

Classification Research

Classification & Authority Control: Expanding Resource Discovery

Daniel Martínez-Ávila

São Paulo State University–UNESP, Department of Information Science,
Av. Hygino Muzzi Filho, 737, São Paulo Marília Brazil (17525-900),
<dmartinezavila@marilia.unesp.br>



The fifth International UDC Seminar, entitled “Classification & Authority Control: Expanding Resource Discovery,” was held in The National Library of Portugal in Lisbon on 29-30 October 2015. This series of seminars is organized by the UDC Consortium to advance research on bibliographic classifications and enable communication between developers and classification users, the Universal Decimal Classification (UDC) in particular. This fifth edition of the UDC Seminars focused on using authority control, classification schemes, and linked data for a better integration of knowledge and resource discovery. The proceedings were edited by Aida Slavic and Maria Inês Cordeiro (2015) and published by Ergon Verlag (also available for purchase at the UDCC website <http://seminar.udcc.org/2015/proceedings.php>). This analysis follows the order of the program. The slides and abstracts of those invited talks that are not included in the proceedings were made available at the UDC Consortium website.

The keynote address, “Classifications, Links and Contexts,” was presented by Michael K. Buckland, Professor Emeritus at the School of Information, University of California, Berkeley (USA). Buckland talked about contexts and the tension between standardized relationships (exemplified by Paul Otlet’s modernist universalism, see 1934 and 1935, and the Semantic Web), as well as the particular, subjective situations in which individuals try to make sense (exemplified by Ludwik Fleck’s emphasis on the influence of local cultural contexts, 1935). In other words, this is another example of tension “between the local and the global.” Buckland discussed different approaches and key aspects of Fleck’s thesis such as meaning, sense and context, vocabulary and context, and links and context, applied to classification, linked data, its relations, and discovery tools. This paper aimed to complement Otlet’s view and to overcome some of the problems of standardized languages.

The first session of the conference was “Past and Future Perspectives on Subject Data Assets,” chaired by Widad Mustafa El Hadi (ISKO France). The first paper of this group was “Complementarity of Perspectives for Resource Descriptions” by Barbara B. Tillett. She reviewed some issues and experiences related to bibliographic data and the multiple perspectives of authority control (including name authorities and subject authority control). Tillett outlined the challenges of mapping subjects across languages, mapping terminology across languages and systems for concepts or topics, mapping terminologies across languages with multilingual thesauri, and even one-to-one mapping of terms within one language (due to regional variations). The author discussed the possibilities of considering the different perspectives to overcome these problems. This approach was inspired by Marcia Ascher’s “Ethnomathematics: a Multicultural View of Mathematical Edeas” (1994), where “different cultures have different ways of counting and measuring things and none are ‘wrong.’” Tillett states that subject terminology and classification numbers do not need to be mapped across languages or systems as multiple numbers can be assigned reflecting those multiple perspectives that, for not being unique, do not necessarily have to be wrong. Finally, Tillett also discusses these aspects for the context of linked data and the future, while calling for libraries sharing globally what they do locally. The following invited talk was given by Maria Inês Cordeiro, Director of the National Library of Portugal, and is entitled “Libraries, Classifications and the Network: Bridging Past and Future.” In her presentation, Cordeiro revisited past

milestones related to classification and subject vocabularies, and related them to present and future challenges of the Semantic Web, Linked Open Data, and beyond. The last invited talk of this first session was “Linking Library Data: Contributions and Role of Subject Data” by Nuno Freire, Chief Data Officer at The European Library at The Hague, Netherlands. This author outlined some of the initiatives on Library Linked Data (LLD) and Linked Open Data (LOD) datasets of The European Library, as well as spoke on the challenges of linking bibliographic subject data and classification.

