

KM Forum Notes 1/19/2012

Notes written by Kate Pugh (thank you for the heavy lifting, Kate) with supplements by Lynda Moulton and Larry Chait.

Introduction from Lynda Moulton:

- Friday breakfast meetings have been suspended.
- Next speaker is Kathy Hagan from Wisconsin. Does knowledge transfer processes in “twin cities”, Midwest. She has been following us. She’s in Boston in February and will be our guest on Feb. 16th.
- Meeting in March will be 3/22 and combined with SIKM Boston.
- Kent state has a survey out on the functions and roles of knowledge professionals:
http://kentstate.qualtrics.com/SE/?SID=SV_2mnGKlxuLKLusyU

Larry Chait

- Chait and associates. KM strategy, process improvement. Doing work with nonprofits. Chait’s client is now developing a knowledge repository of consulting knowledge. Using Salesforce.com.

David Ritter

- Business manager at InnoCentive. Helps companies solve problems using crowd-sourcing. 2001 founded and then spun out from Eli Lilly. Used intranet to collaboratively solve problems. Run competition based problems. Clients include pharma and NASA. Challenges are public on intranet, and also do crowd-sourcing internal to organizations. We have a regimen around IP transfer as well. We are continuing to develop the capability via products. Have a big interest in informatics. Netflix ran the 1MM prize a few years ago. Now you can “gamify” this and create a scoring algorithm and people can see where they stand on the leader board.
- Lynda: InnoCentive did a presentation some years ago.

Lisa O’Donnell

- Genzyme information/records management. Working with R&D to track compounds of interest.

Geoff Bock

- Gilbane group, Seybold group, also Linda’s partner. Collaboration, content management. Also work with clients to develop marketing strategies and analyze biz benefits.

Andy Rittenberg

- IT clients – help develop products. Also certify / teach students on ITIL.

Dan Custer

- InnoCentive. Interested in analytics and informatics type challenges. How do you incite people to participate productively?

Dan Bogaty

- Clinical KM group for Partners Healthcare. We have all kinds of content and forms and we're trying to make it actionable and useful. You often "look at A and have no idea how it relates to B"

Dan Corwin

- Have been working on Search engines for the last 5 years. Semantic search: Revolves around the interplay between words and meanings.

Mark Sprague

- Relay technology. Publishing group spin off, studying drug discovery.

Bob Powers

- Consultant with Predictive Medicine. Do IP analysis. Semantic Web. Technical solutions. (Actually just hired someone from InnoCentive.)

Phil Macphee

- Focus on Data Modeling/Management.

Otto Ritter, AstraZeneca

- Lynda: Tell us about your background
- Otto: Participated in the Mathematical Olympiad as kid. Had a climbing accident in college, and focused on study. Then went into molecular biology. Was recruited by a start up in genetic engineering. I went on to biological research. Cloning HIV antigens.
- Lynda: When did people recognize HIV as a disease?
- Otto: 1983.
- Otto: I always just loved solving problems. I enjoy the freestyle approach. Whatever tool or trick you can use to solve a problem. Some people say, "There is nothing more practical than a good theory." In the abstract many things meet. It's a very useful shortcut. It's appealing to me – even if you use someone else's tool, the intellectual work is parsing the problem and mapping it to a problem already solved.
- Lynda: When did you take up informatics as a discipline?
- Otto: In 1990 I worked in Heidelberg, Germany, and my wife was at the Cancer research – including the pap virus. I visited my wife and was offered a position. I stayed 10 years and could have stayed longer.

