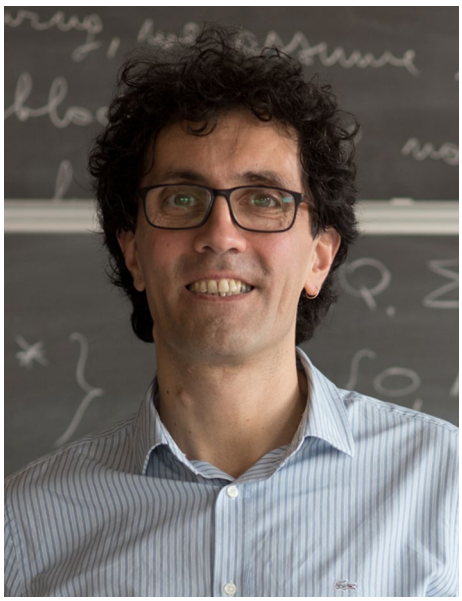
KI - Künstliche Intelligenz
https://doi.org/10.1007/s13218-020-00691-6

INTERVIEW

Interview with Diego Calvanese

Diego Calvanese[1,2] · **Mantas Šimkus**[3]

Received: 10 October 2020 / Accepted: 12 October 2020



Diego Calvanese is a full professor at the Research Center for Knowledge and Data (KRDB), Free University of Bozen-Bolzano (Italy) and Wallenberg Guest Professor in Artificial Intelligence for Data Management at Umeå University (Sweden). He studied Electrical Engineering at Sapienza University in Rome, worked as a software developer for one and a half years in Innsbruck (Austria), and then completed his PhD in Rome. In 2003 he moved to Bolzano and built up an internationally renown research group working on knowledge representation and data management.

✉ Mantas Šimkus
simkus@dbai.tuwien.ac.at

Diego Calvanese
calvanese@inf.unibz.it

[1] Free University of Bozen-Bolzano, Bolzano, Italy

[2] Umeå University, Umeå, Sweden

[3] TU Wien, Vienna, Austria

1 Interview

When did you become interested in science, and specifically in Computer Science? What was your first experience with computers?

As far back as I can remember, I have always been interested in technology and in understanding how things work. Science fiction movies were the ones I liked most, because of the possible future technologies they presented. And computers and programming fascinated me from early on. When I went to middle school, laptops did not exist, and the early PCs were out of reach for me, so I resorted to programmable calculators from HP, first the HP-33C, then the HP-41C, and later the HP-28S. My information source was the monthly CHIP magazine, which presented news and info about pocket calculators, programming, and later home computing. And I was also an avid reader of Scientific American. I still have a subscription to the Italian version, while initially I read the original English version, as a means to practice English.

When did you become interested in logic? Specifically, Description Logics (DLs) play an important role in your research. What was your first contact with DLs?

The interest in logic came later, during my studies in Electrical Engineering in Rome, and was somehow triggered when I read *Gödel, Escher, Bach: an Eternal Golden Braid* by Douglas Hofstadter. The book blew me away, and I immediately wanted to know more about Gödel's Incompleteness Theorem and understand logic better. So I attended a course on Logic by Giacomo Jacopini and one on abstract computing machines by Corrado Böhm, both offered at the Faculty of Mathematics. I discovered only later that the two were quite famous, for a fundamental result they had obtained in the 1960's on the theory of computation. For the course by Böhm, I worked on a project in which I implemented an interpreter for a language based on combinatorial operators. But I lacked background on some basic Computer Science subjects, and I remember that only later, when I took a course on compiler technology, I realized that for that

Published online: 27 October 2020

project I had "discovered" recursive descent parsing, which I had needed to implement the parser for my interpreter. I learned about DLs even later, for the final thesis of my engineering studies, which I did under the supervision of Maurizio Lenzerini.

After my studies I moved to Austria and started working in software development in a small company, where I did both low-level programming of an embedded system, and high-level development of the control and interface layer of a gas analysis device. But I was missing the kind of work that I had done during my MSc thesis and the "logic related stuff" that I had studied and that I had loved. So, when Maurizio invited me to try to compete for a PhD position in Rome, I gladly gave up my job as a software developer and started working again in a research environment, also getting back to DLs. It was a big jump back in terms of salary, but I had much more fun and could work and study what I liked. I never regretted that choice.

