# Elisa Bertino Speaks Out

## on How She Accrued 301 Coauthors, Revitalized a Department, Cut Her Commute to Three Minutes, Enhanced Our Trust in Shared Data, and More

**by Marianne Winslett**



http://homes.cerias.purdue.edu/~bertino/

*Welcome to this installment of ACM SIGMOD Record's series of interviews with distinguished members of the database community. I'm Marianne Winslett, and today we are in West Lafayette, Indiana. I have here with me Elisa Bertino, who is a professor of computer science at Purdue University and the Research Director of CERIAS, which is Purdue's center for security research. Before joining Purdue, Elisa was on the faculty at the University of Milan for many years. Elisa's research interests lie in database systems and information security. She is a former co-editor in chief of the* VLDB Journal, *and is on the editorial boards of* ACM Transactions on Information and System Security, IEEE Internet Computing, *and* IEEE Security & Privacy. *Elisa's PhD is from the University of Pisa. So, Elisa, welcome!*

*Elisa, you have written a* ***lot*** *of papers and you are listed very high in CiteSeer. For example, you have 25 papers listed in DBLP just for 2007, and you have 301 coauthors in DBLP. How do you collaborate with so many people?*

I really like to work with other people a lot. I try very much to understand what other people are working on, because I feel that I learn a lot from them, more than they learn from me. This goes back to my very early experience, when I was a very young researcher, 23 years old. At that time, I had a lot of ideas for research that I wanted to do, but I was always running into some very senior people in

our field who dismissed all my ideas. Later, I discovered that some of those ideas of mine had not been wrong.

Because of that experience, I really try to listen to the ideas of other people and encourage them to pursue their ideas. In many cases, I end up collaborating with them. And I really feel that I can understand research in other fields in computer science as applied to databases or security. So that's why I have all those collaborations. Also, I have been working for many years – that explains that number of coauthors!

*Well, maybe! I don't think I will ever have 301 coauthors! When you advise your PhD students, are you mainly listening?*

I think that each person is unique, so there is no one single approach which is good for everyone. Some students have their own ideas and a lot of enthusiasm, so when I see that, I really try to encourage them. Of course, I try to make sure that they don't take a really wrong direction, but I also really try to understand what could be of value in what they propose. Other students may really need a lot of guidance, because some are shy and don't want to express their own ideas. Sometimes I give the student a problem when I see that the student does not have his own ideas, or is too shy.

*I have heard that you became the department head at the University of Milan when you were very young, and then worked to make the department more research-oriented. For those of our readers who want to build up a research group or a department: what are the steps that you take to do that?*

In Milan, I wanted to invest in people, especially younger people. A major problem in Italian universities is that the faculty are quite old, and we don't have enough young people. So I tried to use the resources to get younger people, to get postdoctoral fellowships, and so on. In some cases, I said that I didn't want to buy more equipment, I wanted to get people instead. That was a big key.

Once you have people, the next step is for the people to have interesting research ideas and collaborate, even across the disciplines. That goes back to my idea that I really like a lot to do research. I think that *everyone* should like to do research, so I usually do my best to make other people do research, interact, and come up with exciting research ideas.

*How did you have time to do research and be the department head?*

That was very tough, to be honest. In Italy, I had to be the head of a department with a faculty of about 50 professors, and also I had to teach. One year I was teaching 20 hours a week, so I was working hard!

Fortunately, unlike US universities, Italian universities hire their own former students, which gives research groups a lot of stability. (But it has other disadvantages, I must say.) When I was the department head, my research group had several assistant professors, and some of them had been my own students. This stability helps a lot, because the older ones can help with the younger students and so forth. But there was still a big stress, I must say!

*Recently some well-known database researchers have moved from the US to Europe, but you have bucked this trend by moving from Italy to the US. What convinced you to move?*

There are several reasons. First of all, even though I was a professor in Italy, I had spent periods of time in the US here, visiting various institutions. In the back of my mind, I had the idea that I wanted to come to really see how it is to be a professor in the US. I had visited Purdue about six years ago, so

I knew several people there, and I knew that the Midwest has really strong research groups in databases and in security. When an opportunity was available for a faculty position from Purdue, they asked me if I had an interest in considering a position here. At that moment, it was something I really wanted to do, and that I had wanted to do for many years. I thought that it could be an opportunity that I shouldn't miss. It was a tough decision, since I had built a lot in Italy, but I also saw it as an opportunity to improve my research, by being able to work with the really strong research groups in the Midwest. You have to follow your instincts sometimes, and that's what I did.