- Lynda: The term “informatics.” When it was coined, were they thinking about unstructured content, content extraction?
- Otto: In the 1990s in Heidelberg we became part of the human genome project. I worked on a problem : integrate info from genome related sources. E.g., nuclear type sequences, protein sequences, genetic and physical maps. There was no sequence at that time! Also I looked at disease information, including inheritance. These sources had their own idiosyncratic structures. They were implemented in files, relational databases, XML-like coding (older). So the problem was: How do you unify semantically, and then how do you connect the data with the methods. You need to make it actionable. Genome informatics and bioinformatics terms emerged.
- Lynda: Did you have to educate people on the value of this?
- Otto: I was very lucky... I just did and people found it useful. I had funding from the European Union and here [?].
- Otto: I think the insight is nontrivial. When you parse something in your head it’s hierarchical. You can encode your understanding as a tree. It has a set of roots... down to the leaf. If you can manipulate such lists, you can recognize what it is and extract and transform and merge it with other things. This is actually what we do with XML documents. In abstract, it’s a simple idea. I was lucky to do it in an area that was well-funded.
- Lynda: Did you develop your own symbolic language?
- Otto: I did something similar at the Brookhaven Labs (NY). Molecular structures – DNA, RNA at the autonomic level. The system evolved over 25 years. It had a structure, file system and there was an industry of software. But the auxiliary information about who does what when, how you develop the grammar... that was new.
- Lynda: You had to work hard to get people to apply the ideas similarly? Were you the first at AZ?
- Otto: I was recruited by Ken Fassman (sp?). I knew him from the genome project at Hopkins and Whitehead. Then he went to Astra Informatics in Cambridge before the 1999 merger with Zeneca. He brought me in. There was a commitment to this. There was a global organization in Discovery. It was somewhat schizophrenic – some legacy app responsibility and some discovery.
- Lynda: Do all pharmaceuticals have informatics professionals?
- Otto: Yes. The jury is out because all pharma companies suffer from declining productivity. Discovery of drugs is an information business. You try out ideas and then decrease your uncertainty. This is not really prospering. The informatics groups have not discovered how to do it well.
- Lynda: Can you use idea people to direct you? They conjecture and then you pursue it?
- Otto: There is an overload of ideas. It’s more about the feedback . How do you follow up and prioritize. You need to know your capabilities. You need to see the difference between your anticipation and outcomes. It typically takes 15 years. >1B dollars. No one is in charge of this.
- Lynda: Are there politics influencing how the results of informatics are used?
- Otto: The issue is due to the long range and the uncertainty. We can learn from natural systems. They increase their fitness by watching the impact of the systems. It’s more than just watching your prey—you have to develop models and the models need to communicate. When you learn the outcome of your plan, you revise. Companies that compete on analytics, e.g.,

insurance, they do this systematically. Pharma companies have yet to learn this. The chemistry and biology need to be connected to the market – perceptions of physicians, investors, markets. It's manageable. If you externalize your representations, your representations evolve.

- Lynda: What about legacy learning? Is that factored in?
- Otto: That's part of it. The anticipation is based on some experience. The cortex uses something to evaluate signals. You build models on top of your models. At the top you have more and more invariants and then it expands into more refined signals into your muscles. It integrates the past, but not uniformly. If important with the future, the ideas are weighted more.
- Otto: In Pharma we don't learn from the past because the loop is too large. We haven't learned how to evaluate information totally.
- Lynda: I'm always astounded at how companies throw out past projects. I try to get them to codify information and state why they abandoned it.
- Lynda: You are a user of Linguamatics.
- Otto: As a company, not the app.
- Lynda: This is a semantic text mining tool developed for the pharma(?) industry. It's largely used for (semantically analyzing) published content in conjunction with enterprise content. A company would license a corpus of millions of documents, merge it with internal content and the engine does computing to discover concepts. Then scientists using natural language query the resulting database index. Are people using this?
- Otto: It's one of the good ways to search and retrieve something. It's like Google. Keyword search is valuable, but the added value is about the connectivity about resources. If things are highly connected they have more authority, weight. Linguamatics, in a simplified way is like several Google searches. It compartmentalizes phrases and gives multiple searches. The value is that it sees the semantics. "I want to see all mammalian genes..." It knows humans are mammals, etc.
- Lynda: Used for unstructured content? [Natural language processing or NLP – computational linguistics underlies this technology.]
- Otto: When I joined AZ 12 years ago, I suggested that we take structured info and make it into sentences and then using the mining information. I would say do it again! Historically space was precious. You had in your head/code what you meant in the columns. Until that source is being used outside the context it's fine, but then you need the metadata. It's a lot of work to add the meta information. Sometimes it's impossible, as the meaning has evolved, so the terms are not semantically consistent.
- Otto: Everything can be put into a relational database. Then you can have views on the representational data. The view is another relation. You can form one or more sentences. Then from a relation you can produce lots of sentences. Then you can discover and connect things more easily. You can more easily reuse someone else's algorithm.
- Lynda: Outside of pharma, where else should we be applying this in the rest of our world? Are you seeing any of this?
- Otto: We are applying this everywhere. It may not be a conscious endeavor, but it's a necessary survival skill. When you think about physical survival – our brains are very smart. We can just

observe the conscious part. I just catch a ball. We've learned to externalize what we know. Instead of building bigger brains we externalize this with computational engines.