Your research is at the intersection of Database Theory and the field of Knowledge Representation and Reasoning (KR&R). What are the reasons for that?

I was introduced to this area by Maurizio, who himself has a background in Databases, but started working on logics for knowledge representation when visiting Hector Levesque in Toronto. For my MSc thesis with him I studied a combination of DLs with Logic Programs, and later during my PhD, we explored the relationship between DLs and data modeling formalisms used in Databases, deepening earlier work that Maurizio had done. So, the connection between logics for KR&R and Database Theory was there right from the beginning when I started doing research. And it was something that intrigued me, since I had always loved Logic, but I also liked to see how it could be applied to the more practical problem of managing data.

Continuing to work on the combination of these two areas was a natural development, and actually it is what I am doing up to now. In some sense we are addressing a setting where we are forced to combine the difficulties that come from both areas. This is the "negative" side, if I may say so: on the one hand we have to deal with the rich structure of knowledge and incomplete information that we find in KR; and on the other hand we have to deal with the huge amounts of data that are typical of the database setting. The "positive" side is that we can also resort to the wealth of techniques developed in both areas: from KR we get, e.g., a proper way of representing rich structures and defining semantics by relying on formal logic, and forms of knowledge compilation; to complement, from Databases we get, e.g., powerful query execution strategies and techniques for query optimization. But the connection between the two areas goes well beyond that, and we have now an increased cross-fertilization between them. It is perhaps interesting to observe that in 2017 the best paper award for PODS, the premium Database Theory conference, went to well-known KR researchers for a paper on ontology-mediated query answering, a fundamental problem studied in KR; and conversely, in 2018 the best paper award for KR, the premium KR&R conference, went to well-known database theory researchers for a paper on incomplete information in relational databases. This is a good sign that the connection between the two communities is growing stronger.

Who are the people who inspired you most?

I owe a lot to Maurizio, since he brought me to research, and since then has been a constant inspiration for what are important problems to look into and research directions worth following. And then Moshe Vardi, who is an incredible person. I am honoured to have worked with him. Moshe has both a very broad vision of many areas of Computer Science—he gave key contributions to Logic, Database Theory, Automata Theory, and Verification, among others—and he also has a very deep technical knowledge and ability to analyze and solve difficult problems. And finally Giuseppe De Giacomo, who more than a colleague with whom I started this adventure in research almost 30 years ago, is a very good friend. There are many other colleagues and friends with whom I have worked over the years and from whom I learned a lot. Actually, I am very happy to be part of such a nice research community.

How would you describe the importance of your research?

Let me mention two reasons why I consider the research work at the intersection of Databases and KR&R that we have been doing with my friends and collaborators important. First, all our work is grounded in solid theoretical foundations. Therefore many of the results we obtain are not ad-hoc and do not depend on specific technologies that might change or fade away, but stay valid and continue to be relevant over time. Second, we have always been guided by practical problems and we have tried to develop solutions that address real needs that people have when it comes to managing complex data and knowledge. This latter point became even more important and prominent in recent years.

An old yet still central challenge in data management is the data integration problem. Why is Ontology-based Data Access (OBDA) so appealing in addressing it?

It is indeed an old problem, however many of the issues that make data integration difficult have not been solved yet. On the contrary, with the incredible growth of the amount and complexity of data that we have been witnessing over the past years, the problems have become even more difficult and solutions are not readily available. OBDA helps in addressing one of the key issues, namely dealing with the semantics of data, and making good use

