



Open Research Online

Citation

Daga, Enrico; Scharnhorst, Andrea and Smiraglia, Richard P. (2023). Ordering the world, ordering our thinking, ordering interdisciplinary collaboration - on knowledge organization and ontology engineering. In: Mieg, Harald A. and Scharnhorst, Andrea eds. Transferability: Reflections on planning and knowledge organization. Bibliographische Informationen Der Deutschen Bibliothek (2022). Berlin, Germany: Wissenschaftlicher Verlag, pp. 133–153.

URL

<https://oro.open.ac.uk/98808/>

License

(CC-BY 4.0) Creative Commons: Attribution 4.0

<https://creativecommons.org/licenses/by/4.0/>

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

ENRICO DAGA, ANDREA SCHARNHORST & RICHARD P. SMIRAGLIA

Ordering the world, ordering our thinking, ordering interdisciplinary collaboration — on knowledge organization and ontology engineering

Kamingespräch, July 6, 2023

Abstract

The more we rely on machines to navigate the information oceans, the more important so-called structured data become, i.e. data that have been given a certain meaning. In this contribution we bring together two specialists in the field: Richard Smiraglia, currently executive director of a non-profit research institute The Institute for Knowledge Organization and Structure, and Enrico Daga, currently Senior Research Fellow at the Knowledge Media Institute of the Open University in the UK, to discuss with each other the history of Knowledge Organization in various fields, and important challenges in the present. The conversation was led by Andrea Scharnhorst, policy advisor at DANS, an institute of the Royal Netherlands Academy of Arts and Sciences.

Dr. Andrea Scharnhorst

Dutch Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Sciences (DANS-KNAW)

E-Mail: andrea.scharnhorst@dans.knaw.nl

H. A. Mieg & A. Scharnhorst (Hrsg.). (2024). *Transferability: Reflections on planning and knowledge organization*. *Wissenschaftsforschung Jahrbuch 2022*. Wissenschaftlicher Verlag Berlin.

Introduction

The more we rely on machines to navigate the information oceans, the more important so-called structured data become, i.e. data that have been given a certain meaning. Such data are used in search engines, for information retrieval, or to train machines to extract structures from unstructured text (and other data). The scientific terms to describe systems which help us order knowledge are various, and they are variously applied in different disciplines. Examples include controlled vocabularies, ontologies, classifications, keywords, metadata schema, data schemas, thesauri, and so on.

In this contribution we bring together two specialists in the field: Richard Smiraglia, professor emeritus in the field of library and information science, and currently executive director of a non-profit research institute, The Institute for Knowledge Organization and Structure, and Enrico Daga, who holds a PhD in artificial intelligence and is currently senior research fellow at the Knowledge Media Institute of the Open University in the UK, to discuss with each other the history of knowledge organization in various fields, and important challenges in the present.

The conversation was led by Dr. Andrea Scharnhorst, policy advisor at DANS, an institute of the Royal Netherlands Academy of Arts and Sciences. Andrea came into academia via physics, did her PhD in philosophy of science, and has since moved towards information sciences.

The conversation centered around three themes: (1) Your own path in academia; (2) Are knowledge organization systems (KOS) unique or universal? (3) The role of KOS for/in interdisciplinary work.

1. Knowledge organization and our own paths in academia

Scharnhorst: Let us devote some time to discuss with each other the history of the knowledge organization and how you experienced this in your own academic journey. Let me start: I am a senior fellow and policy adviser at DANS, which is an institute of the Royal Netherlands Academy of Arts and Sciences, and holds a data repository service. I came to this conversation via

physics and philosophy of science. So I'm neither trained in information science properly, nor in computer sciences.

When using the term *Knowledge Organization* I'm aware that I already set a certain normative tone, as the term *Knowledge Organization* is not neutral and common sense, but has been used after World War Two to mark a specialty in philosophy of science, library, and information science. So, I'm curious to hear about both your thoughts about *Knowledge Organization* in general. Could you both summarize your own path into Academia, and how you came across knowledge organization and what intrigues you?

Smiraglia: I will need several days! I'm a child and nephew of musicians. So I grew up playing at age four with my first accordion and clarinet, then started with the trombone, and eventually, around 1970, when I graduated from high school, I played flute and piano. I went to college here in Portland, Oregon, where I majored in flute playing. But, I also met Edith Kilbuck, who had a major influence on me; she was a brilliant harpsichordist. At that time, the harpsichord was just re-emerging, and I studied with her. And I bring that up because I have just acquired a harpsichord a few months ago, and this is my lifelong dream: here I am, 53 years later, playing the flute, piano, and harpsichord.

