



Connecting
People, Technology and Knowledge
for Agricultural Innovation



Ubiquitous networks and cloud computing – May 10th 2009

Enrica M. Porcari
Chief Information Officer, CGIAR

Abstract

Major changes are in progress in Internet-based computing which will continue for years to come. These changes will offer new potentials for agriculture and agricultural research in developing countries. In this paper we consider: (1) the spread of public wireless data networks, which enable gathering data from sensors and distributing information to rural farmers; and (2) the emergence of "cloud computing," which enables inexpensive processing of massive datasets by any Internet user, lowering the institutional capacity required to participate in research. Leading institutions can act now to accelerate the adoption of these changes in agriculture research.

Introduction

Information technology (IT) continues to advance at a relentless pace, doubling the speed and capacity of computing, storage and communications equipment every couple of years. This progress manifests primarily as steadily dropping costs for computers, storage, and communications, with a gradual spread of IT into new applications as prices fall. From time to time, however, discontinuous changes occur. Sometimes this is due to new software that employs computing power in new or unexpected ways. Sometimes it comes from an accumulation of investment that passes thresholds to enable new applications. This paper examines two of these discontinuous shifts that have already started, and that are certain to have a large impact on agricultural practice and research around the world. References to real developing country applications of these innovations can be found in the reference section of this paper.

Technology Shift 1: Ubiquitous Telecommunications Infrastructure

Thanks to the falling costs of all things digital, there has been a steady flow of investment into communications infrastructure around the world. Cell phone networks carrying voice and Internet data are being deployed in even the poorest countries and with time will expand to cover most rural areas. These wireless networks are sophisticated and easily managed. Multi-purpose public networks will be offered by private telecomm companies and governmental agencies, while self-organizing device networks (such as ZigBee, a low-cost, low-power, wireless [mesh networking](#) standard) can be installed with minimal planning or oversight. Agriculture and agricultural research can increasingly take communications capacity for granted in the years ahead.



Connecting **People, Technology and Knowledge** for Agricultural Innovation



This new infrastructure will enable new applications of communications to both the gathering and dissemination of information by agricultural researchers and practitioners. First, for gathering information, the historical and remotely-sensed data that has been gathered to date can be complemented by near-real time, ground-based data. Sensors can transmit the information they detect through increasingly ubiquitous wireless data networks into Internet-based servers. Radio-frequency identification (RFID) tags can be attached to vehicles, buildings, and selected goods; combined with Geographic Positioning System (GPS) information, objects can be automatically tracked and even audited in real time. The result will be both real-time interpretation of current conditions and longitudinal analyses that reflect up-to-date information. The costs of the sensors, tags, GPS and RFID devices and the communications between them are dropping so rapidly that new data-gathering applications can be expected to proliferate in the near term. Here are some relevant examples:

- Sensors and cameras in fields or on farm equipment
- Sensors of water levels in irrigation or in soils
- Sensors in food storage
- Early detection of pests
- Emissions sensors
- Tagging of livestock
- Tagging of other natural resources
- Tagging trucks and shipping containers
- Market, banking, and distribution data

Like satellite imagery, these new types of data will require considerable processing to ensure their quality and consistency, and to make them comparable from one location to the next. The research community will need to establish processes for validation and distribution of these data, as they have with other public information goods.

The same networks that collect and carry sensor data will also be used to disseminate information into rural areas. Cell phones are already being used at an increasing rate by rural residents. For them, the value of communication is high, and there are many ways to effectively share the fixed costs of phone devices and electric power among numbers of users. As phones get larger screens, touch interfaces, and voice recognition, and as new classes of inexpensive and rugged "netbooks" are developed, many new opportunities for agricultural extension will arise. It will start by providing today's information to new audiences. It will grow into provision of new services that are more localized and more up-to-date, building on the data gathering that is enabled by the networks that feed these devices. With new audiences and new services will come new requirements for assuring the quality of the information provided.

Technology Shift 2: "Cloud" Computing



Connecting **People, Technology and Knowledge** for Agricultural Innovation



The combination of progress in system software, computing hardware, and Internet communications has now enabled the construction of general-purpose data centers that can be reconfigured by command to support any software application in minutes. ("Virtualization" software was the key innovation.) There are already data services that allow a user to have many hundreds of computers at their command, and yet pay for them by the hour or minute, without owning or operating the hardware themselves. The costs are far less than even falling hardware prices would suggest, since the cost of the data center can be shared among many "bursty" users. In effect, the data center acts like a utility, providing as much computing as requested at just the times when needed. Since these data centers are invariably shared over the Internet, they are sometimes called computing "in the cloud," giving rise to the common term "cloud computing."

A shared cloud data center will typically have over 1000 computers, which can support at least 100,000 user "virtual" computers. This is super-computer scale by any standard, so most research centers will not own one but rather will share one with hundreds of other customers. Commercial "cloud providers" like Amazon, Google, and Microsoft already offer service, and some government-run research clouds exist. Shared by many thousands of customers, these are extremely cost-efficient. They employ a relatively small staff of system managers, keep a low budget for electric power, can survive routine equipment failure without service interruption, and adopt continuous modular upgrades of new types of hardware. There are choices in many countries, which allow for flexibility where there are legal restrictions.