of the domain knowledge that is available and that is an important asset and of great value. Such knowledge needs to be encoded, in a declarative manner, in ontology and mappings, and then it makes the content of the underlying data sources understandable and therefore easier to integrate. However, it is still difficult to get this message through where it would be relevant, i.e., with organizations and companies who face serious data management and data integration issues. One reason is also that plug and play commercial solutions for OBDA are not yet readily available, and there is still a lot of manual work involved in setting up a data integration system that is based on this paradigm, e.g., to define the ontology and mappings. This is something on which the KR&R community needs to work, by defining proper methodologies and supporting tools, and by providing means to partially automate the design process. I believe that in complex scenarios it will not be possible to completely get rid of manual intervention and reach full automation. But between a fully automated solution, and the current situation, where the designer of an OBDA integration system has to start from a blank page and manually define possibly hundreds or thousands of mappings, there could be intermediate solutions that are viable, where the designer can rely on supporting technologies that greatly simplify their work.

It appears that over the last 15 years KR&R research has been especially active at developing and studying properties over various DLs. Do you agree with this assessment? If yes, could you explain the reasons for this activity?

I think that the statement is accurate. The DL community has moved from doing work that is of interest mostly to a niche, to addressing real world problems with technically advanced solutions that are grounded in solid foundations, as I have already mentioned before. This has sparked further research, since solutions that in theory seem to work well, face limits, e.g., in terms of performance, when applied to real-world sized problems, and therefore techniques need to be refined and extended. A good example is the work on non-uniform complexity in OBDA and ontology-mediated query answering, which can be seen as a way to refine the analysis so as to discover tractable cases also in those settings where the general problem would be intractable. Moreover, and this is a very recent development, people who have developed or are using AI technologies based on machine and deep learning are recognizing that in many respects they have reached a limit where a pure learning based approach cannot go further, and where one would need to bring in and use structured, symbolic knowledge. I believe that research in DLs will become even more important in the near future because of these new requirements and the developments that will follow.

Did the standardization of OWL influence the development of the KR&R area, and in which way? Were there some disadvantages or unintended consequences? Is there a need for an update to OWL and the standards around it?

I think that the standardization of OWL had an important influence on the development of the KR&R area, and specifically on DLs, since it made this technology usable by everyone, and rather easily embeddable in systems and applications. And since domain knowledge is present and in principle available in many real-world settings, this domain knowledge can now be encoded in a standard way and made usable, using standard technologies. I do not really see disadvantages or unintended consequences of this. However, I think there would be a need to update OWL and related standards. On the one hand, it would need to take into account all the new knowledge that in the meantime we have gained about the three language profiles (QL, EL, RL). On the other hand, some aspects related to querying, the entailment regimes of SPARQL, and also mappings in OBDA are not satisfactory and would need to be considered again in a revision of the standards. Unfortunately, revising standards is a lengthy process, and sometimes not only scientific considerations play a role in the decisions that are being taken.

What are the scientific results you are especially proud of? Are there results we don't know but should know about?

I am proud of the works that directly come from the connection between Databases and KR&R. More specifically, let me mention the line of work that I carried out initially with Maurizio and Giuseppe in a PODS 1998 paper that introduced the problem of answering conjunctive queries over DL knowledge bases. That work was the starting point of a very fruitful research direction that was then taken up and extended by many other researchers, using a variety of techniques, and that later has been called ontology-mediated query answering (OMQA). A different twist to OMQA then came in 2005, when we introduced, also with Riccardo Rosati and Domenico Lembo, the DL-Lite family of DLs and defined the notion of first-order query rewritability. Both DL-Lite and first-order rewritability turned out to be very successful. I never made a count of this, but I would not be surprised if DL-Lite is the DL about which the highest number of papers have been written so far. In hindsight, I find it funny that in the first-paper about DL-Lite, which got rejected also due to a technical flaw that had escaped our attention (and that we discovered shortly after submission), one of the reviewers noted that in fact the logic seemed too simple to be called a DL, being little more than propositional logic but without disjunction and full negation.

As for results that not all readers of this interview might know, let me mention the line of work with Giuseppe, Maurizio, and Moshe on view-based query processing and related problems in the context of regular path queries and their

extensions, which we carried out in the early 2000s. It contains some interesting automata-based techniques, and also connections to the Constraint Satisfaction Problem.

What is your take on machine learning? Is there a place for symbolic methods in the current environment?

