific positioning, on the one hand, and ontological relativity, on the other. It would be useful in the application of the NARCIS classification in the Observatory if it were more flexible in terms of offering means of synthesis for expressing complex concepts.

## References

- Bates, Marcia J. 1999. "The Invisible Substrate of Information Science." *Journal of the American Society for Information Science* 50: 1043-50.



1  
2  
3 Dahlberg, Ingetraut. 1999. "Classification System for Knowledge Organization Literature."  
4

5 *Knowledge Organization* 26: 192-202.  
6  
7

8 Dahlberg, Ingetraut. 2008. "Information Coding Classification: (ICC): A Modern, Theory-  
9

10 Based Fully-Faceted, Universal System of Knowledge Fields" *Axiomathes* 18: 161-76.  
11

12 doi 10.1007/s10516-007-9026-8  
13  
14

15 Fashion. 2018. s.v. *Wikipedia*. <https://en.wikipedia.org/wiki/Fashion>  
16  
17

18 Majid Nabavi, Keith Jeffery, Hamid R. Jamali. 2016. "Added Value in the Context of Research  
19

20 Information Systems." *Program* 50: 325-39. doi.org/10.1108/ PROG-10-2015-0067  
21  
22

23 Smiraglia, Richard P. 2017. "Disciplinary, Asynthetic, Domain-Dependent: NARCIS a National  
24

25 Research Classification in Isolation." *Advances in Classification Research*, 2017,  
26

27 October 27, 2017, Washington, DC, USA.  
28  
29

30 Smiraglia, Richard P. and Rick Szostak. 2018. "Converting UDC to BCC: Comparative  
31

32 Approaches to Interdisciplinarity." *Challenges and Opportunities for Knowledge*  
33

34 *Organization in the Digital Age: Proceedings of the Fifteenth International ISKO*  
35

36 *Conference, 9-11 July 2018, Porto, Portugal*, ed. Fernanda Ribeiro, Maria Elisa  
37

38 Cerveira. *Advances in Knowledge Organization* 16. Würzburg: Ergon Verlag, 530-38.  
39  
40

41 Szostak, Rick and Richard P. Smiraglia. 2017. "Comparative Approaches to Interdisciplinary  
42

43 KOSs: Use Cases of Converting UDC to BCC." *NASKO 2017: Sixth North American*  
44

45 *Symposium on Knowledge Organization*. <http://www.iskocus.org/nasko2017->  
46

47 program.php  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60