- Lynda: If we can informally exchange it, and people don't use it and synthesize it, what's the point? I notice that the average person cannot answer more than one question from my emails. We have a shortening of our focus. It makes it hard to solve problems.
- Otto: The richness of information needs to be met with attention.

Questions?

- Larry: The trick is that my style is to go back into the email.
- Lynda: It's an attention.
- Dan Bogaty: You talked about the tree nature of knowledge, and you talked about the 10-15 year lifecycle of a pharma project. And one of the issues is having the pieces fit together. That reminded me into the "self-similar nature" of things. Preference for semantic database versus database. Similar when you are checking for safety, seems different from 15 years later. The whole process seems fractal.
- Otto: To me knowledge is a degree of uncertainty or certainty. We can work the concepts for which we don't have definitions. We can see the parallels. You could see an image and, if you are not trained, you might see nerve cells. But it could be an image of the galaxy. So many geometric forms go from small to large systems.
- Lynda: You have different reasons for looking at the same data during the Pharm cycle of 15 years.
- Otto: The focus is distributed. All of the decisions are based on some models. We "are" mental models. Some mathematician. We don't have models; models have us.
- Kate: Have you considered using "real options"?
- Otto: When I'm modeling the company in the context of the industry I do this. For example (metaphor) we have a metabolic study that feeds forward info, and feedback information and money with which you buy more inputs and resources. Another metaphor is the marketplace – projects find their way in the market place. The project may be in phase 1, then, you ask what's the value of it. You have to assume that it will get to the market and will generate revenue. It may branch out and there could be many possibilities, e.g., market success, reimbursement. Then when you discount back to the present for uncertainty. Important here is to approximate the decision-maker. If it's a bad project (e.g., future failure, because it's toxic, or inferior to something on the market), then the real option value is low. The drug discovery and development projects are information designs. You have a large structure space of possibilities and you prune that. You need to use the whole context including the future to approximate the value. I in effect use computational biology on our own company. Scientists are economists, but they are not used to those terms. They have to put their equipment and minds to best use when they have lots of options. They have to optimize the net gain for their experiments.
- David Ritter: Some people postulate that Pharma is in trouble. BCG said the return on investment for Pharma is .4 (40 cents to the dollar). You're talking about a way to discuss Pharma more holistically. Can you change the investment paradigm?

- Otto: ROI depends on how you frame it. There are many indicators that this is not such a losing game. Investment into R&D even in today's dollars... there are fewer and fewer drugs being introduced. It would be fine to launch one a decade if it were good.
- David Ritter: Need better decision-making.
- Otto: Yea, BCG.
- Lynda: Are smaller R&D companies better positioned than when you have a monolithic company? Competition for face time, human resources in big companies.
- Otto: Perhaps but good results are still positioned to deliver important products. We live longer, and there are definitely good results with outcomes for society coming out of biotech and pharma.
- Dan 2: You talked about prioritizing ideas in response to struggling productivity. Who does this? Can it be better?
- Otto: It's like politics. That's what brought me to the problem of modeling the decision-making. There are bit opportunities and they are well protected. There are no hard facts, just theories.
- Phil: Is there low-hanging fruit in the knowledge base?
- Otto: Every time you find a new perspective or new tool. Computation is a form of knowledge management. It's like a telescope. When we build new devices we get new ways. I heard a lecture by Andrew Ng at Stanford. He developed algorithms on deep learning. Inspired by the way the brain works. If the brain uses sparse coding, it is selective of the features computed. You can learn and discriminate. The key is to recognize and identify the features. For example, an algorithm can help longstanding disciplines. If you improve by a few percents, you get a paper. (If you feed something with signals to something that compresses and reconstructs it consistently.) A neural network invents these features. These are new developments and they are very significant!
- Kate: Do you use system dynamics?
- Otto: I develop a platform to simulate how things happen in worlds. For example states of an R&D project. I'm using probabilistic system dynamics alongside discrete event simulation. This is impressive, but not the way to represent the world.
- Dan Corwin: You can Google Andrew Ng. Machine learning class is free.
- Geoff: I was struck by your comment about relational databases back when storage was expensive. Now we can have redundancy with a lot of metadata. What kinds of data stores do we need to be focusing on? How does that change the types of problems you can address?
- Otto: You cannot have the metadata in the repository because you cannot anticipate the future. You don't want to close it – other people will combine it with what they have. So, if you are aware of the universality, then you can code the initial state, and it can code for future situations. You can move from relations to sentences. You can do it in a fluid way.
- Geoff: In order to get to these flexible environments, do we come up with new types of data stores? When you scratch away the hidden ideas, then you just put it into data stores. What about the new mongo db(?) formats?
- Otto: You have to ask – is it updating or editing? Data store technologies make tradeoffs. You can't have an optimal system for all options.