I never looked deeply enough into machine and deep learning, so I am definitely not the most qualified person to comment on this. But as mentioned before, I believe that symbolic methods will increasingly play an important role in connection with machine learning. How important, and how long this will take, remains to be seen.

In the past years, you have been combining foundational research with industry-oriented research. How do you strike the right balance?

I believe that also industry-oriented research should be grounded in solid foundations. In that sense, it should be easy to strike the right balance: first develop the foundations, then develop the right tools and carry them over to industry. I am joking, obviously, since it turns out that it is surprisingly difficult to get industries seriously interested in the kind of semantic technologies we are developing. When I say "seriously", I do not mean adoption in the research departments of larger organizations, but rather adoption of semantic technologies in a production environment, having an impact on the business. The latter requires a lot of additional work that goes into directions that cannot be easily pursued within a research group in a university department.

You are a founder of Ontopic. What are the main aims and ideas behind this company?

What I mentioned above is precisely the reason why we founded Ontopic. I believe that the effort to bring research results to industry can only be carried out by a company. A company can invest the needed resources in pure development work, and also in aspects that go well beyond what is of interest in research, such as user interfaces, addressing specific customer requests, and providing customer support. Such work would be difficult to justify in a research group. In Ontopic, we rely on the Ontop system, which continues to be developed as an open source tool in the research group at the Free University of Bozen-Bolzano. But besides consultancy and technical support, our effort is in developing solutions to make Ontop easily usable in complex industrial scenarios.

What are the future research directions that you consider viable for KR&R research? What is missing, what are the most pressing issues?

This is a difficult question. One limitation in the adoption of knowledge based techniques comes from the fact that often knowledge is not readily available or encoded in the proper format, but rather is hidden in textual documents and in data. So, the problem of knowledge extraction, using techniques based on machine learning is definitely an important research direction to pursue. Another grand challenge is how to integrate and make symbolic knowledge available to and usable by systems that rely on machine and deep learning. Or vice-versa, how to exploit machine learning technology not only to extract data and knowledge, but more in general to ease the design and development of solutions that are based on symbolic knowledge, and to make systems that rely on symbolic knowledge more efficient and more effective. These are directions in which KR&R researchers should get engaged, and not only the machine learning community.

What are your future projects? We know that you are also interested in the dynamic aspects of knowledge-enriched databases. Could you please comment on why this is an important research direction?

Dynamics of knowledge is indeed an important research direction, since all systems in the real world evolve over time. So, we need means to model such evolution, and we also need to study the properties related to system dynamics. In the past years, besides the work on OBDA, I have been interested in the verification of temporal properties for dynamic systems where data plays a prominent role and is not abstracted away. Dealing with such systems is challenging, since the possibility of incorporating data coming from an unbounded domain makes them intrinsically infinite state. Therefore inference over such systems becomes immediately undecidable, and one needs to impose suitable restrictions to regain decidability. We have also been looking at variations of this setting where we incorporate not only data but also knowledge. The overall setting, however, is rather complicated since there are many different choices that one can make, so that the space of possible problem configurations is very large. The results achieved so far are very fragmented, and it is also difficult to compare different proposals and get a coherent picture of the problem. It is fair to say that the area is widely open, and there is a lot of interesting work to do, for many PhDs and for many years to come.

What advice would you give to the young generation?

This may sound obvious, but what I consider most important in our job as a researcher is to follow one's passions, and study the problems that one likes. The more you like what you are doing, the easier it will be to invest time and energy into your work, and so you are more likely to make good progress and be successful. Initially, you might not know exactly what you like, so it is important to broaden your knowledge and get solid foundations of your research area in the wider sense, not only look at one narrow sub-field or specific problem that you intend to solve. You should also be brave and explore territories that might seem unrelated to what you are currently doing or have studied so far. One

can discover unexpected connections, and this might lead to innovative ideas and results. There are very good examples of this also in the area of DLs. The most prominent one is perhaps the connection between DLs and modal and program logics, which we now consider obvious. This connection was actually discovered and made widely known at the beginning of the 1990's (although it had been mentioned in restricted occasions already several years earlier), and then led to a very fruitful cross-fertilization between the DL community and modal logicians.

