

**Claudio Gnoli, Andreas Ledl, Ziyoung Park, Marcin Trzmielewski**

## **Phenomenon-based vs. disciplinary classification: possibilities for evaluating and for mapping**

### **Abstract**

The possibility of interdisciplinary research through online library catalogues is today a major challenge for knowledge organization. By examining the different approaches used to index documents in integrated library systems, interdisciplinary researchers may find that classifications based on academic disciplines are less appropriate for this type research when compared with phenomenon-based systems. Based on this consideration, the aim of the present paper was to compare two systems having different conceptual structures: the Dewey Decimal Classification (DDC), which is based on an epistemological approach, and the Integrative Levels Classification (ILC), based on an ontological one. We also explored how ILC phenomenon classes can be mapped to corresponding DDC disciplines and how DDC can be used in phenomenon-like ways. This work highlights the importance of future studies in phenomenon-based classification for the development of new indexing schemes and for more powerful applications of already existing disciplinary systems and opens a new way for improvement of interdisciplinary research.

### **1. Introduction**

An important trend in recent research in knowledge organization (KO) is the criticism of academic disciplines as the main dividing principle of classification schemes (Beghtol 2010, 1056; Broughton 2015; Gnoli 2016). There have been many claims that disciplinary classifications may act as an obstacle to interdisciplinary research (Szostak *et al.* 2016), and may leave out such important applications as serious leisure, services for citizens looking for practical information associated to real-life events (Bonner 1982), particularly in e-government, or vendors' catalogues of such products as clothes and food, which do not fit in any academic discipline. However, phenomena seem to be a more general unit of knowledge than disciplines in ontological terms, so that general systems could be primarily defined in terms of phenomena, and a discipline could then be defined as a phenomenon belonging to the class of human culture.

Experimental phenomenon-based classification systems have been developed since the times of James Duff Brown's Subject Classification (Beghtol 2004). An in-depth enquiry towards a general classification of phenomena was developed by the Classification Research Group (CRG) (Austin 1969; CRG 1969) and, although this did not lead to any definitive system, two schemes were produced by CRG members, namely the Bliss Bibliographic Classification 2nd edition and the Broad System of Ordering, which do provide some main classes (3 and 088 respectively) under which phenomena should be classed along with classes for disciplines. CRG research has also inspired the Integrative Levels Classification (ILC) project, which has now been in

development for a decade and currently offers a general scheme with more than 8,000 classes and facets (Gnoli 2017).

On the other hand, the classification schemes used in most libraries and their online catalogues are still the classical disciplinary ones, like the Dewey Decimal Classification (DDC), the Universal Decimal Classification (UDC) or the Library of Congress Classification (LCC). This situation suggests that a comparative evaluation of phenomenon-based vs. disciplinary indexing, as well as an exploration of possible ways to create cross-references and mappings between the two kinds of systems, may have important implications for the future of classification. This paper explores such possibilities by specifically considering DDC as an example of a disciplinary system and ILC as an example of a phenomenon-based system.

## 2. Methods

Preliminary tests of comparison between DDC and ILC were performed by indexing small samples of items using both systems and considering how the resulting orderings and retrieval possibilities differ. An initial sample consisted of books on nature conservation held at the University of Pavia ecology library (Szostak *et al.* 2016, 104-106).

Experiments have recently continued by applying both DDC and ILC to items recorded in the Basel Register of Thesauri, Ontologies and Classifications (BARTOC) maintained by Basel University Library (Ledl and Gnoli, 2017). As a directory of thousands of knowledge organization systems (KOS), BARTOC covers the whole spectrum of knowledge fields, but does not require a very detailed indexing as the coverage of a KOS is usually quite broad. All BARTOC items, as they are input into the directory, are indexed by one or more three-digit DDC classes freely combined and by one or more EuroVoc terms. Additionally, using one or more ILC classes, also freely combined, we have indexed a KOS sample consisting of the first 200 Top-Rated Systems (that is the systems which have been voted as most relevant by BARTOC users) plus all the ten remaining KOSs dealing with health care; the latter having been chosen in order to test ILC within a specific domain of knowledge. To allow reference to a stable source, the first edition (ILC1) of ILC was used, despite a second edition (ILC2) currently being developed which includes various changes and improvements. Update of classmarks in BARTOC to ILC2 is envisaged in the medium term.

