Open data: enabling food systems transformation

Robin Lougee, IBM Research
IBM Research: the world is our lab
Deep Thunder and the Flint River Partnership

In Georgia, where agriculture has a $72 billion economic impact, farmers are turning to ground-breaking technology from IBM to help meet ever-increasing food production demands and leading the way in conservation measures, improving agricultural efficiency by up to 20 percent.

Published April 23, 2014

Deep Thunder and the Flint River Partnership

In Georgia, where agriculture has a $72 billion economic impact, farmers are turning to groundbreaking technology from IBM to help meet ever-increasing food production demands and leading the way in conservation measures, improving agricultural efficiency by up to 20 percent.

Published April 23, 2014

Open Data used for US applications includes

- NOAA/MADIS near-real-time surface and upper-air observations
- NASA high-resolution (2km) sea surface temperatures (SST), which include Lake Surface Temperature (LST) analysis over the Great Lakes
- NASA high-resolution (90m) Shuttle Radar Topography Mission (SRTM) terrain elevation
- NASA MODIS 1km 20-category land use data
- NASA 4km dynamic (daily) VIRIS Green Vegetation Fraction (GVF) data
- NASA 3km land surface fields for initialization
- NOAA/NCEP Rapid Refresh (RAP) 13km analysis for background fields
- NOAA/NCEP North American Model (NAM) 12km or Global Forecast System (GFS) 25km) for lateral boundary conditions
Foodborne illness is a global issue

**USA**

1/6 of the population had food-related illnesses in 2011, causing 50 million illnesses, 3000 deaths and costing $80 billion

Source: 2012 The State of Food Insecurity in the World... FAO, IFAD and WFP; collateral information from WHO; CDC Morbidity and Mortality Weekly Report June 10 2011; FAO Global Food Waste 2011

**UK**

Estimated 1 million cases of foodborne illness each year, resulting in 20,000 hospital admissions and 500 deaths


**Africa**

2,000 people die in Africa each day from food safety related illnesses. Sub-Saharan Africa loses estimated $4 billion each year in post harvest grain losses, which could feed 48 million people


**Australia**

The cost of the estimated 11,500 cases of food poisoning daily in Australia was calculated at AU $2.6 billion annually, or roughly 2B USD.

Source: WHO, 2008
Food contamination presents a significant threat to food security and sustainability

**Human and Health Impacts**
- 25% of food crops are aflatoxin contaminated
- Mycotoxins, aflatoxins and other microbial threats affect key crops consumed by hundreds of millions of malnourished people
- Diarrhea related to food borne infections is a leading killer of adults and children (2 million annually…more than TB, HIV/AIDS and malaria)

**Economic and Social Impacts**
- Small family Farmers represent 500 million of 570 million farms; 1.5 billion people
- Accounts for almost 60% of agriculture production
- Future food security progress is linked to family farms
- Face multiple challenges to grow and deliver food
- Contamination reduces the supply and value of food globally
- 40% loss developing countries at post harvest and processing
- Globally ~ 1/3 of food (1.3 billion tons) is lost or wasted
- Degrades food quality
- Lowers crop values, reduces farmer’s income, restricts trade opportunities
- Restricts sourcing local commodities for nutritional purposes
- Hinders the ability of food companies to procure local food ingredients

Sources: FOA; iel Kertesz WHO; Liu & Wu 2010; IFPRIForum Brief # 9 2013; IFAD website 2014
Consortium for Sequencing the Food Supply Chain

IBM Research and MARS Corporation tackle global health with food safety partnership

1 in 6
Americans each year suffer a food borne disease (CDC)

2 Million
Annual deaths in emerging areas due to food borne infections

3,000
Annual food borne disease deaths in U.S. (CDC)

$80 Billion
Annual cost of losses and illness caused by food borne disease

Surveillance, risk assessment, and diagnosis of food borne pathogens across the food supply chain using a scientific approach

Metagenomics for food safety

With DNA and RNA sequencing, we are able to profile communities of microorganisms—the microbiomes—in the supply chain anywhere along the process from farm to table.

Consortium for Sequencing the Food Supply Chain

IBM Research and MARS Corporation tackle global health with food safety partnership

1 in 6
Americans each year suffer from food borne disease (CDC)

3,000
Annual food borne disease deaths in U.S. (CDC)

2 Million
Annual deaths in emerging areas due to food borne infections

$80 Billion
Annual cost of losses and illness caused by food borne disease

Surveillance, risk assessment, and diagnosis of food borne pathogens across the food supply chain using a scientific approach

Open Data used includes

- Ensembl
- NCBI NT nucleotide and NR protein data bases
- 100K Foodborne Pathogen Genome Project

With DNA and RNA sequencing, we are able to profile communities of microorganisms—the microbiomes—in the supply chain anywhere along the process from farm to table.