So I say all that because I also worked in libraries, in high school, universities, and in college. And when it became time to get a master's degree in music, I thought: Well, let's do something to make money, instead. So, I went to Indiana University, Bloomington to the then School of Library and Information Science, and worked in what is now called the William and Gayle Cook Music Library¹, one of the world's leading music libraries – I think the sixth largest. From there I went to work at the University of Illinois at Urbana-Champaign as cataloguer of printed music, and that's the world's fifth largest collection of music. So I had, in addition to my musical career, all these interactions with the *stuff* of music: I mean, I had millions of scores. And one of the problems we were dealing with that day was shifting old collections that were classified using the *Dewey Decimal Classification*, re-binding

1 <https://libraries.indiana.edu/music-library>

them, and putting them in a new building, using the Library of Congress *Classification* for music. And so this meant that if I had a work by Karlheinz Stockhausen² there would be one score. But, if I came to something like a Beethoven symphony, there would be 400 scores, and there would be different editions over the course of centuries, and I had to spread them all out on the table and arrange them. And I became very aware of the problem of: When is this work a symphony and when not? When is this the score and when is it not a score? What are all those variations? In information science, we say – thanks to the next 20 years of my work – ‘How do we disambiguate the cluster and provide context within it?’. So that became my PhD, you know: What is a *work*? Specifically, what is a musical work? How do we disambiguate a cluster once we've given it a name? So, at the time that was considered a problem in what librarians call descriptive cataloging. But, as you see it, it is essentially the problem of the gathering and disambiguating.

And that is the problem of what Ingetraut Dahlberg called: *Knowledge Organization*. How do we bring together things that are alike, that are like each other, and then within this cluster disambiguate for precise retrieval? My dissertation was about a study of what's called a derivative bibliographic relationship. Which is essentially to find the ‘work’ in quantitative terms so it could be studied empirically. We discovered that these clusters, like almost everything else in information, follow power laws. We discovered that what we're doing is essentially classifying with alphabetical terms. So, we call that an alphabetico-classified system or structure.

Later, I became a professor of information science. I taught knowledge organization, and research methods, and did this for 45 years at University of Illinois, Columbia University, Long Island University, and finally the University of Wisconsin, Milwaukee.

So that's my academic career. I worked on that problem of what is a *work*, and in 2001 I published a book called “The Nature of a Work”³, which pulled together all the science existing at the time. After that I began working

2 https://en.wikipedia.org/wiki/Karlheinz_Stockhausen German composer, known for his groundbreaking work in electronic music.

3 Richard P. Smiraglia (2002) *The Nature of 'A Work': Implications for the Organization of Knowledge*. Scarecrow Press.

on problems of epistemology and the phenomenology around this. So, to give an example, we say ‘chaconne’⁴. But how do we know what a chaconne is; and do you and I agree on what we think a chaconne is ... how that has to do with people's experience.

So that's basically, where I was when I met Andrea. I was invited to the Virtual Knowledge Studio⁵ when I went to work on the book ‘The Elements of Knowledge Organisation’⁶. At that time, Andrea was working on a project about Wikipedia⁷. They had downloaded the whole of it, I think, yes, and wanted to know what the underlying knowledge structure was. I just sat there and laughed. It was too big a problem, and I just kept saying: What is the research question? And the others kept saying, indeed, what is the question? But, in the end we did a lot of good work with that, by connecting to the Universal Decimal Classification (UDC)⁸, which is a faceted classification and it's almost a linguistic-style combination of work representations.

I taught the team⁹ that the point of classification is to remove language from the problem, so that we can work with pure ontology. That's why the UDC classification uses symbols, because the symbols simply represent the

4 Type of musical composition, see <https://en.wikipedia.org/wiki/Chaconne>.

5 A research institute at the Royal Netherlands Academy of Science, 2005–2010. Wouters, Paul, Anne Beaulieu, Andrea Scharnhorst, and Sally Wyatt (Eds.). (2013). *Virtual Knowledge. Experimenting in the Humanities and the Social Sciences*. Cambridge, Mass.: MIT Press.

6 Richard P. Smiraglia (2014) *The Elements of Knowledge Organization*. Springer.

7 Suchecki, Krzysztof, Almila Akdag Salah, Cheng Gao, and Andrea Scharnhorst. 2012. “Evolution of Wikipedia’s Category Structure.” *Advances in Complex Systems* 15 (supp01): 1250068–1. doi:10.1142/S0219525912500683. (Preprint Physics and Society; Digital Libraries. <http://arxiv.org/abs/1203.0788>. A. Akdag Salah, Cheng Gao, Andrea Scharnhorst, and Krzysztof Suchecki (2011). Design vs. Emergence: Visualisation of Knowledge Orders. MAP. Courtesy of The Knowledge Space Lab, Royal Netherlands Academy of Arts and Sciences. In “7th Iteration (2011): Science Maps as Visual Interfaces to Digital Libraries,” *Places & Spaces: Mapping Science*, edited by Katy Börner and Michael J. Stamer. <http://sci-maps.org>.