Many observers believe that cloud computing will soon be the lowest-cost option for nearly all types of data center computing. Cloud providers are already more cost-effective for "bursty" high-performance computing, like video and image processing, bioinformatics, and most types of scientific data analysis. We can expect research centers in agriculture to have accounts on several cloud providers, and to select them at different times for different purposes.

The shift to cloud computing is a good thing for today's researchers, by cutting the total cost of scientific computing. But it also brings two new opportunities for international agriculture. First, it completely separates the utilization from the operation of computing facilities. In other words, users of data centers no longer need the capacity to procure and operate them. As long as one has a browser on the Internet, one can "order up" essentially any computer software at any scale, and pay only for what is used. As a result, many more organizations will be able to take advantage of large-scale advanced computing.

A second implication of cloud computing is an increased impetus to share data among researchers. It is a common pattern today to move large data sets, such as satellite images or longitudinal data sets, from one data center to another for use in different



Connecting **People, Technology and Knowledge** for Agricultural Innovation



projects. The transfers add delay and can be error prone. By contrast, cloud data centers are a natural repository for public information goods like shared data sets, so that users in any location or institution can instantly access, analyze and interpret data without the need to move it to their own facilities. This reduces the need for high-speed or high-capacity network connections, since much less data moves between the users and the source of the data. A researcher with a moderate-speed connection to the Internet can work with data as well as other researchers regardless of location. In addition, researchers will normally leave the results of a cloud analysis at the cloud data center, allowing potential re-use by others. Properly managed, this can enable new kinds of collaboration and project organization.

Implications

Leading institutions in agricultural research have an opportunity to flesh out these possibilities today, and thereby create templates for future models of progress. Here is one illustration.

A research center that works with a crop could choose a group of similar varieties that have been cultivated for a long period at one of their facilities. The center will already have basic long-term data across many seasons, along with much bio-informatic data. These data could be stored in one or more cloud computing facilities, and could be supplemented from now on by extensive sensor data, collected and made available in near real-time from the fields where the varieties are cultivated. In effect, these fields become a "bio-observatory" for those varieties. In addition, one or more regions where those varieties are currently cultivated by farmers could also be instrumented with some sensors, and the markets in those regions could employ tags or other methods for continuous data collection. Once a data collection like this is available in a cloud data center, a series of analytical studies could be commissioned at various developing-country institutions around the world. These institutions would be chosen for having familiarity with the varieties but currently lacking the facilities to do their own extensive data analysis or interpretation. In addition, adaptation and extension projects could be commissioned at additional national organizations to produce materials for delivery into the areas where these varieties are grown.

Like any collaborative research, this kind of project would have to confront issues of data harmonization, accessibility, and ownership. Part of the value of this project would be the demonstration of solutions to these issues, as a pattern for future projects to follow.

Naturally, this entire scenario could be adapted in many ways to the other agricultural research topics. For example, a project could treat livestock instead of crops, or could extend a system like Fishbase for a class of fish. Genetic studies of crop pathogens, patterns of water supply and utilization in a watershed, and forest growth and



Connecting People, Technology and Knowledge for Agricultural Innovation



production patterns, all lend themselves to this sort of project. There will be limitations to the effectiveness of any single project; but the first projects are likely to provide key lessons to light the way for the research community in utilizing the next wave of technological changes.

These technological shifts are opportunities to “turbocharge” our research efforts to help smallholder farmers. Opportunities that the agriculture research must not miss!

References to background information

The leading cloud computing provider is Amazon Web Services, with its “Elastic Compute Cloud” or EC2 service (<http://aws.amazon.com>). One example of cloud-resident data sets is GenBank, currently 250 gigabytes in size, described as the “annotated collection of all publicly available DNA sequences including more than 85.7B bases and 82.8M sequence records” (<http://developer.amazonwebservices.com/connect/entry.jspa?externalID=2261&categoryID=246>).

Sensors and RFID tags in agriculture have been covered by CTA and other organizations. Examples:

- RFID tags on livestock: <http://ictupdate.cta.int/en/Feature-Articles/LITS-tracking-Botswana-s-livestock-using-radio-waves> and http://www.idtechex.com/research/reports/rfid_food_and_livestock_case_studies_000131.asp
- The use of sensors and computer interpretation to track localized CO2 levels: <http://cleantech.com/news/4268/forget-carbon-emissions-haymet>
- Computers in “precision agriculture,” combining remote and local sense data with farm equipment: http://www.freshplaza.com/news_detail.asp?id=41931 and <http://ictupdate.cta.int/en/Feature-Articles/Farming-from-space-precision-agriculture-in-Sudan>

Biological observatories, using intensive sensor-based data collection and real-time data dissemination, were first proposed for biodiverse regions. One example was the US Ecological Observatory Network, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=13440&org=DBI.

A general review of the promise of cloud computing for developing country research can be found at <http://www.newsweek.com/id/195734>