Erich Neuhold Speaks Out

on industry research versus academic research, funding projects, and more

by Marianne Winslett and Vanessa Braganholo



Erich Neuhold
http://cs.univie.ac.at/mis-team/infpers/Erich_Neuhold/

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 Indianapolis, site of the 2010 SIGMOD and PODS conferences. I have here with me Erich Neuhold, who is a professor at the University of Vienna. Until 2005 he was the director of the Fraunhofer Institute for Integrated Publication and Information Systems in Darmstadt, and a professor at the Technical University of Darmstadt. Erich has also worked for IBM and HP, both in Europe and in the US. Erich is an IEEE Fellow and a Fellow of the Gesellschaft fuer Informatik. Erich's PhD is from the Technical University of Vienna. So, Erich, welcome!

Thank you very much for having me!

What does it mean to have an integrated publication and information system?

I think in order to explain, I will add a little background. When I took over the directorship of the Fraunhofer Institute (formerly a GMD Institute), I felt that, coming from the database area, we

needed to add the human to this whole approach. In this situation, the human is a person that uses information, and then maybe creates new information. This is a cycle, and that's what integration here means: it is the cycle of locating information, processing it, finding the human user, creating new information, depositing it again. That is what we call the integration of information in this publication and information cycle. So that gave rise to the name, and we got stuck with it in reality, and couldn't change it because at some point publication was not anymore the right word, and the Web came along, and you couldn't call publishing what you now do with the information. But we got stuck with it, we had a brand name "IPSI" and we didn't want to change that, so we stayed with it, never finding a new paradigm for it.

Well, you had all these different groups. You were mentioning this really big Institute with 120 people total and a $10M/year budget. So what was it like to manage something that big?

Before I started with this institute, I was at HP as the director of one of their research labs. I had also about 60 people working for me. So anyway, that provided experience. I mean, if you step from, let's say, a small research group with 3 or 4 PhD students, to manage 120 people, some of them are older than you, and some of them are the same age, and some are really young, then this needs adjusting your behavior. But I felt it as a challenge, and I took up to that challenge when I took the job, and I think it went quite successful.

I'm actually a new director of an institute myself, so maybe I should ask you: do you have words of advice for me, and for the other people out there who are taking on these new roles?

"We distributed all our business cards. Everybody was interested". And my question was, "how many business cards did you collect"?

I think one has to look for contact with the people. I mean, in some ways, it is a little different from industry. When I was a director in industry, for example, it was not a good style to go down and work with researchers, because of my management team. I had two lines of management between me and them. They would come to me and say "don't you trust that we can do it?", whereas in a research community, this is different because as a professor, I was essentially also the PhD advisor of those younger people. And of course as a PhD advisor, you have to look for personal contact, and not delegate it to your management, because that would not really work. So this is a very different behavior I found between management in industry research and management in academic research.

What do you think of the way all this XML stuff has turned out?

I think when semi-structured data came along, let's say documents were put into databases (let's call it that way), I found (talking with publishers and other people) that actually storing the structure of these documents in the sense of XML or SGML at that time, was important. Then, of course, HTML came along with the Web, and proved that this concept was important, and became successful, I believe, because it was simpler than SGML. XML has a lot of SGML and HTML, but it is complicated. It is simpler than the first one, but more complicated than HTML. For that reason, I believe that it was a steeper step for people to go in this direction, and that is why XML wasn't as successful as I believed originally. I thought it would essentially take over the Web.

Well, I guess a long time ago, people thought object oriented databases would play this role, and then the relational vendors just added object oriented features. So is it the same way with XML, or did something different happen there?

I belong to the people that started out with relational databases, and then got very enthusiastic with object oriented databases, because in order to talk about entities, you shouldn't flatten them out into a table. You want to combine the attributes of different properties together, interconnect entities and also add manipulation functions. So I thought that was the way to go… And it was at a similar time when C++ came along in the programming world, and developed in about the same speed as it did in the database world. But then, object orientation in databases wasn't so successful. First of all, there were these strong vendors of relational database systems already, so there was steep competition. But I believe also that, in my opinion, the first commercial object oriented database systems never really worked well. They didn't really have the performance, the scaling, and things like that. And so people got disappointed with the vendors behind it and the established RDBS vendors pushing pressure on them said "there is never going to be any better [product than relational databases]". A self-fulfilling prophecy...