8 <https://udcc.org/index.php/site/page?view=about>

9 The project was called ‘Knowledge Lab’ and the team consisted of Krzysztof Suchecki, Almila Akdag, Cheng Gao, and Andrea Scharnhorst.

levels of ontology without the problems of language. So there you remove the phenomenological issue.

Around 2017/2018 I created this research institute, The Institute for Knowledge Organization and Structure. We received our Digging into Data grant¹⁰, which ran for four years. As soon as that was finished, I left the university and now I just work privately in this research field.

Scharnhorst: Thank you very much, Richard. The Wikipedia project was all about the evolution of knowledge organization systems, starting with the Wikipedia categories (and category pages) but also looking into the evolution of the UDC. I remember that I told my father-in-law about it: Jürgen Scharnhorst, who also worked on dictionaries as a language specialist. And he said: Andrea, this is nonsense. This will never work. It is usually done top down. When making a dictionary, we extract the terms, for which we really find evidence in mundane communication, or whatever the collection/domain is for which the dictionary is made. How could this be achieved by collective, unsupervised editing, when even the understanding of the English among the different editors will be different. But, in the *folksonomic* nature of the Wikipedia categories (English Wikipedia), we found structures and change. My own interest goes back to my physics background, where I was studying in the field of statistical physics and thermodynamics of nonlinear systems. And that's all about dynamic processes, so that's all about change. I was very intrigued to find such changes in classifications, something which I naively thought is kind of God-given and stable. I'm still kind of puzzled by this: How volatile knowledge organization systems are in practice, although they are supposed to be stable, standards and give us guidance and be our reference systems – like the latitude/longitude system does to measure Earth.

10 Digging into Data was a funding stream of the Trans-Atlantic Platform for the Social Sciences and Humanities, which enabled collaboration between North American and European scholars. <https://www.diggingintodata.org/>, <https://diggingintodata.org/awards/2016/project/digging-knowledge-graph>, <https://web.archive.org/save/https://diggingintodata.org/awards/2016/project/digging-knowledge-graph>.

KOS classifications *should* help us to navigate the seas of knowledge as the former help us to navigate Earth. That's a good bridge to Enrico I think.

Daga: Richard, I certainly can understand how it could take days to talk about it all, and I really hope in the future there will be opportunities to hear more from you. I share with both of you a kind of nonlinear trajectory. I mean, when I first engaged with academia in my undergraduate studies, I was studying performing arts and the history of theater and music. Actually, my bachelor thesis was on the use of music in contemporary performing arts. And so what I was looking into is how you can take a structuralist kind of approach to entangling the texture of the performance. And what was the role that music was playing in the rhetorics and in the performance. I also share with you, Richard, that I was playing music and acting. I was doing a lot of art. But at some point, I realized I needed a job. At that time the Web (world-wide-web) was rising, and there was a lot of need for web developers. I also realized that in the 10 years I was working as an artist, I had developed two or three digital libraries for smaller organizations, as a kind of aside job, and that was really very enjoyable. I 'jumped' to computer sciences when I became engaged in web technologies and particularly semantic web technologies. This happened by chance. I was working as a web developer, and that happened to be in a research organization, and then I just got engaged with projects on semantic web technologies. So, my background is really from that niche area of ontology engineering. I was developing tools for ontology engineering. So, the terminology is one of the first things I heard, let's say in my academic career. Then, I moved to the UK, did a PhD, and so forth.

But what really attracts me, of all these stories we are telling, is that in the end – whether we look into the software engineering part (how we support humans with computers); whether we are looking into describing human things with formal methods; whether we are looking into communicating with machines at different stages – the key is always, and this is technology independent: How do we organize and structure knowledge? This happens from 'How do we organize and structure a website?' to 'How do we organize a structural digital library, a database? How do we organize and structure the information that we can use to understand how complex systems work and

what users do with these systems? That's what I'm particularly looking at in this phase of my career, is: How we can describe the journeys of the data in complex systems, particularly in systems that are user- and data-intensive. So, how we can describe them as, and how ontology, engineering, or knowledge organizational seasons can help in mastering this complexity. I just sketched a few notes while listening to both of you, and think there is a big embarrassment concerning terminology. I find this very often in computer sciences, which is my kind of reference academic area, and even more specifically ontology engineering, and the semantic web. I'm used to reading and reviewing computer sciences papers. I write papers for computer sciences on venues. I'm always, as a humanist, very irritated by the lack of consistency in terminologies and also by the lack of interest the typical computer scientist has in terminology. However, I think the devil is in the details.