The second session dealt with “Data Models and Semantic Structures” and was chaired by Clément Arsenault (Canada). The first invited talk for this session was given by Maja Žumer, University of Ljubljana, Slovenia, and Marcia Lei Zeng, Kent State University, USA, entitled “Application of FRBR and FRSAD to Classification Systems.” A starting point for their paper is the possibility of extension of the FRSAD conceptual model beyond controlled vocabularies (its original focus) to model classification data. The paper uses the *Dewey Decimal Classification* and the *Universal Decimal Classification* as case studies to test the applicability of the FRSAD model for classification data and the applicability of FRBR for modeling versions, such as different adaptations and different language editions. These applications are intended to be reviewed and approved in 2016. Aspects such as the differences between classification (e.g. *DDC*) as a work and edition (e.g. *DDC* 22) were discussed, as well as the importance of thema-based mapping (i.e. between classes) instead of nomen-based mapping (i.e., between notations associated) for semantic interoperability. The second paper of this session is entitled “Relational Aspects of Subject Authority Control: The Contributions of Classificatory Structure” by Rebecca Green, assistant editor of the *DDC* at OCLC. The paper begins introducing the differences and similarities between thesauri and classification systems. Among the latter, the author lists the hierarchical relationships, equivalence relationships, and associative as relationships paradigmatic relationships. These relationships in the *DDC* (both the tables and the Relative Index) are discussed for the context of FRSAD. Syntagmatic relationships in the *DDC* are also studied by Green. The last paper of this session is “Distributed Person Data: Using Semantic Web Compliant Data in Subject Name Headings” by Violeta Ilik. The starting point of her paper is the importance and need to associate person IDs and URIs with subjects when a named person is the subject of the document. However, the author is concerned with the fact that libraries are not always taking advantage of all the data sources on authority control for names of persons that exist, especially of those that are non-

traditional library sources. Provided this, Ilik proposes the ideal solution of using VIVO, a Semantic Web discovery tool developed by Cornell University that connects researchers “across disciplines, institutions, geography and time” and uses a central VIVO Registry. More specifically, the author proposes the enhancement of name authority records by adding VIVO person URIs.

Closing the first day of the conference, two more papers were presented on the session “Authority Control Design and Classification,” chaired by Claudio Gnoli (Italy). Dagobert Soergel, University at Buffalo, and Denisa Popescu, World Bank Group, presented a paper entitled “Organization Authority Database Design with Classification Principles.” The authors provided a paper on the unified treatment of all authority data (including subject authority control/classification, places, events, persons, organizations, etc.), using as example the case of the World Bank Group (WBG) information system design. Following this, Ulf Schöneberg and Wolfram Sperber, Leibniz-Institut für Informationsinfrastruktur, zbMATH (Germany) presented a paper entitled “Machine-Learning Methods for Classification and Content Authority Control in mathematics.” This paper focused on the domain of mathematics, stating that “until now publications are the most important resource of mathematical knowledge and they are also the basis for knowledge management in mathematics.” Following the introduction, the paper covers a brief historical remark of mathematics and challenges for content analysis in mathematics. Then, the authors proceeded to describe the automatic tools and practices on authority control and classification at the bibliographic database of mathematical literature zbMATH, including the development of machine-based concepts and tools to create controlled vocabularies and to improve the Mathematics Subject Classification (MSC) scheme.

The second day of the conference began with the presentations of the forth session, entitled “Classifications in Subject Access Authority Control” chaired by Maria Inês Cordeiro. Marie Balíková, (Czech National Library) gave an invited talk on “Subject Authority Control Supported by Classification: The Case of National Library of the Czech Republic.” She reviewed the reasons to use universal tools in libraries and discussed some cases related to subject authority, such as the use of the Czech Subject Authority files (CZENAS) in the Conspectus categorization scheme, and the Interoperability in Memory Institutions (INTERMI) project. The second presentation was “Multilingual Subject Access and Classification-Based Browsing Through Authority Control: The Experience of the ETH-Bibliothek” by Jiri Pika and Milena Pika-Biolzi (Zurich, Switzerland). This paper continued other studies on subject searching

and classification browsing in OPAC interfaces (e.g., Casson, Fabbrizzi and Slavic 2011; Slavic 2006), while introducing the use of subject authority control and multilingual subject access in NEBIS (Netzwerk von Bibliotheken und Informationsstellen in der Schweiz). The OPAC interface NEBIS is said to include a classification scheme and a multilingual subject descriptor system, allowing users to search in English, German, and French, access to a subject authority record, and a semantic search expansion. Ana Vukadin (National and University Library in Zagreb, Croatia) provided a paper on the “Development of a Classification-Oriented Authority Control: The Experience of the National and University Library in Zagreb.” The author presented the experiences and challenges encountered during the development of the UDC authority database. Vukadin also discussed the advantages and characteristics of the UDC as an analytic-synthetic scheme also for authority control. In the view of the author, some of the tensions and problems present in these systems can be addressed with the distinction between syntagmatic and paradigmatic combinations of concepts. The papers also concluded that “the possibilities of verbal searching, already improved by linking UDC captions to the bibliographic records in the OPAC, can be significantly enhanced by connecting authorized classification data to subject heading authorities.”

