Richard Beckwith, Intel

What is to be done?

This document will propose that ethnographic and other qualitative methods can be useful in establishing metrics for selection, analysis, and evaluation of ubiquitous computing systems. After summarizing the work we've done in this field, we will end with some specific recommendations for further research.

We do application-led research because we believe that this is the quickest route to understanding how ubiquitous computing -- and allied its research -- must develop. In answer to the question "What do we need to do to enable ubiquitous computing?"; we say that application-led research is a partial answer. And we do need to know how to enable ubiquitous computing if ON World, Inc.'s prediction of a seven-billion dollar market for wireless sensor networks in 2010 is to happen.

The question framing this workshop is different, though. "What makes for good application-led research in ubiquitous computing?" This is a more difficult question. First, application-led research encompasses research across a broad temporal span ranging from brief concept studies to longer-term trial deployments. We'll restrict comments to the latter since we believe that long-term iterative design is essential. Our multiple year research program with wireless sensor networks in agriculture will be offered as an example. Despite all this work, we will not have discovered how to enable ubiquitous computing until we have gone much further. For example, real, useful deployments will not occur until we have solved issues related to model building and sampling frequencies (to be detailed below).

Simply focusing on long-term iterative design is not enough. Success depends upon research that pushes up against and must conform to the real world phenomena associated with our application. We must, for example, understand the domain of our application. "Good metrics for selection, analysis, and evaluation of ubicomp applications" follow from that understanding. "Good approaches to longer-term iterative design in which applications are refined and scope expanded" require working with and paying attention to the people on whom our emerging technologies are about to emerge.

In some ways, the problem of "what leads to good application led research" is reminiscent of Plato's Meno dialog where Meno asks Aristotle how he can inquire into something he does not know, what should be the subject of inquiry, and how will he know that he's discovered what he set out to learn. Plato goes on to claim that we can do the inquiry if we start with beliefs that we can then evaluate. The point to be made here is that these beliefs should come from good qualitative work in the field and the evaluation should follow from field trials. To that end, I will offer a couple of general points that come from my group's application-led research in wine grape-growing.

What should be the subject of the inquiry?

Metrics for selection. Our research began before we selected the subject of our inquiry. We knew we wanted to look at sensor networks. We looked for a domain where a sensor network would make sense. We chose agriculture because it seemed a likely area of early adoption. In agriculture, sensors are common and sensible data are recognized as central to the enterprise. Focusing on the domain, allowed us to constrain the technological solutions that we might entertain, but understanding the domain was paramount for doing this well.

How do we inquire into something we do not know?

Ethnographic Methods. We used ethnographic methods to determine the kinds of deployments that would be attractive to our potential users. We began with the simplest methods. We spoke with the people working in this area. We began with semi-structured interviews of various people in the wine production value chain. We spoke to vineyard owners, vineyard managers, wine makers and their assistants, wine marketers, wholesalers, and retailers. These interviews addressed not only day-to-day activities but also the economics of their end of the business. This allowed us to focus on areas where there might be potential ROI from a deployment. On the basis of the interviews we identified a large number of potential applications ranging from tracking the work done in the vineyard (e.g., spraying or leaf-stripping) to monitoring conditions of the finished product as it was shipped. In each of these applications we were able to establish the parameters that would be needed for analysis and evaluation of the performance of a deployed system. In addition to this, we were also able to consider the form factors that we could easily deploy for a "deep dive" and selected one sub-domain for our trial deployment. We selected the agricultural side because this was one area that we thought would be shared with many other potential applications.

Participant Observation. Despite our decision to look at the agricultural side, we still needed to know just what we would do. We began participant observation, where we worked alongside those who would be most directly in contact with the technologies that we had in mind. We were involved in various aspects of fieldwork before harvest, harvest, crush, and cellar maintenance. Following this phase of the research, we decided to put sensors in the field to monitor climate and, we hoped, improve fruit quality and/or control over fruit quality for the winemaker. This was an application that was of interest to people throughout the value chain and we hoped would transfer easily to other crops.