Further work has consisted in mapping between DDC and ILC main classes, by creating an additional "DDC map" field in the MySQL database of ILC2. DDC appears to be a suitable reference for mapping to a disciplinary system, given its worldwide use. For similar reasons, terms in the Nuovo Soggettario subject heading system, developed by the National Library of Florence and based on a structural conception not far from that of ILC, are also mapped to DDC.

At the time of writing this paper (January 2018), the field has been compiled for 25

ILC main classes labelled *a* to *y* and for 100 DDC subdivisions labelled *010* to *990*. DDC numbers, taken from WebDewey, the updated current edition of the scheme, have been used in full, so that, for example, ILC *f* "molecules, chemical substances" has been mapped to DDC *541.22* "molecular structure (in theoretical chemistry)".

Availability of DDC classmarks in BARTOC can subsequently allow leverage of ILC-DDC equivalences to automatically produce additional ILC classmarks, although these are expected to be less accurate than the classmarks produced intellectually because (1) the DDC classes used are very broad, (2) the very conceptual structure of the two systems is different. Concerning the latter point, some considerations are developed in the following part of this paper.

### 3. How to map a phenomenon to a discipline

As a disciplinary system, DDC scatters a given phenomenon, such as "water", into a variety of different disciplinary classes (chemistry, hydrology, hydraulic engineering, architecture, ...). This can easily be observed by checking a term for the phenomenon in the DDC Relative Index, where references are provided to its occurrence in the various classes. However, the phenomenon "water" cannot be separated from the final class numbers assigned to the information objects. On the other hand, in ILC, water has a place of unique definition (Farradane 1950) at *febb* under class *f* of chemical substances; it can then be combined freely with any other concept, e.g. *wkf7febb* "fortifications with water".

So, should *febb* be mapped to a DDC class in chemistry, in hydrology, in hydraulic engineering, or in architecture? Concepts like "water" have been described by a DDC editor as *topics*, which could indeed be useful for "identifying equivalence and hierarchical relationships between Relative Index headings, one of the steps required to fully implement a topic-based data model for the DDC" (Green 2014). Topics can thus be another relevant unit for classification, in some way complementary to disciplinary classes:

"As a knowledge organization system, Dewey is an analytico-synthetic classification system first and foremost organized by disciplines, in which topics and concepts are scattered throughout. What can be done with Dewey as a classification system (that is, because it groups topics into classes) that couldn't be done if it were strictly topic/concept-based, like the typical thesaurus? [...] Many of the topics within a class occur as Relative Index terms. Within WebDewey, these are formatted in Marc 21 for Authority data and include references. Can/should Relative Index terms be used (as supplements to subject headings or independently) in indexing?" (Green 2017)

Green's idea involves the establishment of topic-to-topic relationships, which could improve search in DDC applications, such as a library catalogue: users could thus navigate from a water-related class to another in a different discipline.

Although topics are not defined precisely and are labelled by a term other than *phenomena*, it seems that the two units can work in similar ways, at least functionally.

Thus, Green's suggestion opens up the possibility of using DDC in some phenomenon-like ways.

Among other things, topics can work as a better reference than disciplines in mapping DDC to different systems, including ILC itself or subject headings and thesauri, which usually index phenomena rather than disciplines. A search for "water" in the Web Dewey Relative Index yields many different classes, including 551.35 "water--geological agents", 354.36 "water--public administration", 714 "water--landscape architecture", etc. However, all these are preceded by a number for the topic "water" alone, 553.7, also known as its interdisciplinary number.