*What do you miss most about Italy?*

That's an interesting question! What I would miss in the past, but not today in the Internet age, was not to know what was happening in Italy. When I was living in San Jose many years ago, you wouldn't be able to get an Italian newspaper unless you would drive up to San Francisco. But today, with the Internet, this is not an issue. You can get TV online, and you can be really up to date on whatever good and bad happens in Italy. Sometimes I miss a little bit the fact that in Italy you can easily go to the sea, to the mountains, and I miss a little bit the food. But other than that, I don't miss much, because you can get almost everything today, the world is so connected.

*Wasn't it strange to move from such a big cosmopolitan city like Milan to Lafayette, which is a fairly small and quiet place?*

This is a question that everyone asks me! Actually, I still feel like I am on vacation being here. Milano is a nice city because it is cosmopolitan, but it is also very stressful, with a lot of traffic jams. I had to get up very early, at 6 o'clock, to get to my office. You don't meet your friends there very easily because you have to drive two hours. And here, everything is so close, so comfortable; there is much more social interaction. And it may seem minor, but in Milan, there is an excess of noise, and you cannot see the stars because the sky has always some pollution. Here sometimes there is a perfect silence, which I really like a lot.

I think life goes in cycles. Maybe in 10 years when I'll be tired of being in a quiet place, I'll go somewhere else!

*How long is your drive to work now?*

Oh, it's maybe three minutes.

*Mine's four minutes, so you have it better!*

*Would you recommend that new computer science PhD graduates in the US consider a job in Europe?*

I think that working in Europe could be an important experience for young people. There are very good institutions there today. Universities care much more about ranking than they used to before, so there is much more competition there now. Some countries in Europe are very good at computer science and they are trying to attract a lot of talented young PhD researchers. The European Union is trying its best to help people moving from one country to another.

Of course, there are differences among different countries. In Italy, you do need to speak the language, because otherwise you wouldn't be able to teach, except for very few places. The funding may also be different depending on the country. Typically, each European country has its own national funding. The national funding can be very low in some countries, and then you have to work to get funding from the European Union. I feel that getting EU funding is much more difficult than getting funding in the US, because you have to write big proposals with a lot of countries.

Participating in those proposals is very interesting, and you get exposed to different cultures.

*You were born in Rome, which at that time did not have a tradition of work in computer science. What led you to get into computer science in general, and databases and security in particular?*

Both my parents have doctorates in mathematics, so I saw mathematical symbols all the time. I wanted to study mathematics since I was three years old! But at that time in Italy, if you studied mathematics, you would end up being a high school teacher. My mother thought that was not an exciting career. She said that I needed to study computer science instead, and I followed her advice.

Now, about databases, how did I get into that? I studied in Pisa, and at that time the professors there were very theoretical, doing work on formal semantics, programming languages, or complexity theory. I tried to find an advisor, but they were all busy with other students. So I ended up working on a thesis on databases with someone in a research institution in Pisa.

When I went to IBM in San Jose as a postdoc, I was working on database views for security. Once I started to work on database security, I learned more and more things about security. I got into other security topics, like digital identity management. I keep going back and forth between databases and security, in a way.

*You were a postdoc at IBM Almaden, and I have heard that there are some interesting stories from your time there.*

That was a very exciting time. I was just out of school from Pisa when I joined the System R* group. There were people there like Bruce Lindsay, Laura Haas, Pat Selinger (she was the manager of my group), Paul Wilms, and Guy Lohman. There were so many good people that I would walk around, looking at the names on the doors, and I couldn't believe my eyes to be there! It was very exciting! I really learned a lot from them. I had to do a lot of programming, which really made me learn the inside of a DBMS.