Do you believe in the Semantic Web?

No. I mean, this has to be qualified. The Semantic Web, in some way, is a vision. A vision in which many of us didn't believe in from the very beginning, because we felt it being too ambitious. You will not be able to describe all the semantic information that is hidden in documents, in their structures, and in any context information, in the way it was assumed by the semantic Web people, in order to be able to make automatic deductions and all kinds of other things. For a long time I have worked in data integration and as a curator of databases, and terminology discrepancies and all kinds of other problems have not been solved. So it was quite clear they are not going to solve them in a Semantic Web, and they didn't. Despite that some of their algorithms made sense or made contributions… For me, it is like in artificial intelligence. Artificial intelligence always had such high goals. And then it didn't succeed, and then funding stopped, and everybody kind of said they were not successful. But if you look a little more carefully, a little deeper, they actually made many contributions. There are quite a number of fields where their algorithms, like machine learning algorithms, have moved into systems. Supervised learning and unsupervised learning have played a role in many other applications. But of course, we are not recreating human intelligence with those machines.

Have the Web standards bodies played an important role in the creation of the Web as we see it today? The W3C…

I would say definitely not at the beginning. Tim Berners-Lee himself, he was working at CERN at the time when essentially the Web was created. He and his colleague Robert Cailliau actually wanted to build an SGML engine, but their manager – who told me the story – advised them "you need two years and three people to build that, but I'll give you half a year, and you two do it". So they had to cut down on SGML, simplified it and developed HTML as a hypertext system. There was no standard for doing it. It was just a simple code that allowed exchanging documents in the high energy physics domain, but they were very successful. And of course interfaces and tools like Mosaic and other things, made it useful for everybody. Again, Mosaic was not built on any standards or anything. Later on, when the field became a little more mature, and many other people joined in, I think standards started making more sense. If you don't have standards, for example, if you look at an HTML document, how to visualize it? You have many choices. So in that way, Mosaic was an ad-hoc standard. People just keep following whoever has the first successful approach. But then later you have to have some control of that. I think, then some of the standard things are very valuable, but you shouldn't go too early, because otherwise you will restrict the development.

How is it different to do research in the US versus in Europe?

This is a difficult question, but I will try to answer it. I think for example, research in Europe has similarities now within the universities at both places. When I started out as a professor, a long time ago, a professor in Germany had a number of research assistants, a number of teaching assistants, provided by the universities. So essentially, you had a budget and you didn't need to go out for project money or industry things. That changed. Of course, even then ambitious professors would always go for additional money. Starting out I had 3 researchers and 2 technicians from the University in Stuttgart, but in the end I had a team of 15 or 20 people. So I was able to have large projects. But the pressure was different; you didn't have to do it in order to be successful. So, that gave you an advantage, because you had a larger number of people, but you were free to do with them whatever you liked. There was no funding available in distributed databases when I started working on it. I started working on it because somebody in the institute bought about 10 PDP 11 computers and then we didn't know what to do with them, because the guy who was supposed to do something with them left. So they were sitting around with 64K addressing space, and as I was in the database field, I said "wouldn't it be good to distribute a database over these small machines, and utilize these machines in a distributed system"? There was nobody who would say that was a great idea, at least not at that time.

It is amazing to think of doing something significant with 64K of memory.

Oh yeah, but that was the time. I mean, I was talking to Nicholas Wirth at some point and he built the personal computer LILITH. We discussed, and he said "oh, 64K of memory is plenty for writing programs! Nobody writes larger programs". He was at Zurich, at ETH at that time. And I told him "this is not enough, I work in databases, and where do you store the data"? And he said "oh, you bring them in as you need them". And then a year later he came back to me, and said, "You were right, and I need a larger memory". But it was not because of the data. It was

because of the visualization he wanted to do on his screen. The machines were slow, so you didn't do graphics on the fly, you prepared the screen in the memory, and he needed as much memory as he had (64k) just for the higher resolution screen. So this was a very interesting observation.