To be precise, even the number 553.7 is not discipline-neutral, as it belongs to the hierarchy of 550 "Earth sciences and geology", which is a discipline too; still it is indicated by the Relative Index as the most general number available for the topic "water". As such, 553.7 is supposed to be used to class interdisciplinary works, like a book dealing with water from many different perspectives. Many topics in the Relative Index are provided with such interdisciplinary numbers. These are also used to map the terms of Nuovo Soggettario to DDC (Lucarelli pers. com.). Farradane would say that the place of unique definition of water in DDC is Earth sciences. A phenomenon-based system would be ontologically more precise, claiming that the place of unique definition of water is terrestrial materials or substances, which can be studied typically by Earth sciences but also by a variety of other disciplines including chemistry, astronomy, engineering, architecture, etc. In the ILC schedules, the most typical disciplines which study a certain class of phenomena are indicated in square brackets, *e.g.*:

*f*"molecules, chemical substances, pure substances" [chemistry, molecular physics]

#### 4. Comparing the structures of DDC and ILC

Our mapping work also provides interesting opportunities to compare the disciplinary system and the phenomenon-based one. For example, the order of main classes of phenomena in ILC2 can be compared with the order implied by the corresponding DDC numbers:

<i>a</i> forms, mathematical objects [formal sciences, logic, mathematics]	≈ DDC	510, 160
<i>b</i> spacetime		530.1
<i>c</i> branes [string theory, M theory]		539.7258
<i>d</i> energy, elementary particles, waves [quantum physics, high energy physics]		539.7
<i>e</i> atoms, chemical elements [atomic physics]		541
<i>f</i> molecules, chemical substances [chemistry, molecular physics]		540
<i>g</i> continuum bodies, bulk matter, states of matter [classical mechanics]		530.4, 530
<i>h</i> celestial bodies, cosmic structures [astronomy, astrophysics, cosmology]		520
<i>i</i> minerals and rocks [geology]		550
<i>j</i> land, landforms, regions [physical geography, topography, geomorphology]		910
<i>k</i> genes, genetic code [genetics]		576.5
<i>l</i> bacteria, prokaryote organisms [microbiology]		579.3, 570
<i>m</i> organisms, eukaryote organisms		570

<i>n</i> populations, groups, wildlife [ecology]	577
<i>o</i> instincts, action patterns, behavioural modules [ethology, neurosciences]	591.5
<i>p</i> consciousness, subjects [psychology]	150
<i>q</i> languages, idioms [linguistics]	400, 410
<i>r</i> rituals, tradition, mores, spirituality, religion, sacred, holy	200
<i>s</i> communities, civil society, complex society, groups, people [sociology]	300
<i>t</i> politics, political power, governments [political sciences]	320
<i>u</i> economics, economic systems [economics]	330
<i>v</i> technologies, technological systems, production systems [applied sciences]	600
<i>w</i> artifacts, tools	620
<i>x</i> artworks, art, expressions [aesthetics, art criticism]	700
<i>y</i> public knowledge, recorded knowledge, intellectual works, cultural heritage	001

Inversely, ILC classes (of which only *a* to *j* have currently been updated in ILC2) can be rearranged using the corresponding DDC notation:

000 ≈ <i>V</i>	<i>general class, the universe of knowledge</i>
004 ≈ <i>wtu</i>	<i>computers [computer science, information technology]</i>
010 ≈ <i>y99x</i>	<i>indexes, bibliographies, catalogues</i>
020 ≈ <i>spgiWl</i>	<i>library and information services [library and information science]</i>
030 ≈ <i>y99y</i>	<i>dictionary, encyclopedia, directory</i>
050 ≈ <i>y99j</i>	<i>academic journal, research serial</i>
060 ≈ <i>so</i>	<i>organized civil society, NGOs [organizational studies]</i>
070 ≈ <i>y99e</i>	<i>newspapers</i>
080 ≈ <i>y976t</i>	<i>references [documentation]</i>
090 ≈ <i>y97</i>	<i>document instantiations, knowledge items</i>
100 ≈ <i>yy</i>	<i>philosophy, wisdom</i>
110 ≈ <i>aWz</i>	<i>world, universe [ontology, general philosophy]</i>
120 ≈ <i>p98</i>	<i>objects, phenomena, contents, qualia, representations</i>
130 ≈ <i>yg</i>	<i>magic knowledge</i>
150 ≈ <i>p</i>	<i>consciousness, subjects [psychology]</i>
170 ≈ <i>p938</i>	<i>values, sakes, goals, happiness</i>
180 ≈ <i>yy99</i>	<i>philosophical schools</i>
190, 140	≈ <i>yy99m</i> <i>modern Western philosophy</i>
200 ≈ <i>r</i>	<i>rituals, tradition, mores, spirituality, religion, sacred, holy</i>
210 ≈ <i>r99</i>	<i>ritual kinds</i>
220 ≈ <i>r97s</i>	<i>scriptures, religious texts</i>
230/280 ≈ <i>rt</i>	<i>Christianity</i>
290 ≈ <i>rX</i>	<i>faiths, doctrines</i>
300 ≈ <i>s</i>	<i>communities, civil society, groups, people [sociology]</i>
310 ≈ <i>s97</i>	<i>members</i>
320 ≈ <i>t</i>	<i>polities, political units, governments [political sciences]</i>
330 ≈ <i>u</i>	<i>economics, economic systems [economics]</i>
340 ≈ <i>t93</i>	<i>regulations, institutions, legal order, laws, norms, rules, acts [law]</i>
350 ≈ <i>t95</i>	<i>procedures, protocols, administration, management, power exercise</i>
360 ≈ <i>sp</i>	<i>public institutions, public bodies, services of general interest</i>
370 ≈ <i>spe</i>	<i>schools [didactics, pedagogy]</i>
380 ≈ <i>wn</i>	<i>vehicles, means of transport</i>
390 ≈ <i>r95</i>	<i>rites, practices, worships, uses, habits</i>
400/410 ≈ <i>q</i>	<i>languages, idioms [linguistics]</i>

420 ≈ qemi	<i>English language</i>
430 ≈ qemk	<i>German language</i>
440 ≈ qeir	<i>French language, Langue d'oïl</i>
450 ≈ qeit	<i>Italian language</i>
460 ≈ qeiiWk	<i>Castilian language, Spanish language; Galician language; Portuguese language</i>
470 ≈ qeia	<i>Latin language</i>
480 ≈ qeg	<i>Greek language</i>
490 ≈ qX	<i>language families</i>
500 ≈ bWo	<i>nature [natural sciences]</i>
510, 160 ≈ a	<i>forms, mathematical objects [formal sciences, logic, mathematics]</i>
520 ≈ h	<i>celestial bodies, cosmic structures [astronomy, astrophysics, cosmology]</i>
530 ≈ b	<i>spacetime, events, pacha</i>
540 ≈ f	<i>molecules chemical substances, pure substances [chemistry, molecular physics]</i>
550 ≈ i	<i>minerals and rocks [geology]</i>
560 ≈ m9ly	<i>fossils</i>
570 ≈ m	<i>eukaryote organisms</i>
580 ≈ mp	<i>plants [botany stricto sensu]</i>
590 ≈ mq	<i>animals, metazoa [zoology]</i>
600 ≈ v	<i>technologies, technological systems, production systems [applied sciences]</i>
610 ≈ sph	<i>health care institutions, healthcare [medicine]</i>
620 ≈ w	<i>artifacts, tools</i>
630 ≈ vaWp	<i>agriculture, primary sector</i>
640 ≈ vq	<i>cooking, cuisine [food science, gastronomy, cookery]</i>
650 ≈ u5o	<i>corporations, legal entities, companies, firms</i>
660/680 ≈ vt	<i>industry</i>
690 ≈ wh	<i>home appliances, domestic appliances</i>
700 ≈ x	<i>artworks, art, expressions [aesthetics, art criticism]</i>
710 ≈ wl	<i>settlements [town planning]</i>
720 ≈ wk	<i>buildings [architecture]</i>
730 ≈ xd	<i>ceramics, pottery</i>
730 ≈ xb	<i>sculptures</i>
740, 760 ≈ xg	<i>drawings</i>
750 ≈ xf	<i>paintings</i>
770 ≈ xh	<i>photographic art, reproductive arts</i>
770 ≈ xs	<i>films, movies, motion pictures, cinema</i>
780 ≈ xm	<i>music [musicology, musical criticism]</i>
790 ≈ xx	<i>sport, games</i>
800/890 ≈ xl	<i>literature [philology, literary theory, literary criticism]</i>
900, 940/990 ≈ t91	<i>historical periods [history]</i>
910 ≈ t92	<i>jurisdictions, territories [political geography]</i>
910 ≈ j	<i>land, landforms, regions [physical geography, topography, geomorphology]</i>
920 ≈ pY	<i>persons, individuals</i>
930 ≈ t91d	<i>Antiquity, 999 BC to 499 AD [ancient history]</i>