And the people were all very young. I remember working with Laura Haas, who had just started also with IBM. One day we were trying to debug some code and we were looking at the value stored in a particular location in memory. Before a certain instruction executed, there was one value in that location, and afterwards there was another value there, and we couldn't understand why. We tried everything we could think of, and then in the end, we went to talk to Bob Yost, who was the system manager there. We told him that the IBM 370 must have some hardware problem! Bob told us that this is not possible, go back and look again... then finally we found out that the problem was in the way the compiler would translate the move character instruction. Once we realized this, then we fixed the problem.

*After working at IBM Almaden and MCC, you left industry. What makes you prefer academia to industry?*

I feel that in academia, you are free to do what you like. You are your own boss. You can really decide your own research directions. And also, in Italy, we don't have industrial research labs where you can go to work. There is nothing there like an IBM research lab, or a Microsoft research lab. So if you are in industry in Italy, either you work for consulting companies, or you do applications.

*What impact do you think September 11 had on security research?*

Of course, that attack has shown that there are terrorist groups who will do what they can do to harm a country, and therefore, protecting cyber infrastructure is a key component of an overall security solution. But also, the attack has shown the importance of databases for security. After September

11, people understood that from data you can get a lot of useful information that can really be used to try to prevent those types of attacks, and to manage emergencies. People understood that all this large amount of data needed to be integrated, to be analyzed, so that you can extract useful knowledge. You need to be able to make relevant information available very quickly, so people can have it whenever they need it – but without being flooded with too much data, which is the other really big problem!

*What do you view as the most important problems in information security today?*

At the end, I think that the really important security problem is securing data and being able to trust the data. More and more people provide content and share data. However, can we really trust this data? Can we really rely on this information? Many years ago, the European Union started a research program to provide better and higher quality information on the Web. They observed that in a lot of cases, people go on the web to find medical information, and then believe this information. How can we make sure that this information is good? A lot of people use Wikipedia now; can you trust Wikipedia? And we will have the same problem for data in databases, which are more protected. They say that your network can go down, or your machine can be broken, but as long as your data is safe, you are okay. I am really interested about how we can assign trust levels to information.

*So what approach do you think we could take to solve this information quality and information trust problem?*

It is a really interesting research challenge. I would say that you have to combine a lot of different solutions. For example, semantic integrity is a very well known approach we have in databases, but it may not be enough. You may also need to use approaches to improve the quality of the data. You may need to keep track of provenance information, and in fact there has been a lot of work done in this area. But somehow this type of work has never considered the possibility that you may have malicious parties who may try to change the provenance information, and then may try to deceive you by providing the wrong data.

My group has been examining approaches where we look at multiple sources that supply information on the same topic. We decide how similar the information is that they provide, and determine the distance between them. Based on that, we can rate the trustworthiness of a party that provides information. We also need to consider that somebody will be using this information to make a decision. There can be policies for deciding whether to use a certain data item in making a decision, based on its trustworthiness. In the end, it is better to tell a user, e.g., that this data we couldn't verify and it's not trusted, so you may not want to use it if you have to make a critical decision. Then the user will be aware that the data is not so trusted, and can decide whether to use the data or not. There can also be policies for deciding whether you want to admit certain data in your databases, based on how trusted it is.

*Do you see differences between the database and security research communities, for example, in the way program committees function or the way work is evaluated?*

I have been in the program committee of the Oakland conference (the IEEE Symposium on Security and Privacy) and CCS (the ACM Conference on Computer and Communications Security). Oakland and CCS are both very selective conferences, but they receive fewer submissions than we do in the database conferences. Oakland usually has an in-person PC meeting.

A very interesting difference is that a lot of security research looks at attacks and defense, and we don't do that in databases. Many database security papers don't have a security analysis, where you show the possible threats to your approach or techniques, and possible mitigation. We always have a security analysis in papers we send to security conferences. While the type of research I do is more

on policy-based access control, I think it is also fun doing this type of research on attacks and defense. Sometimes I wonder whether the emphasis on attacks and defense is a US attitude, because the US had its Wild West, the frontier to conquer. Perhaps now there is the idea that you are the West, reasoning in terms of enemies trying to attack you.