We should actually try to use the terms consistently, and to question ourselves about why we are using that term and not another; why we are describing this as a knowledge graph and not as an ontology; why, we are describing this as a taxonomy instead of a database or a concept scheme, or a term list and so forth. What is a keyword and what is a label? This is a useful place to reflect about what are the terms that we use.

Also because language has a key role in ontologies. Here, Richard, I would like to connect to what you were saying before. You pointed out that we want to have classification systems because language is ambiguous, so we want to get rid of language. The problem is that we need language to understand those classification systems and to interact with them. So, the word we choose becomes important because it's what we attach to our own internal kind of machinery [our brain]. And we use those words for making sense of the knowledge organization systems. I remember that I had a debate a few months ago with a colleague. They built a taxonomy of topics for computer sciences, and they derived this from papers. They built this kind of topic-subtopic, a very rich and interesting ontology. So, ultimately, you can browse the scientific literature along topics and subtopics. I remember that there was one topic on semantic web related to a subtopic of Linked Data, and one related to a subtopic of web of data, something like that. That thing kind of puzzled me. So what is the semantics of this topic-subtopic relationship? The

answer was that the relationships they found, the semantics was just the result of the algorithm. So essentially, the algorithm was clustering topics and articles, it was considering relationships to subtopics in an extensional way. So, there are articles that typically have the topic X, but are also about this other topic Y. And if there are more articles having topic X, this topic is supposed to be more general. That was the kind of trick. But this has nothing to do with the definition of the concepts. This has more to do with how the concepts are used, or even more, in the assumptions that we make when we take these concepts and we put them together. And sometimes, how we use them together is not even expressed in a knowledge organization system. So people put things together, but it's very hard to know why they made those choices, for example.

So one of the first things we started 10 years ago, when I was more involved in the ontology engineering part, was to look for ontology design patterns¹¹. This idea was borrowed from software engineering, from object-oriented software engineering, and it was brought to ontology engineering by Aldo Gangemi¹². You have design patterns, and those are well-known, good-quality design solutions that you can use. And the idea was that ontologies are the same. They are design patterns, and we can reuse those to make sure that ontologies are well designed. But, in the end, we don't know why an ontology pattern is good or not. Yes, there has been literature on the quality of ontologies. But, at the end, we just agree that it is good, or some people agree that it is good, or the reference community or the domain expert says it's good. And here I finish. I don't know whether this is useful.

11 Presutti, Valentina, Enrico Daga, Aldo Gangemi, and Eva Blomqvist. "eX-treme design with content ontology design patterns." In Proc. Workshop on Ontology Patterns, pp. 83-97. 2009.

12 Gangemi, Aldo. "Ontology design patterns for semantic web content." In International semantic web conference, pp. 262-276. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

2. Knowledge organization and interdisciplinary work

Scharnhorst: I find it very inspirational. Because you also touch a string of my ‘scientometric’ heart. In scientometrics, where people look at quantitative traces of scientific communication, namely, article publications, and how you can cluster them. They find clusters, give them names. And there's a huge debate about how to do that properly.¹³ And computer scientists (and physicists) are very much intrigued by this kind of data mining. Those papers and references form complex networks, you see. But those working on better clustering algorithms are sometimes also a bit *laissez-faire* in the way of attaching labels to those clusters. Richard, how does this resonate with your experiences? What would you like to say in response?

Smiraglia: Well, the thing that I noticed most clearly was, I think we call this *siloining*. I have been very much aware of that, being an information scientist who works with computer scientists on occasion. Those two communities do the same thing, but call it something different, and don't ever talk to each other. And this really is not a new problem. I had the good fortune of having two PhD advisors because my first advisor, Arlene Taylor¹⁴, she was renowned for being one of the first to bring empirical methods to the problems of the distribution of names in large library systems. But then she left the University of Chicago. And so Don Swanson¹⁵ took over as my advisor, and he was very much aware of this problem [of silos]. This also had to do with him having a specific medical condition that different medical communities couldn't treat, even though two of them were both studying it simultaneously, but would not talk with each other. One of his most famous papers is called

13 Gläser, J, Glänzel, W & Scharnhorst, A (2017). Same data – different results? Towards a comparative approach to the identification of thematic structures in science Special issue. *Scientometrics*, vol. 111, no. 2.