Several remarks can be made from examining these comparative schedules. An obvious one is the different arrangement of the two systems, already discussed by Ledl and Gnoli (2017): while ILC follows the order of integrative levels in nature, thus taking an ontological approach, DDC follows the traditional arrangement of disciplines of reason, of imagination, and of memory, taking an epistemological approach. The

difference in approach can be seen clearly in such classes as DDC *001* "knowledge" or *030* "dictionaries", which in a DDC epistemological view are premised to the whole of knowledge, but from the ILC ontological perspective are simply kinds of high-level cultural phenomena in class *y*.

Another consideration is that some main classes of phenomena in ILC, such as *d* "elementary particles and waves" or *l* "prokaryote organisms", are represented in DDC by several digits as they are only subclasses quite deep in the hierarchy; conversely, main DDC classes such as *800* "literature" or *410* "English language" are just subclasses (of *x* "arts" and *qem* "Germanic languages") in ILC. Indeed, the top level classes are very different between the two classification schemes.

A further point of interest is that some entries in ILC are mapped to more than one entry in DDC:

*190, 140*  $\approx$  *yy99m* modern Western philosophy  
*230/280*  $\approx$  *rt* Christianity  
*400/410*  $\approx$  *q* languages, idioms  
*510, 160*  $\approx$  *a* forms, mathematical objects  
*660/680*  $\approx$  *vt* industry  
*740, 760*  $\approx$  *xg* drawings  
*800/890*  $\approx$  *xl* literature  
*900, 940/990*  $\approx$  *t91* historical periods

## 5. Cross references by phenomena

Topics, as introduced above, are also an implicit basis for cross-references within the DDC (Green 2014). These can be recorded in WebDewey as "see-also" references, or in library systems as links between classes provided in interfaces for navigation.

The latter kind of relationships is, unfortunately, not common in library OPACs, but has been successfully implemented in the SciGator web-based browsing interface (Lardera *et al.* 2017). This application allows navigation between DDC main classes and a selection of their subclasses before launching a search by DDC notation in the University of Pavia online catalogue.

SciGator's most original feature is query expansion using cross-references between DDC classes or their mapping with local shelving schemes. Topics, and indirectly phenomena, are indeed an intuitive basis for the identification and establishment of these cross-references in SciGator. Behind links between fluid dynamics in physics, hydrology in Earth sciences and hydraulic engineering in applied sciences, which are created by librarians in Pavia for SciGator users, stands the concept of water as a phenomenon. Such cross referencing has the potential of being more formalized and connected with topics listed in the DDC Relative Index.

