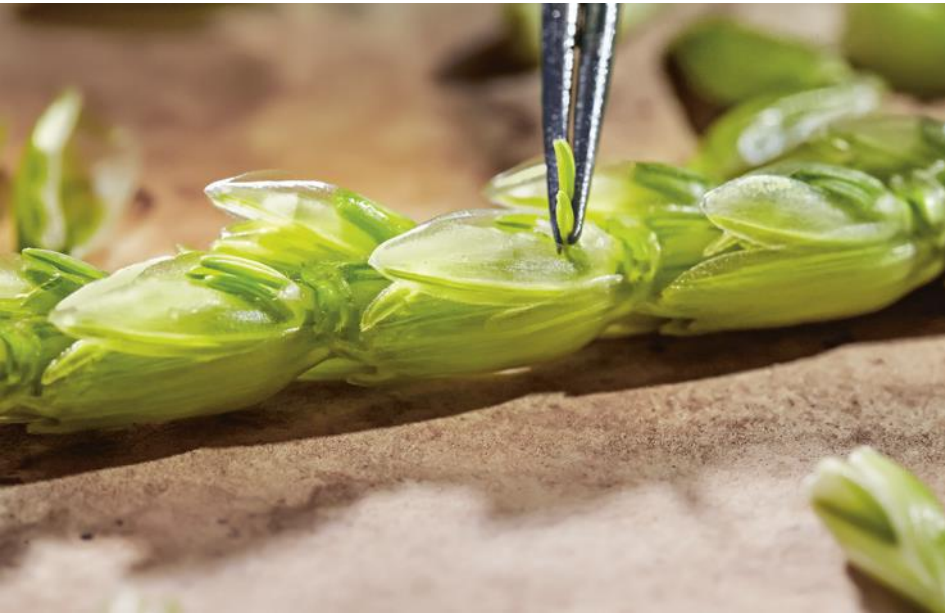




Science For A Better Life



Case Study in Geospatial Analytics: Building a Global Platform for Agro-Environmental Analysis

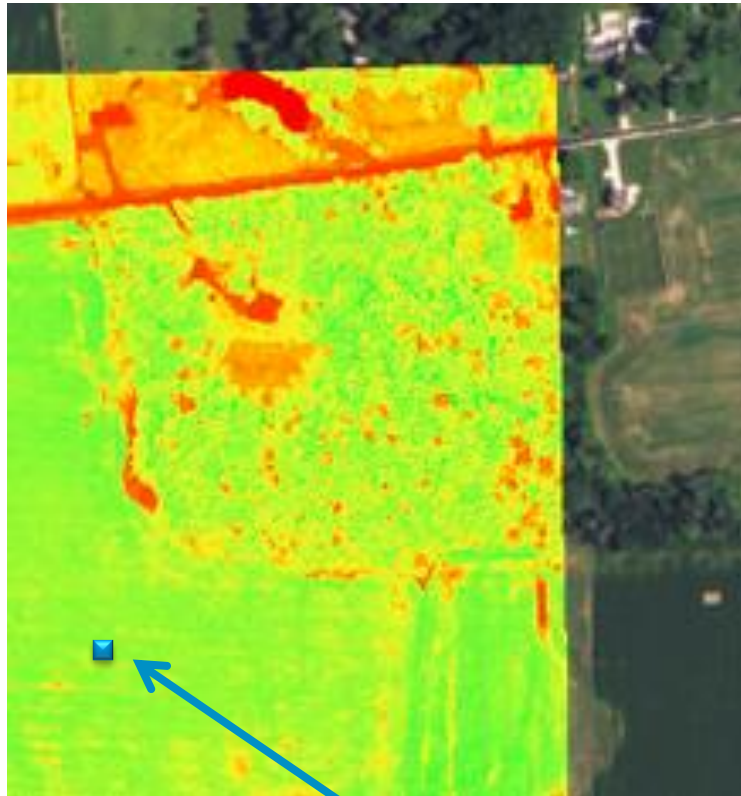
OGC Agriculture Domain Working Group 2016

09-22-2016 / Kris Matson / Version 2

© Bayer CropScience • September 2016

Bayer CropScience On-Farm Trials

Studying Real-World Product Response



30ft x 30ft research trial plot

Must understand crop protection product effects in real-world situations

- Efficacy trials on 30 X 30 foot plots do not easily translate to ½ mile X ½ mile fields
 - Not enough variability in the small plot
 - Large fields are not homogeneous
 - Practical conditions differ from R&D

Why answer questions on product response in the real world?