14 Taylor, Arlene G. "A Five-year Projection of the Impact of the Rules for Form of Heading in the Anglo-American Cataloguing Rules, Upon Selected Academic Library Catalogs." PhD diss., University of North Carolina at Chapel Hill, 1981.

15 https://en.wikipedia.org/wiki/Don_R._Swanson

‘Undiscovered public knowledge’¹⁶. And it's all about these things that are actually known and common, but nobody knows that they are known and common.

So, one of the most recent examples of that for me was during our ‘Digging into the Knowledge Graph’ project. One of our computer science people from the VU (Vrije Universiteit Amsterdam) started to give a long talk about versioning of things. And I had to leave the room, because I've been writing about that for 40 years. But he had no idea that we ever thought of it. He thought it was a brand-new idea. Oh, my goodness. And what's the other one? The use of terminology in computer science is often puzzling to me. ‘Semantic shift’ – that's it. Yeah, it has nothing to do with anything semantic, right? I had the group in Amsterdam explaining it to me, you know, and I was following what they were doing. But then I said: Well, which terms shift? And they answered: “Oh, it doesn't have anything to do with specific terms.” Well, then that's not semantics.

I once worked with the CIDOC-CRM¹⁷ expert group; I did so for two decades. And there was that famous ontologist (Nicola Guarino¹⁸), and he just kept standing up, waving, and saying ‘No, no, no, that's not what that is’. So I think it's a really serious problem – the silos that academic disciplines form.

I also worked together with Rick Szostak. He's a Canadian economist and a professor for knowledge organization as well, and also works in our institute [IKOS]. He has created an interdisciplinary classification that's phenomenon-based. A classification that does not use disciplines or disciplinary structures, but that relates phenomena empirically. So there is a movement, trying to escape the siloing.

Scharnhorst: I also have a background in science and technology studies, and in philosophy of science. In those research fields, people have studied interdisciplinary working extensively. What made (still does) a deep impression on

16 Don Swanson (1986) Undiscovered public knowledge. *Library Quarterly* 56(2): 103–118

17 https://en.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model

18 https://en.wikipedia.org/wiki/Nicola_Guarino

me was a talk that Peter Galison gave at the WZB¹⁹ in the 1990s. He introduced the vision of the concept of a *trading zone*²⁰. He spoke about the fact that if people come together in an arena or marketplace, from very different areas, and they have different languages, they have to find a way to communicate with each other. To be able to do this, everybody needs to loosen up a bit, and finally they ‘invent’ a kind of pidgin language²¹, a mixed language that isn't as pure any more as it was in their home communities. They kind of trade concepts, and terms to describe them. They agree about the use of terms and their meaning, and this way they create their own community and language. I remember that the GDR philosopher and novelist John Erpenbeck²² also spoke about this. John said: When innovation is born, an invention or a new idea always starts in the head of one person. We don't invent collectively, we don't think together. But, to be able to fly and conquer more than one brain, this new idea needs to be communicated.

For that it needs a sender and receiver, right? Somebody needs to understand your idea. And that's where a shared language becomes important. Throughout my career, most of which was interdisciplinary, I observed that everybody is kind of solving this problem by *doing*, based on tacit knowledge and experience. There is no guidance. There is very little teaching, very little best practice. When people become aware of terminological problems they often start to do glossaries, and then they get tired of doing glossary, because it's also a lot of work to do one properly. So, they do this as a preparation, and in the meantime they figure out their ‘pidgin’ and leave the documentation. One could also say, it is a kind of illness that accompanies the immense growth of knowledge and the science system. What I would really be interested in, is to see if one could formalize the ‘trading zone formation’ more. I hear this in what Enrico said: We wanted to build a *good* ontology and repeat a good way. You search for a reference system and try to formulate this, and I hear this need also in Richard's stories.

19 <https://www.wzb.eu/en>

20 Described also in his book. Peter Galison (1997) *Image and Logic*. University of Chicago Press

21 <https://en.wikipedia.org/wiki/Pidgin>

22 https://de.wikipedia.org/wiki/John_Erpenbeck

3. Are KOS unique or universal?

Scharnhorst: On the other side, there might also be a logic in ‘starting from scratch again’. I think there is something inherent to any new research collaboration. It's not that people are just stupid and not willing to learn from each other. Let me give you an example: In the Polifonia project²³, at the consortium meeting in Bologna, Albert Meroño Penuela made a remark that was an eye-opener for me: When thinking about how to build an ontology for this project, at some point we need to start from scratch. Yes, there are other ontologies about musical objects around, but if we follow them too much, too closely, we cannot think outside the box, think openly any more. So, we need to ‘invert’ the process, coming up with our project-specific ontology and then see how much this overlaps with already existing ones.