The fifth session of the conference dealt with “Strategies and Innovation with Classification in Libraries” and was chaired by Rebecca Green (OCLC). Victoria Frâncu and Liviu-Iulian Dediu presented “TinREAD—An Integrative Solution for Subject Authority Control.” The authors began with a historical review of automatic uses of the UDC (with projects such as AUDACIOUS (Automatic Direct Access to Information with the Online UDC System), ETHICS (Eidgenössischen Technischen Hochschule Information Control System), BSRS (Basic Semantic Reference Structure), DOBIS/LIBIS, and NEBIS) and followed with the introduction of TinREAD (The Information Navigator for Readers), an integrated library system developed by IME Romania that allows the assignment of verbal index terms mapped to classification numbers in bibliographic records. This system supports subject authority control from two authority files: subject headings and UDC. The TinREAD system is said to take advantages of both intellectual indexing (from the UDC notations assigned to the documents in the past) and the automated indexing resulting from the integration of the UDC-based thesaurus. Olívia Pestana (University of Porto, Portugal) spoke on “Alignment in Medical Sciences: Towards Improvement of UDC.” The author revisited some of the arguments on the revision of the UDC Class “61 Medical sciences,” and conducted a comparative analysis with the National Library of Medicine (NLM)

Classification. The results of this paper point out some of the problems and difficulties of alignment vocabularies and interoperability. In the third paper of the session, Claudio Gnoli (Italy), Rodrigo De Santis (Brazil), and Laura Pusterla (Italy) presented a paper entitled “Commerce, See Also Rhetoric: Cross-Discipline Relationships as Authority Data for Enhanced Retrieval.” The starting point for the paper is Hugh of Saint Victor’s observation on interbranch relationships of subjects within a tree-like hierarchical classification system, also previously reported by Olson (2010). The authors then begin discussing cross-discipline relationships in library classifications (usually described as “see also” references), stating that in terms of authority control, cross-discipline relationships can be recorded in special fields of the classification reference database accounting for relationships of a class with others in a different hierarchy. This leads to the introduction of SciGator, an interface for exploring cross-discipline relationships in the DDC used at the The Science and Technology Library of the University of Pavia, Italy. The authors also study special kinds of relationships, namely existential dependence between classes, in which the extension of a class depends on another for its own existence (for example building and architecture, as architecture cannot exist if there are no buildings where it can be applied) and often belong to different hierarchies in a classification scheme. Finally, the authors also discuss the use of OWL for representing this existential dependence between classes in the Integrative Levels Classification (ILC). The final paper of the session was “Managing Classification in Libraries: A Methodological Outline for Evaluating Automatic Subject Indexing and Classification in Swedish Library Catalogues” by Koraljka Golub, Joacim Hansson, Dagobert Soergel, and Douglas Tudhope. The authors reported on a project that aims to evaluate automatic indexing to Swedish textual resources with the DDC (primarily) and the Swedish Subject Headings (SAO). The project considers the comparison of automatically assigned index terms with end-user assigned index terms and catalogers’ assigned index terms. Domain analysis (Hjørland and Albrechtsen 1995; Hjørland 2002) is used as a theoretical framework and qualitative complement to a post-study questionnaire. In the words of the authors, “domain analysis is used to take into consideration the social and disciplinary context of the documents used in the study and the catalogers, subject experts, and end-users.”

The sixth and final session of the conference was entitled “Issues and Opportunities for Classification Data,” chaired by Dagobert Soergel. A paper by Attila Piros (University of Debrecen, Hungary) reported on “Automatic Interpretation of Complex UDC Numbers: Towards Support for Library Systems.” Piros dealt with feasible ways of building

a UDC-specific XML schema for describing complex UDC numbers. A state-of-the-art method of parsing and converting UDC numbers is provided. The XML schema for UDC number descriptions, when finished, is also said to be made available online under a Creative Commons license. Andrea Scharnhorst, Richard P. Smiraglia, Christophe Guéret, and Alkim Almila Akdag Salah presented “Knowledge Maps for Libraries and Archives—Uses and Use Cases.” This invited talk discusses functions of visual explorations and knowledge maps in libraries and archives. The authors work with the concept of macroscopes (Börner 2010; 2015) as heuristic devices in research on the evolution of KOS. They analyze and compare the catalogs Worldcat, the library KU Leuven, PORBASE (the union catalog of the Portuguese libraries), the catalog of the BNP (National Library of Portugal), and BND Domínio Público (a dataset of the National Digital Library public domain files). Shenghui Wang and Rob Koopman (OCLC, Leiden, The Netherlands) gave an invited talk entitled “A Second Life for Authority Records.” The starting point for their paper is the benefits of authority control although they also point out that authority records are not used in the way they should be used, thus making it difficult to leverage the benefits of authority files. The authors analyze some cases of UDC usage in Worldcat and conclude the importance of context and the need of acknowledging bias.