- Dan Corwin: We just had IBM's Watson play Jeopardy last year. We know that they break docs down into triples. Problem is that they didn't use RDF technology. The ontology was more important than the data store itself. I think this is a clue that something like that wouldn't be a bad way to represent enormous information from enormous information sources.
- Lynda: Watson was using conventional language. Genomics is more complex.
- Otto: You could find situations that could go either way.
- Kate: What about the use of these tools for military intelligence?
- Otto: Information fusion in law enforcement e.g., good traffic vs bad traffic in cyberspace isn't dissimilar from finding disease versus health. You'd think it's easy to find the genes for a disease for pathways or proteins or organ shapes (?), but it turns out that the more spaces you explore the more exponentially large. But it could also be that the combination of signals diminishes the size. If you look at incidents of crime and you see how people move, it's complicated. But if you find that someone was at two crime scenes and they had been communicating. It's similar with the genes.
- Phil: So you need to look at the data with a context. They have used humans before.
- Lynda: Too much volume.
- Otto: 20 years ago the mission to have Chess grand masters compete with computers was a no contest, now computing has the capability to compete strongly.
- Dan: ___ to provide a vehicle to aggregate these cross-pieces. Vioxx was pulled after launch because the data were later aggregated.
- Otto: Similar for drugs that are now being reused for other purposes. New things are being re-classified.
- Dan (InnoCentive): What's on the top of the wish list? For example, in your org., how do people find each other if they don't know each other exist?
- Otto: My wish list? It would be helpful to accept the value of information. We know how to quantify things, but few people recognize that. If people could recognize the value of information related to their goal, then you would have a more efficient self-organized market. It's an issue of culture. If we had a better currency for selling and buying information, this could be more productive.
- Lynda: How about the education system? Maybe we should be starting at a younger age to teach people to appreciate information.
- Otto: Some people say the education system destroys kids learning. Paul Lockhart's Lament (<http://www.maa.org/devlin/LockhartsLament.pdf>) went to teach math and felt it was taught incorrectly. It's an art form. How do you articulate a problem and see what parts are valuable? If someone simply just matches to a pattern, and it's rote manipulation, then there is no story.
- Lynda: Right. You are not teaching thinking.
- Otto: Picasso said, "Art is a lie that helps you to better understand the truths of the world." Similar to models. Maps are similar. They are wrong. But still maps are very useful.
- Otto: I see science as a discovery, but the engineering (and math) are an art forum. You discover through [...]
- Lynda: What types of persons do you like to work with?

- Otto: Someone not dogmatic. Math is about simplification. [Talk about semantic processing and triple stores – grammatical representation: <http://en.wikipedia.org/wiki/Triplestore>)
- Dan x: We learned that the genome was the answer to all our needs. Then we learned more. Is this the missing knowledge?
- Otto: We are too analytical and too specialized. We are not good at mixing things together. The brain needs to use both hemispheres. What's fascinating for me is the junk DNA that was thought to be the evolutionary junk. The insight is that the genes are like blocks, but the blue prints for putting it into the "pirate ship" aren't simply in the sequence. It's the noncoding parts in the short RNA molecules that combine and conduct, and orchestrate the processes. They are what we've found to be the more important genome information.

Lynda: Thank you (Otto). Usually our speakers come with the anticipation that they will learn something new. I hope our questions gave Otto some new ideas for his work because we're not all alike in the way we think. That's what's fun about KM.