So that brings me to the question: Can or should ontologies be universal, or can they not? Do we have to live with the dichotomy of universal versus specific? Or can we define functions, situations, locations, points in time in research processes where the more universal nature is needed, or where the more unique nature of ontologies, or Knowledge Organization Systems in general, is needed?

Daga: It is a lot to react to. I will come back to answer your question about universal vs unique at the end. I'm intrigued by your remark, Richard – about a classification system that can work with features and not with concepts, which works in a kind of inductive way. This might be a way to kind of eventually overcome the problem of fixing terminologies. Because the fact that different communities talk about the same things in different ways leads to situations where: one puts people together, and some people get fixated on the terminology. They can't get out of those words or terms. I think there is a big problem. Not that we actually can really solve this, because it's inherent in the system, and it is really about the tools that we use. It's this kind of

23 <https://polifonia-project.eu/> Permalink: <https://web.archive.org/web/2023072011153/https://polifonia-project.eu/>

semiotic problem. So when we point to something – let’s say we point to the Moon, and we give it a name: Is this name for the finger-pointing or the Moon? So, if it is the finger we need to name, then the idea of a Pidgin language might work. Because the other person knows already in her/his head – assuming that in some way the other person knows already in his/her head what we're talking about, and we just need to reach it, find the quickest way to reach what they're thinking of. But the problem is – especially for knowledge and ontology engineers – they curate ‘fingers’. So they create names; and their organization – making it a kind of standard – and they want to forget what it stands for. They make use of these systems on behalf of the real thing, as a kind of simulacra. So those labels become important, and communities get attached to those labels. And if you change those labels, people don't understand any more. It is the same when we refer to professional language, and it can go down to everyday workplaces.

I remember last year I was involved in this working group around the university [Open University] to systematize the terminology that the various departments use, from recruitment to new students, to finance, budgeting, and people that work with claims or policies. There was an office that only specialized in retainment. That means you need to avoid students leaving before completing their studies, which is a big issue at the Open University because we have 200,000 students all over the world. And we are a distance learning university. So there are a lot of people that just drop out after a while. So the problem is that the same terms really mean different things in different departments for very strong reasons. For example: The accountancy rules that require – that impose on us – the use of one term rather than another, even though that other term would mean something much more flexible and broad. So when you build statistics and you integrate data, then it's a really hard thing to do. Because sometimes you get numbers that don't match, and you don't know why. But the reason, is because the concepts don't match and the terminology doesn't match, and people don't talk to each other because they just do different jobs. So these are very fundamental problems of knowledge organization systems.

Now, this is why ontology engineering started with the idea that an ontology was based on agreement. So, a certain community decides to structure

and organize its own discourse, agrees on an ontology, and describes this piece of word in this way, because that makes sense to them. So, universality and re-use are big problems, and these have been studied in ontology engineering. At least, that's the area I know. I'm sure that there's much more outside of ontology engineering, dealing with very similar issues. But in my area, there were heated debates about whether there should have been one ontology. Some people thought we should agree on one – THE ontology: Put all the big brains of the world all together, and agree on the ontology. I remember Nicola Guarino²⁴, who is one of the main founders of the ontology engineering discipline and developed the OntoClean methodology. You see, it is in the name: 'OntoClean'. But other people disagreed, and said it's ok if people see things differently.

The thing that we need to reuse – and this is what we need to find out – is: What are the good ways of doing these things? Because ontologies are artifacts, and they have a life cycle like any other artifact. You can't think that you can just make one big ontology that covers everything.

Although, you can abstract in many different ways. And then ontology engineering solves this problem by having foundational ontologies, or top ontologies. So, you work on very abstract patterns that nobody will use because nobody cares to distinguish perdurance and endurance. That's because you don't need that when you need to organize books or employees, or whatever in your system needs to be organized.

Let's focus on these abstractions, because those will necessarily be reusable. So if you go to the root of these schemas, this must certainly be reusable. I think that here we find great results. I mean, DOLCE²⁵ is fantastic. You look at DOCLE and you really see an effort on systematizing the foundations of our conceptual thinking by digging into philosophy, description logics, all of the kind of big thinkers. Trying to encode this into an ontology. But the problem is that when we need to do an ontology, we don't need to

24 https://en.wikipedia.org/wiki/Nicola_Guarino

25 'Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is a foundational ontology designed in 2002 in the context of the WonderWeb EU project' Quotation from https://en.wikipedia.org/wiki/Upper_ontology#DOLCE, and <http://www.loa.istc.cnr.it/dolce/overview.html>

say that a person is a ‘perdurant’²⁶; we just need to say that it’s a person. And so then the pidgin language layer enters and saves us. Okay, because otherwise things get too complex for no reason whatsoever. Nobody cares in the end.