Throughout these early phases of the research, it became increasingly clear to us that the kind of data that we could use a sensor network to collect had never been available to these practitioners before. This lack of history with such data had an interesting consequence. Some were interested to see what it could do for them, others doubted that it could add much – the research was just not in. Now we saw that we had to do the research to prove the value of the data.

Trial deployment as participant observation. Normally, a trial deployment would not be considered participant observation. Who would we be observing but ourselves? However, in this case, we needed to work with domain scientists so that we could establish the value of the data. So, as we came closer to understanding what we could do, we started working more directly with domain scientists. In fact, we sought out working scientists (notably, a PhD in plant physiology with a focus on wine grapes) to further refine the application area and make our data useful.

This phase of the research was in some ways the most exciting. We really didn't know what would work and only hoped that our bet would pay off. In the end, we found that we were able to predict a significant amount of the variability of the fruit in the vineyard we monitored. We were able to predict variation in pH, titratable acids, and berry weight. We found that we could predict boundaries for frost damage (and measured parameters for some kinds of damage). We were also able to define areas that would be amenable to growing more valuable crops. We could have done none of this without the input of a scientist working in the field.

We are still working on ways to make these data useful by developing different visualization that allow people with different interests to consume the data in a form they find useful. Still, there's much more to do.

How do we know we've discovered what we set out to learn?

Model Building. In the course of our research, we haven't discovered all that we need to do to enable ubiquitous computing. We didn't learn the answer but we did learn how to get closer to it. We have been able to determine some very specific lacks in the research that we've done. One of the most interesting aspects of working with domain scientists was that we finally really understood what it meant that no one had ever collected this kind of data before. The working scientists finally drove home the notion by explaining that it wasn't clear that finer grained data would be relevant because the existing models of how plants respond to climate used a very loose measurement for climate. Climate was presumed to be homogeneous and could be measured from one point in space, that is, one measurement with no variation. Practitioners and scientists were used to single measures with no variation. Our sensor network offered a measure of the variation in climate with variation.

Consider the way in which extant models have been developed. In order to represent the population of plants in a test plot, agricultural scientists have worked hard to ensure that they have a broad sample of plant products. They use various methods of plant sampling ranging from random to stratified. This ensures that they have the full breadth of variation with which to characterize the crop. On the climate side (where we were introducing technology), scientists had always relied on a small number of measurements to characterize the area the plant products were sampled from. In fact, they usually relied on just one measurement. That is, they had variation in the plants but not the climate. This meant that models predicted a broad range of plant variation that could be expected in a particular climatic situation. These models would hold that our variance in climate could not predict variation in compositional chemistry because the variance in the plants is so high. (Of course, if we believed those models, we would not have gone on with this research.) Which is to say, the models to really support this work will, in many cases, be non-existent. The work that we did started to address the issue of model development but did not go far enough. Two years of data are the norm in agricultural research before findings are even considered publishable. What does this kind of timing mean for the model development that would have to be completed for ON World's seven billion dollar market to be a reality in 2010? We have to develop those models and there isn't much time.

Following from this dearth of models we have a related problem. How finely grained shall we sample our variables?

Nyquist Frequencies. Optimal sampling rates are not well understood. At which spatial frequency should we sample? At which temporal frequency should we sample? The simplest way of describing this is that we do not know the spatial or temporal Nyquist frequencies for the domains in which we want to see deployments. This may be more significant than appears at first blush. In our experience, we have found that different services require different densities of measurement. However, you can't know what the optimum density for any service might be without over-sampling and then looking to see how sparse your sampling could have been while at the same time still reflecting the appropriate level of variance.

Data for each of these issues can be presented at the workshop.

Summary

When we need to understand our application domain enough that we can develop metrics for selection, analysis, and evaluation of ubicomp applications, we believe that ethnographic methods can be reasonable tools. Having set these metrics we can hope to understand when we have learned what we have set out to learn. However, we will add at this juncture that (at least in the case of wireless sensor networks) we will not be done before we have determined exactly how dense our infrastructure needs to be or have built models to support the analysis of the data we have collected.