Coming back to doing what one likes most, it is also important to understand not only the right area in which you want to work, but also the kind of problems you like to look at: whether you are more inclined and interested in solving theoretical problems, or whether you rather prefer practical work, implementation, and experimentation. This might also depend on your background and your previous studies. But also here you should not be shy and be ready to try out and acquire new skills. You might end up liking this more than what you were doing before.

A further advice that I want to give is to be open and talk about your work and the problems you are looking at to your colleagues, even if they work in different fields and you might think that they are too far away from what you do. You might get unexpected inspirations or learn about some techniques or results that might prove useful for your work. For example, during my PhD I was developing a technique based on linear inequalities for finite-model reasoning in expressive DLs, and I had only a bound on the solution space of the inequalities that would give me a non-optimal complexity bound for reasoning. While talking to a colleague in my office (at that time, I was sitting in an office with six other PhD students and researchers) about a problem she was looking into, she mentioned by chance a paper dealing with a specific form of systems of linear equations. Precisely that paper turned out to contain a result that I could exploit to obtain an optimal complexity bound for my problem. In the end this led to three papers that we published in three top conferences, and that made up an important part of my PhD thesis.

Due to limited funding in science and education, many young researchers find it challenging to advance in their academic careers. How should one stay motivated?

We all continue to hope that our political leaders get wiser and understand that investing more in research and education would be a smart move that in the end pays off multiple times. But the signals we get from politics and the decisions that are taken tell us a different, much more depressing story. Still, young researchers should not give up because of this, and should stay motivated. You might think that it is easy for me to say so, since I have a tenured position and do not need to worry about what I will be doing next year. Indeed, in hindsight I somehow feel that I was lucky in my professional career with respect to how things developed. But this does not mean that I always saw things in that way, that everything fell precisely in place, and that my career developed smoothly with the perfect timing that one would expect. For example, I already said that I worked for one and a half years in a company in Austria, and that I quit my job there to start a PhD in Rome, more precisely to apply for a PhD position. What I did not say is that the competition was tough and I did not get the PhD position that year, but only one year later, after applying again. I had already quit my well-paid job and had moved to Rome, however, so I experienced right from the beginning, even before the PhD, the insecurity of the academic path. And after finishing my PhD, I went through various short-term contracts and was not always covered by a salary. It took me almost 5 years to get a stable position as an assistant professor in Rome. The salary I then received was quite low, and was not sufficient to sustain me, my wife, who did not have a job, and our two small kids. So, I had to do additional teaching.

A decent salary came only when in 2003 I moved to Bolzano as an associate professor. Also there, it took me 12 years to be promoted from associate to full professor, although I believe that I had the right qualifications much earlier. The reason is that we had a difficult situation in the faculty, due to internal contrasts, and this costed a lot of energy. The situation improved only when two professors (including the former dean) were fired, precisely for all the reasons I and other colleagues had been in strong contrast with them. But my attitude has always been that when there were problems or delays, but I saw a light at the end of the tunnel, I was ready to wait and continued to work towards my objectives. Instead, when the situation looked unclear and a different opportunity emerged, I was ready to take that opportunity and change. For example, this brought me from Innsbruck to Rome, back to Innsbruck, to Rome again, to Bolzano, in between to Vienna for one year, and now part-time to Sweden for the next 4 years. And I never regretted my choices. It might be that different choices could have led to a different, perhaps better, outcome. But this is totally irrelevant, since one has to do the best out of the taken path.

It is true that typically in academia it takes us much longer than for other jobs to reach a relatively stable and satisfactory situation. But this is a price we should be ready to pay for a job that gives us a lot of freedom and fun. One should not despair when the perspectives do not look good, and instead continue to work at the best of one's capacity. If one does a good job, this will be visible, and in the end things will turn out right. Let's try to be optimistic, even in these difficult times. This is the final message that I want to give.

Thank you for this inspiring interview!

Thank you for giving me this opportunity to reach out to our community!