So there is a way or a question, which is my favor, and which makes all those complex discussions very shallow. And this is the big question: What is good enough? You cannot decide what’s good enough if you don’t have a task. So every time that we talk about ontology engineering, and discuss ‘Let’s do an ontology’, my question is always: What is the task? So what is it the ontology needs to support? It is not ‘we need to describe the domain’, because – and now I contradict myself – nobody needs to describe the domain. What we need is to describe a domain because this helps us in understanding it. And that’s where Albert’s point makes sense.

So, Albert’s point makes sense. But, not in the way: ‘we need to start to think anew, because otherwise we can’t be creative’ [which was Andrea’s summary of it]. I don’t believe that. I believe in the fact that we need to start anew because otherwise we don’t think; And then this process of building an ontology becomes useless to us. What we want is that the process of ontology building becomes useful to us, because in the end we want to understand something more about this domain. And here then the creative process is necessary. Otherwise, it is only painful. If you just have to try to understand what others were saying and put together something, that just does not work.

Scharnhorst: Another reply, another short round, and then we should kind of wrap it up, Richard.

Smiraglia: In the beginning, Andrea, you asked about the term ‘Knowledge Organization’, and I think that I have to remind us that that term came from Ingetraut Dahlberg, and that she was not in information science. She was a documentalist, and when she created the term Knowledge Organization she translated it from the German word ‘Wissensorganisation’; and you have to

26 perdurant in (ontology) A happening; an entity that only exists partially at any given point in time (see <https://en.wiktionary.org/wiki/perdurant>)

read her writing, to understand why she chose this term, also to distinguish herself from mathematicians because she was trying to work with the concept of pure, universal ontology. Her society, which is now the International Society for Knowledge Organization, was intended to be a division of the ‘science of science’. She saw this as a substrate of all science. And so, there is that tension in the field, 50 years later. This is a tension between the poles in universal classification or ontology, something that describes everything in one language and, on the other side, what's now called the postmodern approach, which is to say that each discourse community – which can be also be just one person – has its own ontological structure. And what we have to look for is a translation device – a crosswalk. So, those tensions still exist in the field.

Enrico, when you were saying – in connection with the purpose of ontology – that you will always ask ‘what is the task?’ I started to laugh, because it's exactly the opposite of what we do in my institute. Our purpose is to ask ‘What is the underlying knowledge structure of the domain under study?’ And we try to seek that by sort of an ethnographic method. What are these people [in a domain] saying, and what do they mean when they say it, and how do they use it, and to describe their domain. But of course both things are necessary, right?

And then a colloquial example: Just in the news, as you know, there are many shooting events in the United States. Lately, they've started to say ‘it was random’. I always shake my head when the news says ‘it was random’, because I remember my professor Abraham Bookstein drumming into my head: “random” means with equal, known, and non-zero probability. What they mean is ‘chance’. It wasn't random if someone chose to do it. I bring this up because it's the problem of colloquial language²⁷ differing from scientific language.

A musical example is this problem with *melody* that my institute is about to study. You know, users want to classify music by melody, but musicologists don't know what a melody is; or, rather: We can't agree what a melody

27 ‘colloquial language, everyday language, or general parlance – is the linguistic style used for casual (informal) communication’ (<https://en.wikipedia.org/wiki/Colloquialism>)

is. I mean, we have an idea what it is: it's that tune. But we can't nail down what a melody is, to be able to classify it. So the question is not to have a classification that is universal or domain-specific, but rather it's culturalism. That's what it really is: Where the two different understandings are not shared. To determine where the intersection is, that's an interesting problem to go forward.

Scharnhorst: Thank you very much, Richard. Maybe we can have a last round, and share what we three do when thrown into a situation (as we always are in our kind of daily practice) of engaging with a new community, when we are thrown into a new project, or into new 'clinics'. Richard does these beautiful workshops, called clinics. So how would you get the dialogue going? What would be your favorite 'recipe', to tackle the problem of people not understanding each other? How would you do this? What would you do?

I can tell you what I do. I try to understand the academic backgrounds of the people involved, to which *academic tribe* they belong, which epistemic formation they had in their career. Am I talking to someone who came into academia via mathematics, physics, chemistry, social sciences, philosophy, or art? Yes, it becomes increasingly complex to 'science-locate' a person (as Katy Börner would call that) as academic curricula themselves become a mixture. The big disciplines are broken up and recombined. But often, for me, that helps me to at least have a hypothesis about the 'epistemic mindset' of that person, and with which method a scholar is most familiar: quantitative, qualitative, text, statistics...