Finally, the conference also included some posters and short papers that are also listed in the proceedings. Nuno Freire, Valentine Charles, and Antoine Isaac presented a poster on “Subject Information and Multilingualism in European Bibliographic Datasets” in which they relate the experiences with the integration of knowledge organization systems in cultural heritage data. Suzanne Barbalet presented a short paper entitled “Enhancing Subject Authority Control at the UK Data Archive: A Pilot Study Using UDC.” Barbalet investigated the suitability of the UDC as a flexible knowledge organization tool to strengthen the archive’s resources “in preparation for future metadata challenges,” consequence of “big data.” This paper also examines the adoption by the UK Data Archive of the Data Documentation Initiative (DDI) as an international metadata standard for social science data and the application of the UDC. The short paper “Towards the Creation of Integrated Authority Files in the Domain of Science and Technology: an Italian Use Case” by Elena Cardillo, Iryna Solodovnik, and Maria Taverniti reports on the creation of a local name authority file that integrates authority lists in the domain of science and technology for their application in the CNR S&TDL project. This paper addresses the problems of “locally developed authority lists mined from the resources managed in local repositories

and not aligned with external trustworthy sources” that do not allow sharing, re-use and interoperability of data. Andreas Ledl (University Library of Basel, Switzerland) reviewed the “The Basel Register of Thesauri, Ontologies & Classifications (BARTOC),” a bibliographic database of knowledge organization systems, developed by the University Library of Basel, Switzerland. One of the main features of BARTOC is that it provides a search interface for all types of KOS from any discipline. Darija Rozman (The National and University Library, Ljubljana, Slovenia) reported on “Experience with UDC Updates: the Slovenian Perspective.” The author reviewed some practices and the cataloging environment in Slovenia. Rozman introduced an authority list of UDC notations published in Slovenia called “UDC summary,” and related the update of this authority file and how occasionally UDC codes have been extended and edited following requests from librarians (usually following new releases of the UDC Master Reference File in Slovenian). Finally, Agnieszka Maria Kowalczyk, Łukasz Skonieczny, and Małgorzata Wornbard (Poland) authored a poster on “Visualization of a Library Collection Based on UDC: Research in the Warsaw University of Technology Main Library,” and Alenka Šaupel (Slovenia) presented a poster on “UDC as a Standardisation Method for Providing Titles of Documents.” Abstracts of all posters as well as all invited talks are included in the proceedings.

References

- Ascher, Marcia. 1994. *Ethnomathematics: A multicultural view of mathematical ideas*. Boca Raton: Chapman & Hall.
- Börner, Katy. 2010. *Atlas of Science: Visualizing What We Know*. Cambridge, Mass.: MIT Press.
- Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, Mass.: MIT Press.
- Casson, Emanuela, Andrea Fabbrizzi, and Aida Slavic. 2011. “Subject Search in Italian OPACS: An Opportunity in Waiting?” In *Subject Access: Preparing for the Future*, edited by Patrice Landry, Leda Bultrini, Edward T. O’Neill and Sandra K. Roe. Berlin: De Gruyter, 37-50.
- Fleck, Ludwik. 1935. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Basel: Schwabe. English ed.: Fleck, Ludwik. 1935. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.
- Hjørland, Birger. 2002. “Domain Analysis in Information Science: Eleven Approaches-Traditional as Well as Innovative.” *Journal of Documentation* 58: 422-62.
- Hjørland, Birger and Hanne Albrechtsen. 1995. “Towards a New Horizon in Information Science: Domain Analysis.” *Journal of the American Society for Information Science* 46: 400-25.

- Olson, Hope A. 2010. "Earthly Order and the Oneness of Mysticism: Hugh of Saint Victor and Medieval Classification of Wisdom." *Knowledge Organization* 37: 121-38.
- Otlet, Paul. 1934. *Traité de documentation*. Brussels: Editiones Mundaneum.
- Otlet, Paul. 1935. *Monde: Essai d'Universalisme*. Brussels: Editiones Mundaneum.
- Slavic, Aida. 2006. "The Level of Exploitation of Universal Decimal Classification in Library OPACS: A Pilot Study." *Vjesnik bibliotekara Hrvatske* 49, nos. 3-4: 155-82.
- Slavic, Aida and Maria Inês Cordeiro, eds. 2015. *Classification & Authority Control Expanding Resource Discovery: Proceedings of the International UDC Seminar 29-30 October 2015 Lisbon, Portugal Organized by UDC Consortium, The Hague*. Würzburg: Ergon Verlag.