Coming back; on the other side, I feel that having had NSF projects in the United States, and EU projects in Europe, the overhead in Europe is much higher and slows down research behavior.

When you say overhead, do you mean like management overhead?

Yeah, yeah, administrative overhead.

Is that because of the meetings?

This might be because of the multi-national teams, and of course, the idea behind it is that cooperation over Europe should be encouraged. On the other side, you travel a lot, the behavior, knowledge and culture of the different nationalities are different, and you spend a lot of time in meetings arguing about things which you would much better solve in your own home place. My institute had many researchers and in some of the areas we were involved in, we could have done the whole project much more efficiently ourselves, but we wouldn't have gotten any money for it without international partners, whereas in the United States, as an institute, I probably would have gotten the money.

The first commercial object oriented database system never really worked…

You've moved back and forth between academic and industrial research. What leads you to make those changes?

Opportunities. I mean, I was in industry, I worked for IBM at the beginning, and then in IBM, I moved into the field of databases. Then when I moved back to Europe, after being in IBM (I was in New York at that time), computer science in Germany was building up. Some friend of mine, whom I knew from earlier days, told me "Oh, I became a professor. Don't you want to become a professor too"? So I just applied, and I got 2 offers, one in Darmstadt and one in Stuttgart. In the end I took the Stuttgart position. Amazing enough, because you mentioned my Darmstadt association, I went to Darmstadt 20 years later. So it was Stuttgart, and I left IBM, and I built up a team there, and then I got contacted by Joel Birnbaum, I think he was vice-president of HP, whom I knew when he was the head of computer research of IBM in New York. I had spent some sabbatical leave there, so he knew me, and he sent me e-mail where he asked "do you know somebody who could lead a research group as research director in my team at HP in the database field"? I thought he meant me! And I applied. And it turns out that he didn't mean me because when he knew me in IBM, I was actually working on compilers. I even wrote a book on

compilers, so he thought I was a language person, and was very surprised, but he hired me anyway. This was kind of the reason I switched the second time into industry. And then I stayed in Palo Alto, and I enjoyed it very much working for HP. But I had a slight problem with my family, because they didn't want to come. I could have tried to force them but without ever applying, I got offers from Europe, as a professor again. So essentially, as a test I asked all that I wanted from the universities. If they gave it to me, I would take the job. If they didn't do it, I'd stay with HP. They gave me everything I asked for and I went back to Europe.

So there is one more switch to Fraunhofer, right?

Yeah, yeah, right. Actually, I went to Vienna as a professor from the United States. The Austrian president came to California and told me "we need people like you in Austria". So I got tricked into that, since when I came to Austria others there told me "we don't really need you here", but more polite, of course. As a result, I was not so happy, and I started looking around again. Soon, some other colleagues knew that I was looking around. I looked into the United States, looked at HP actually, as I thought to go back and even got an offer. But I also looked at academia at that time, and GMD came to me with an institute proposal. Later GMD (German National Research in Computer Science), was merged into Fraunhofer. GMD was more research oriented than Fraunhofer. Fraunhofer is really applied or contract research and GMD was more like basic and applied research. It did not as much research as the Max Plank Society. For example. Max Plank doesn't have to go out for projects. They have enough funding to perform their research; consequently they do more basic research. GMD was an in-between organization. At that time there was an outside research institute which actually fell apart working in the documentation field. They were chartered to develop things for online databases, like CAS (Chemical Abstract Services) in the USA, or FIS (Technical Information Systems) in Germany, but were not successful. In the end they were dissolved. A number of people left, and GMD was supposed to take over the remaining parts, and thus needed a leader for that. That's how they approached me. It was a challenge to build up from 20 people to the kind of what was later an institute of 100, 120 and more people, more than many American university CS departments.

So did the whole thing become contract research?