- Help growers to be more profitable – and grow more crops
- Help growers to be more sustainable – limit off-target effects

Helping growers to be more profitable and sustainable

Bayer CropScience On-Farm Trials

Challenges in trial data collection and management



Communicating and Managing Trial Protocols

- Originally communicated only at season start
- Insufficient data collection guidance
- Lack of visibility into protocol workflow & issues

Timely Collection of Protocol Data

- Bulk data egress at end-of-season → #epicfail

Missing Metadata

- Metadata is required for on-field activities as well as geospatial data

High-Variety & High-Volume Data

- Equipment, sensors, & FMIS software format all aggregate data differently

Result: Analysis and Modeling At-Scale is Difficult or Even Impossible

- **For analytics at-scale, standards are not optional**

- End of year 1 - complete data from only 50% of trials.
- Question: How to scale from 30 to 200 fields?
- Answer: **STANDARDS**

USB is a standard...
but what about the files?!



What We Are Building

A Field Trial Protocol Management System → FTPro



Software platform for the systematic collection, processing and analysis of trial data from commercial farm operations

Provides standardized data to visualization and analytic tools:

- To understand product performance and ROI
- Analyze multiple fields and seasons

The screenshot displays the Bayer FTPro software interface. On the left, a sidebar lists data collection methods: Imagery, Seeding, Applications, Soil Sampling, Scouting, Harvest, and Weather. The main area shows a satellite map of a field with a color-coded overlay representing trial data. On the right, a panel displays protocol details for 'WestSp Corn' starting on 04/01/16, and a 'Job Status' section with progress indicators for Preplanting, Planting, and other tasks. Below that, 'Current Conditions' are shown: Temperature 81.6 F (27.6 C), Wind 2.0 mph Southeast, Precipitation 0.0000 in, Leaf Wetness 15, and Cumulative ET 2.930 in (Month). A callout box on the right lists various analytics and modeling tools: R, ENVI, python, esri, tableau, and sas, all grouped under 'Ag-Enviro Models'.

Inspiration for this Presentation

Vision for Geospatial Analytics via Open Standards

Three hexagonal graphics arranged in a cluster. The top-left hexagon shows a satellite map of a landscape. The top-right hexagon shows a 3D topographic map. The bottom hexagon shows a computer monitor displaying a data visualization.

The Future of Geospatial Analytics through Open Standards

George Percivall
CTO, Chief Engineer
Open Geospatial Consortium
percivall@myogc.org



© 2016 Open Geospatial Consortium



OGC Future of Geo-Analytics



Lack of interoperability is a serious technical debt

Lessons from the success of Apache Spark...

interchange is necessary for the ecosystem

major use cases tend to build their own ML libraries – despite a case where a majority of committers tend to support a common vision and encourage use of a canonical library (**MLLib with DataFrames**)

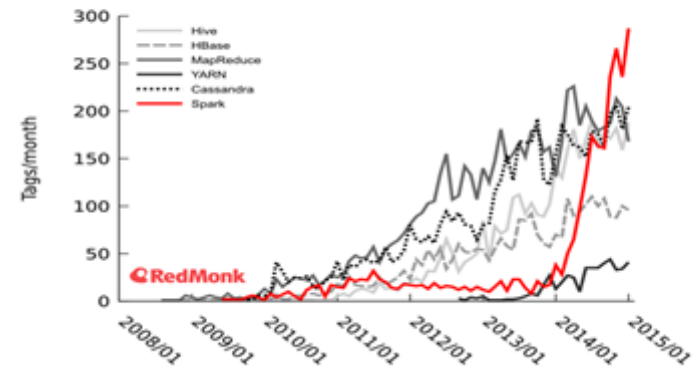
when a successful business grows over time, challenges arise by definition: managing separated teams, mergers and acquisitions, increased audits, regulations, etc.

therefore, lack of interchange for analytics represents a serious technical debt and potential liability



Source: "Use of standards and related issues in predictive analytics" KDD 2016, SF 2016-08-16 Paco Nathan, O'Reilly Media