Engineers, I found, usually really wanted to build something; So this is different from a more theoretical approach. Whatever their discipline – natural or social sciences or humanities – at the end, everything is connected. But your own research practice determines your view, and therefore also your behavior in interdisciplinary work. So I try to encourage people to share information about their own academic trajectory; And I try to share my own experiences, making explicit what determines my worldviews. Because, at the end, you want to achieve something in such an interdisciplinary encounter or maybe you have to write something together, such as a project deliverable.

So, Enrico and Richard. Do you have the secret recipe?

Daga: In the first round, I would ask: Why we are here? What is the task? And a first task could also be to have a formal understanding of the domain of these two groups that might be in such an encounter. That's absolutely fair, just to clear the air. But the reason why I emphasize the need to determine the task is because, often, knowledge engineers get into such a collaboration by following the approach to build a knowledge organization system top-down. They try to understand the discourse in a certain domain, build a system, and then pretend that this system can solve the problem at hand. Or they recommend a service or product. And that is when I kind of get nervous. Because your concern – as a researcher, as a knowledge engineer, which is ‘I want to understand the domain, I want to do the *right* ontology, I want to use all my tools’ – becomes a burden; baggage that doesn't help the final goal. Okay, so this is why I put my software engineering and systems engineering head on and say: Look, why we are here? If we are here to help those guys, we need a knowledge organization system, but that system has to feed the task. If we are here just to do something else, then we do something else, and then that might be exciting as well. It's probably more exciting than solving the problem of the domain.

And I think that the problem of interoperability and interdisciplinarity is anthropological. In the sense that if you want to understand how to bring together different communities, you need to understand language and norms. And these are different for different communities.

So if I need to write a computer science paper, it must have a certain, specific language; and often, this language is also very specific to the sub-community. So I need to pay attention to what language I use. I mean, we look at a “call for papers” to decide what types of words we need to use, to write about a subject [for that community]. So, we do the work. Then we ask ourselves, where we submit the paper for publication? There, there, there or there? Oh, there is this one. Okay, what are the topics? Oh, yeah, we can express our work in a particular way, and we use relevant words. We can take the same work, then do another step such as another piece of the work; then choose another venue, with different words, different problems. Oh, yeah,

we can use these words. And we move on with our research. Again, all those words are *fingers* they are not the Moon.

And the assumption is that we know what this 'Moon' is about. So we know that I'm working on ontology engineering methodology. That's actually a place where terms probably remain the same. But when we work on recommender systems or user models, what is important is language and the rules of the game. The rules are different between communities. So what are the ingredients that the paper needs to have, etc? What type of discourse is considered to be scientific?

And this changes dramatically. Even within computer science, it's completely different between the area of computer interaction to that of databases. Okay, I don't know if I answered anything

Scharnhorst: Interesting. Richard would you like to give it a go? So, the initial question was: If you are thrown into an interdisciplinary setting, what's your recipe to get it going, to look in it, to organize yourself in the group, to organize the knowledge exchange in the group?

Smiraglia: I think I said it before. I realized now why you brought us together. I'm beginning to see it. Enrico was starting with: What is the task? My version of the starting point is: What is the research question? And that's where I always go, and that comes from my training in Chicago. Once we find the question, then we know how to begin.

So what I was thinking about, though. Well, again, back to the CIDOC-CRM, which is a cultural heritage ontology, and it's designed primarily for use in musea and archives.

But I found that the meetings are excruciating. Because, there will be 30 or 40 scholars together in the room, all around one big table, and they're all from different disciplines. And so they all have different vocabulary. And whatever the problem is, there will be 15 or 20 different solutions, because everyone places their own one on the table. And I was describing this once to my colleague, Paul Wouters, and he sent me to a book by Collins²⁸. The

28 Randall Collins. 1979. *The Sociology of Philosophies*. Belknap Press.

main idea is that you have a limited number of schools of thought [in the history of philosophy] and they grow and grow until they fall apart. And then they come together, and compromise around a truth or a hypothesis that they can live with. And this is what happens in the CIDOC-CRM meetings. At the beginning, there will be 30 answers to the question, and then over the course of five days it narrows down, and we get to the gist of the true research question, or you could say: the task. What is it that we try to describe here?

And this is the method we apply in our own institute. We begin with a meta-analysis, we bring a group of scholars together – they're not all from the same discipline, they are offering their own data to the conversation – and we lay out all the questions and get the big confusion. And then, over time, we narrow it down to the critical question. This is what phenomenon-based classification does. What is the thing? What is its nature? How do we describe it? So that's, I think, the intersection with interdisciplinarity.

Scharnhorst: Thank you both very much, Richard and Enrico, for your contributions.