No, it was part of it. I mean, at GMD, my institute had about 50% that was directly government funded, and another 40% was typically EU funded with joint projects the EU was offering, and some from the German National Science Foundation. But this was a smaller part. Industry played a small role: a maximum 5-10% of the funding came from them. At Fraunhofer, this changed. They actually expect you to have about 30-40% of industry money, and then another let's say 30-40% of EU money, and only 20% would be given to you, directly. It is a very different model, especially as the pressure is on getting industry money. If you are not successful with the industry money, you are losing basic money also. You are in a situation where failure here will actually lead to your institute being out of money even if successful in other areas.

Well, how do you go after industry money? I think a lot of our readers might be interested in how you get money out of industry.

It is a lot of footwork. We went to fairs and exhibitions with demonstrations and showed them our capabilities. I'll give you the attitude of, let's say, a university person, and the attitude of somebody who wants industry money. My people came back from a fair and happily told me "we distributed all our business cards. Everybody was interested". And my question was, "how many business cards did you collect"? And they would just look at me, and say "huh"? Of course, as any industry person will know, you have to collect the business cards, because you have to make calls after that, and then you say "Yes, you were at my stand, and you saw my demonstrations. Are you interested in talking to us? We can probably help you". In Germany this is difficult, but also in Europe altogether. I believe that happens in computer science, not in machinery or manufacturing or other fields because there is very little primary industry in CS, nothing like Microsoft, IBM, and all those. What we have are application industries, banks and logistic companies and other similar ones. They don't want to take your prototype because people want product quality software. Essentially, they want your consulting, but you need money for doing research, and this is a stretch, it is not easy to bridge. You have to really learn how to do that. We had training. We had people that came in to teach us when GMD was taken over by Fraunhofer. One of them was a friend of mine. He headed IT for a bank in Germany. He would come and sit there and my people would come and make their 'sales' presentation. If he didn't like it he would say "thank you very much for an excellent presentation, don't call me, I will call you". This of course was a clear indication he wasn't impressed. But then he also analyzed the talk and told them why he wasn't impressed.

Even in the era of the web, the computer science is not topmost in the minds of the people.

Oh, that would be very helpful! Why is it so hard to get the scientific community to accept computer science as a first class discipline?

Because they are used to use computers as a service. They think computers are there, and computer science is just the industry that builds computers, but of course it is not. Very often, in research environments, the researchers write software themselves. Take a physicist, for example, in CERN, or in Stanford. Most of the software written there is not written by computer science people. It is written by physicists. So they feel that computer science is a service discipline for them, and not a standing by itself research field.

So do you agree with them, are we a service industry?

Of course I do not agree, but we had the same problem in Fraunhofer. Fraunhofer spreads to all disciplines and the largest institutes, the very successful institutes are in, let's call it mechanical

engineering. They work with the Germany car manufacturers, and they work with the famous German tool manufacturers and so on. Fraunhofer was created after the war to actually help build up the German industry again. And you know the success story of Germany is in the manufacturing industry. So these institutes considered that we, the IT institutes, would be essentially a service for them. They essentially came to us and told us "can you adjust this for us, or can you do this for us?" They were not interested in a partnership. They looked at us as a service. The Fraunhofer directors have a meeting once a year, where all directors are together for two or three days, and usually we have discussions about that view, because we IT people of course would not accept such a secondary role.

Did they ever change their mind, or is there something we need to do in our discipline to change their minds for them?

I think it is difficult. I think even in the era of the Web, the computer science is not topmost in the minds of people. If you go to Facebook and play around, or if you go to YouTube, or Twitter, or whatever, you are not thinking that behind all this is a computer scientist. You think of it as if someone just writes software that you can use, but seeing it as a discipline that analyses the data, for the good or for the bad (I don't want to comment on that), and a whole scientific research field behind it, I don't think is in front of the mind of the people. Maybe it needs time. I think mathematics may be in a very similar situation. However mathematics is accepted as a separate discipline, but they have a 3,000 years old history, and we have 40 years of history.

What do you think will happen with the conference system way of publishing in computer science?