Another difference is that in databases, we give a lot of attention to performance; performance is the key. Some security papers talk about performance, but usually this is not the main aim of security research. The main emphasis is to make sure that you analyze the security of your technique and your system.

*Traditionally, it has been hard for professors to collaborate with researchers in industry, due to intellectual property (IP) issues. Can you tell us about your new Open Collaborative Research (OCR) project with IBM? (See http://www.cs.purdue.edu/news/12-14-06IBMocr.htm .)*

The goal of IBM's program was to try to develop common research but make the research results and artifacts such as software available in the open source community. And therefore, we didn't really have IP rights to consider. Purdue was very open minded and agreed to these terms.

Our team's collaboration is going very well. Our project includes Purdue and IBM, and CMU as well. CMU is doing interesting work on usability for security and privacy techniques. So we are getting to learn new things, while functioning as one team.

*Are other companies doing that, or is it just IBM?*

I think that the OCR program grew out of an agreement that was signed by several companies, including HP, IBM and others. These companies noticed that IP is getting in the way of doing good joint research which would be a benefit to companies and universities. I don't know about the other companies, but IBM really tried to implement this agreement, by starting the OCR program.

*What have you learned from your PhD students?*

They really provide me with a lot of enthusiasm to continue my research. It is because of them that I never get tired. It is really exciting to see their ideas, and discuss with them. It is really a learning process.

It is also very interesting is to observe that people are very different. For example, I am proud that I have had a lot of female PhD students. Sometimes they are different from the male students, perhaps more shy but more dedicated to their work in a way. But again, as I said before, each person is different from each other.

*How do you make time for everything?*

I do as everyone else does. At least here at Purdue, everyone works a lot. Even after so many years, I really like research, so I don't feel like I am working. As a consequence, I work a lot. And so this is how I manage to do many things, I guess.

*Do you have any words of advice for fledgling or mid-career database researchers or practitioners?*

I think today we live in an exciting time. Computer scientists are changing the modern world. Probably we never anticipated that our field could have such an impact. I think about the Web 2.0, the new Web 3.0, there are so many applications!

I feel it is really important to have a vision and keep track of the trends in the world. I think that you

have to think about the future, because there are unprecedented opportunities and possibilities. Thinking about that makes you realize that you are really doing something exciting.

*Among all your past research, do you have a favorite piece of work?*

I have one nice piece of work on query optimization with unstructured data [Elisa Bertino, Fausto

Rabitti, and Simon J. Gibbs. Query Processing in a Multimedia Document System. *ACM Transactions on Information Systems* 6(1): 1-41 (1988)]. In 1988, nobody was working on unstructured data at all. I still remember that I was thinking to send the paper to SIGMOD, but then I decided not to. I was sure my paper would be rejected. The PC members would say, what is this "unstructured data"!

The paper is about a query optimizer for a multimedia document management system which had text and structured fields. We had to define a query language which had the equivalent of XPath expressions, and I had to develop the optimizer for it. Fortunately, I had learned a lot at IBM, so I implemented all of the optimizer and wrote my first ACM Transactions paper. I really liked that work.

Right now, in a lot of cases, the software produced by my research projects is developed by my students. At that time, I did everything by myself: the theory, the idea, the organization, the implementation, the testing, everything.

*If you magically had enough extra time to do one additional thing at work that you are not doing now, what would it be?*

I would try to become a *cordon bleu* chef. I like cooking a lot! Especially in Milan you get spoiled. Milan has a lot of restaurants, and very good places from which you can take away food. For ten years, I didn't cook at all. But now here at Purdue, I do cook a lot, which I like anyhow. I like to improve.

*If you could change one thing about yourself as a computer science researcher, what would it be?*

I would like to do hardware-oriented research. Dennis Tsichritzis gave me a very important piece of advice more than 20 years ago. He told me that if you want to understand where new research in software has to go, look at the technological advances in hardware. Those advances will really tell you where you have to direct your research. I think that nanotechnology is really an exciting field. I wish I could have done a lot of work in hardware.

*Thank you very much for talking with me today.*

Thank you, Marianne. I really enjoyed our conversation.