Solution for precision irrigation at E&J Gallo Winery

- Satellite images are analyzed to determine
  - greenness of vines (NDVI)
  - surface temperature
- Predictive weather model
- Weather station data
- Evapo-transpiration modeling
- **Control a smart variable rate irrigation system**

*Double drip line with control electronics*  
*Control electronics box and node*
After two years of precision irrigation, yield maps show significant improvements

- 26% more yield in the precision area
- 11$ higher water efficiency
- 50% higher uniformity
- Improved quality index (Brix value)
- $120M of annual value (based on 100K acres)

Gallo & IBM: 2014 Vintage Report Innovation Award for prototype system

20 Jan. 2015
After two years of precision irrigation, yield maps show significant improvements

- 26% more yield in the precision area
- 11$ higher water efficiency
- 50% higher uniformity
- Improved quality index (Brix value)
- $120M of annual value (based on 100K acres)

**Open Data used includes**
- Landsat satellite data
- CIMIS data

**Gallo & IBM: 2014 Vintage Report Innovation Award for prototype system**

20 Jan. 2015
PAIRS: A big data platform for scalable spatio-temporal data and analytics

- Data bus feeds in (near) real-time open spatio-temporal data into PAIRS (Physical Analytics Integrated Repository and Services)
- Full data curation process (e.g., filtering, classifying, aligning, resampling, reprojecting) at ingestion.
- Large-scale Hadoop / Hbase system for efficient distributed data store and processing
- System allows complex queries
  - "Find all real estate in California with elevation gradient and high rain fall and certain soil type."
- Access to new layers of analytics
  - Irrigation forecasts
  - Improved weather forecasts
- Curated data and analytics accessible as a Service via an integration layer
  - REST APIs to run queries
  - Basic web interface to run queries
PAIRS incorporates many open data layers

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Origin</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satellite</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landsat 7</td>
<td>USGS</td>
<td>16 days</td>
</tr>
<tr>
<td>Landsat 8</td>
<td>USGS</td>
<td>16 days</td>
</tr>
<tr>
<td>Modis /Vegetation Product</td>
<td>NASA</td>
<td>16 days</td>
</tr>
<tr>
<td>Modis Surface Reflectance</td>
<td>NASA</td>
<td>2 days</td>
</tr>
<tr>
<td><strong>Weather</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Forecasting System(GFS)</td>
<td>NOAA</td>
<td>Daily + 60h forecast</td>
</tr>
<tr>
<td>North American Mesoscale Forecast System (NAM)</td>
<td>NOAA</td>
<td>Daily +10 day forecast</td>
</tr>
<tr>
<td>European Centre for Medium Range Weather Forecast (ECMWF)</td>
<td>EU</td>
<td>Daily +10 day forecast</td>
</tr>
<tr>
<td>CIMIS data</td>
<td>State of California</td>
<td>Daily</td>
</tr>
<tr>
<td><strong>Survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Elevation Data</td>
<td>USGS</td>
<td>Static</td>
</tr>
<tr>
<td>Soil data (SSURGO)</td>
<td>USDA</td>
<td>Static</td>
</tr>
<tr>
<td>Land use (Cropscape, MODIS)</td>
<td>USDA</td>
<td>1 year</td>
</tr>
<tr>
<td><strong>Analytics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM Blended Forecast</td>
<td>IBM</td>
<td>Daily</td>
</tr>
<tr>
<td>IBM Evapo-transpiration/ Irrigation</td>
<td>IBM</td>
<td>Daily</td>
</tr>
</tbody>
</table>
Key Findings

“The need for more data scientists strongly emerged across discussions in the different working sessions.”

Source: Summary Report, p4
Master’s degree programs in Analytics and Data Science are growing

Cumulative number of programs in the U.S. by year of initial enrollment

Source: Institute for Advanced Analytics, North Carolina Statue University
Pro bono data science

Do Good with Data
The challenge is to determine how a farmer can make seed variety decisions that reliably reduce risk and increase yield.

Nurturing interest in agriculture analytics

The Syngenta team which won the INFORMS Edelman prize is using their prize money to encourage INFORMS members and others to tackle agriculture analytics challenges.

Source: https://www.ideaconnection.com/syngenta-crop-challenge/
Corporate Service Corps
A triple benefit
Communities have their problems solved
IBMers receive leadership training and development
IBM develops new markets and global leaders
Developing data science that benefits humanity

- For students, summer and postdoctoral fellowships
- For partner organizations, projects are conducted in collaboration with nonprofit partner organizations who put forth the problems and commit to using the results.

Open source for the operations research community

Incentives in a reputation economy
Thank you

rlougee@us.ibm.com
@RobinLougee