I think we are in a time of changes. Because of the internet and the electronic availability of documents, now proceedings and the conference documents appear in a digital library essentially at the same time as the conference takes place, so the need to access documents by going to a conference will go away. Before, it was precisely that need, especially in the beginning, when we didn't have many journals. The way of publishing high quality work was through a conference. This was the main role, and that is changing now. The conference becomes more and more a social gathering place, because you can take the proceedings home, or you can download them later, but you are meeting peers, you go to sessions to discuss. All the conferences have now a tendency, like SIGMOD this year, to have the papers shortened in order to have more interactive sessions. There are people who have a paper and also have a presentation with a poster. We don't call it poster sessions now, but it is an additional possibility to discuss with the authors directly for better interactions. This social gathering aspect is increasing. But of course the other aspect, namely to have an achievable document in the database conference instead of a database journal will decrease because of that. At least, that is my opinion. But there is another situation coming up. I believe that maybe for the next generation of researchers (the one that is now 10 years old), physical gatherings may be in danger. This is because they have all this Facebook stuff. I see it with my grandchildren. Funny as it is, they are sitting in the same room, and talk to each other via Facebook, and not directly. They have this group of Facebook friends, and they chat and chat, and exchange all kinds of information. When they meet the same people physically, they don't know what to talk about. In a way, I believe there is a danger that the conferences as a social gathering place will also disappear, but then what is remaining and what are we loosing?

Definitely having a drink or a coffee in a bar together. I am watching this development very carefully, but I see a tendency in this direction and I believe it is wrong.

There is a danger that the conferences as a social gathering place will also be gone.

Do you have any words of advice for fledgling or midcareer database researchers or practitioners?

My advice is that you look around. See "hanging in the air" problems. I mean, at the conferences, sometimes good keynotes will raise issues. Look around your own environment and see some problems, and then you try to formulate a goal. What would you like to see happening here? And then try to find an approach of how you can solve it or partially solve it, if it is a mega problem. I always keep saying to my PhD students, "you should also be willing to say no to your advisor", because he is caught in his own context, and he may not see some developments. I changed fields a number of times. I worked not only on the database field, but earlier in programming languages and later also the in the digital library field. But when you are older, it is very hard to establish yourself in a new field. So for that reason, you get kind of stuck in your field. But young people or midcareer people are still free. They can still make this change. So that is my advice: follow your own intuition and not necessarily your advisor or a prospective (short term) job situation.

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

I would really go and look at the whole issue of interoperability and semantics and modeling and try to make some sense out of the many conflicting and many repeating approaches. I'd try it for myself. First of all, find out whether there was any progress in the last 20 years in the field. But then also try to find out what would be a good thing to really help a human centered approach, to really help the human to find things she/he does not know at all. Most of the approaches in semantics, even if they are semi-automatic and with human feedback and all such stuff, always assume the human knows. But if I don't know, I can make no judgment. If I go to, let's say, a medical wiki and I believe I have this and that sickness, and I find treatments in there, how do I know I can trust this information? I mean, I can only do it when I know doctors that have a high reputation or experience, and they back these descriptions. But I don't know the doctors because the whole subject is new to me. My human feedback is not going to help either. The question therefore is: how can you build systems that do not assume the human knows what he/she is doing?

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

I don't know. I started in electronics, I have to say, and I switched to computer science only when I joined IBM. I became a programmer, and the group itself was theoretically oriented,

consequently I did formal language stuff for programming. But when I got in contact with databases, I realized that one thing that would have helped me a lot, even in the early days with SQL and relational approaches would be to know more about linguistics. I feel knowledge about linguistics, concepts of speech, and natural languages, etc., is very helpful. That is even more helpful and more important nowadays. When I started IPSI, I had a linguistics group as one of five research fields. But we needed additional research money and eventually I couldn't maintain the group. I would get projects, but I would not get linguistics projects because 'true' linguists at the university would get that part, even of cooperation projects. Here I regret that I had no linguistic background myself. I was an outsider, and the funding agencies would not give me the money. They gave me the implementation aspects, the software aspects of the project, but not the linguistic aspects. If I had been a linguistic, even as a second kind of expertise I probably would have been more successful.

Thank you very much for talking with me today.

Thank you very much for having me.