## 6. Conclusions

Our experience shows that, while both DDC and ILC may be used in a post-coordinate way to assign a plurality of classes to each knowledge item, *e.g.* to each KOS listed in BARTOC, the general ordering of items and the display of search results are very different if phenomena are considered as primary grouping criteria instead of disciplines. The same can be observed in the schedules of DDC and ILC main classes if arranged by one or the other system.

Work on mapping the two systems in more detail needs to be continued, as does indexing of BARTOC by ILC both intellectually and automatically, so as to make it more complete, accurate and updated as the development of ILC continues. Evaluation tests with a more formal methodology are planned.

Overall, the experiments we have performed using both disciplinary DDC and phenomenon-based ILC for indexing books (first test in Pavia), library catalogue items (SciGator) and KOS directory items (BARTOC) look promising. In particular, some notions implicit in existing disciplinary classifications, such as topics and interdisciplinary numbers in the Relative Index, could be analyzed and formalized more precisely to provide bridges towards phenomenon-based systems like ILC or Nuovo Soggettario.

These experiences suggest that continued research in phenomenon-based classification is an important requisite both for the development of innovative KOSs and for more powerful applications of existing disciplinary systems.

## References

- Austin, Derek (1969). Prospects for a new general classification. *Journal of librarianship* 1(3): 149-169.
- Beghtol, Clare (2004). Exploring new approaches to the organization of knowledge: James Duff Brown. *Library trends* 52: 702-718.
- Beghtol, Clare (2010). Classification theory. In *Encyclopedia of library and information sciences*. 3rd ed. Edited by Marcia J. Bates and Mary Niles Maack. London: Taylor and Francis, 1.045-1.060.
- Bonner, Robin A. B (1982). Community information classification research project: user oriented empirical methods of classification construction. In *Universal classification: proceedings of the 4th international study conference on classification research, Augsburg, 1982*. Edited by Ingetraut Dahlberg. Frankfurt: Indeks, 227-234.
- Broughton, Vanda (2015). *Essential classification*, 2<sup>nd</sup> ed. London: Facet.
- CRG: Classification Research Group. 1969. *Classification and information control*. London: Library Association.
- Farradane, Jason (1950). A Scientific theory of classification and indexing and its practical applications. *Journal of documentation* 6(2): 83-99.
- Gnoli, Claudio (2016). Classifying phenomena, part 1: Dimensions. *Knowledge organization* 43(6): 403-415.

- Gnoli, Claudio (2017). Classifying phenomena, part 2: Types and levels. *Knowledge organization* 44(1): 37-54.
- Green, Rebecca (2014). Moving towards a topic-based DDC. In *25th SIG/CR Classification Research Workshop*. Advances in classification research. Seattle: University of Washington. Available at: <http://journals.lib.washington.edu/index.php/acro/article/view/14905>.
- Green, Rebecca (2017). What more do we want? In *EDUG 2017 Meeting, 27-28 June, Paris, France*. <http://edug.pansoft.de/tiki-index.php?page=2017+meeting>. Accessed January 4, 2018.
- Lardera, Marco, Claudio Gnoli, Clara Rolandi and Marcin Trzmielewski (2017). Developing SciGator, a DDC-based library browsing tool. *Knowledge organization* 44(8): 638-643.
- Ledl, Andreas and Claudio Gnoli (2017). Indexing KOSs in BARTOC by a disciplinary and a phenomenon-based classification: preliminary considerations. In *Faceted classification today: theory, technology and end users: proceedings of the International UDC Seminar, 14-15 September 2017, London, United Kingdom*. Edited by Aida Slavic and Claudio Gnoli. Würzburg: Ergon, 109-117.
- Szostak, Rick, Claudio Gnoli and María López-Huertas (2016). *Interdisciplinary knowledge organization*. Cham: Springer.