Selected Big Data activity on Stack Overflow

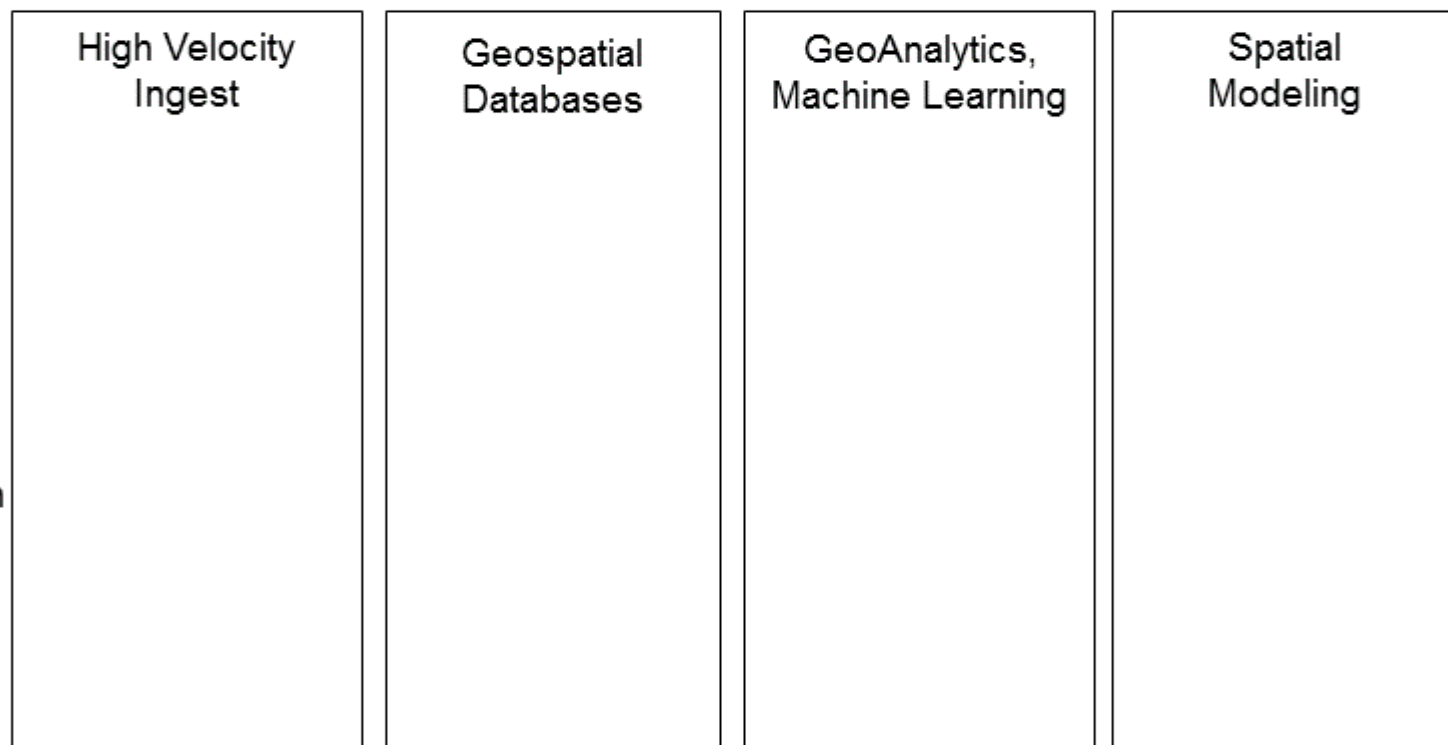


OGC Big Geo Data Analysis Use Case

Presented at ENVI Advanced Analytics Symposium



Big Geo Data




Observation Sources


Users and consuming apps

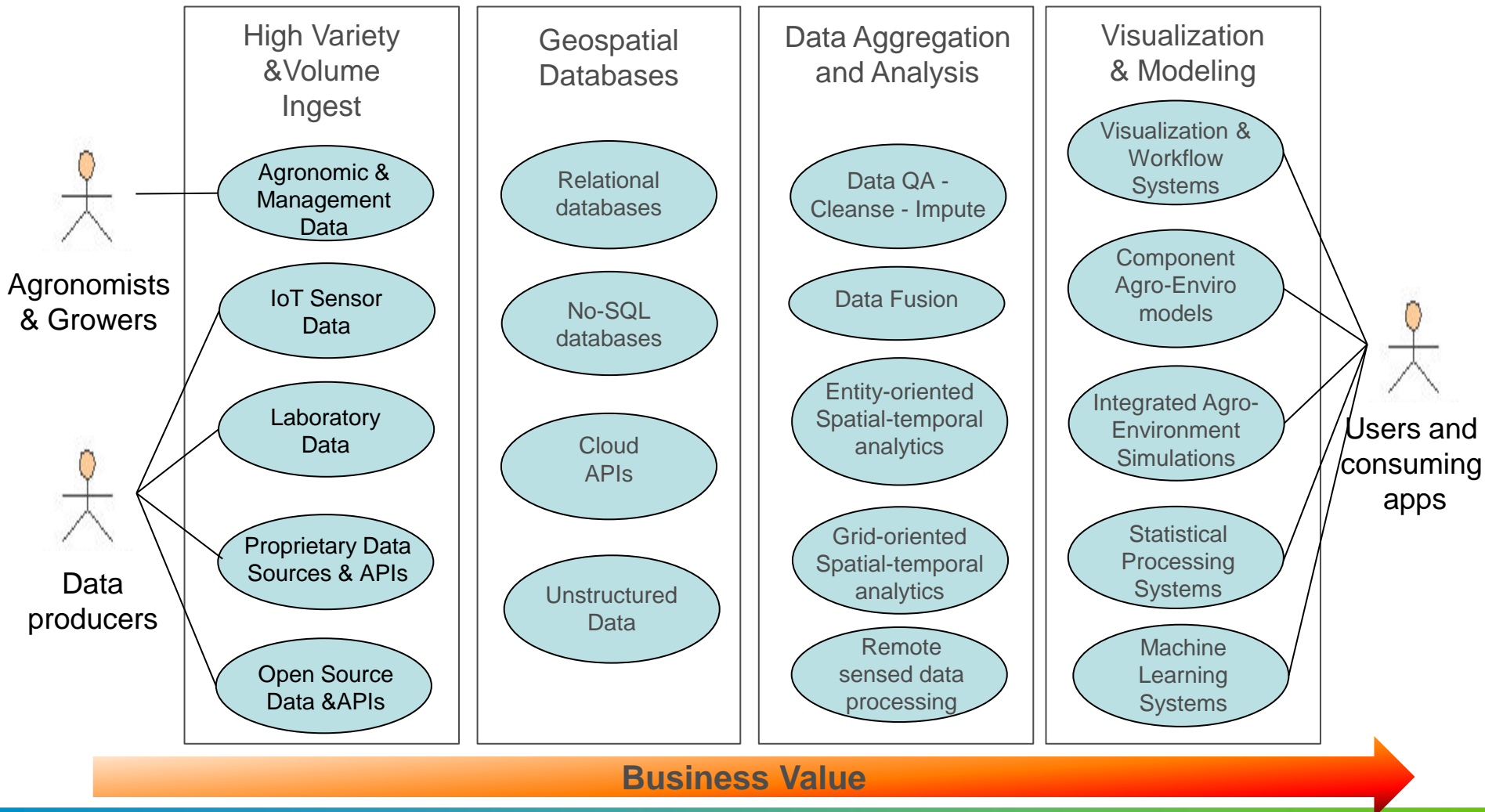


© 2016 Open Geospatial Consortium



Big Data Use Case for Ag R&D Trials

High-Variety & High-Volume Analytic Pipeline



Typical Field Data Sources

Agronomic, Management, and Spatial data



Agronomist + Grower Collected Data

- Field Boundary
- Field Scouting
- Soil Cores
- Soil Chemical Analysis
- Crop Tissue Samples
- ...

Public Sector Data

- Elevation
- Soil
- Landsat
- ...

IoT Equipment Generated Data

- Planting (Seeding)
- As-Applied Fertilizer
- As-Applied Herbicide
- As-Applied Pesticide
- As-Applied Insecticide
- Harvest (Yield)
- Weather Stations
- ...

Multi-Source Imagery

- UAS, Air-borne, Satellite
- NDVI and other derivative products



Agriculture IoT Data

Equipment data capture and aggregation



Planting



Spraying



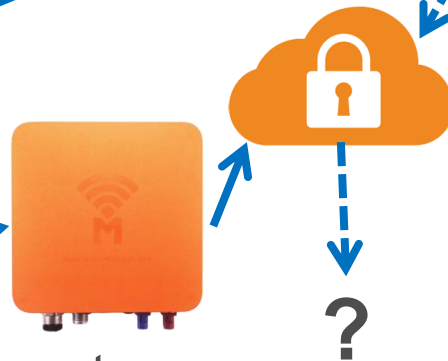
Harvest

Data Logging



Typical Ag equipment data issues:

- Was the seed variety entered into the Planter?
- Was the applied product entered?
- Was an applicator nozzle clogged?
- Was the harvester calibrated?
- **Data egress in cloud or as files?**



Agronomist & Grower Collected Data

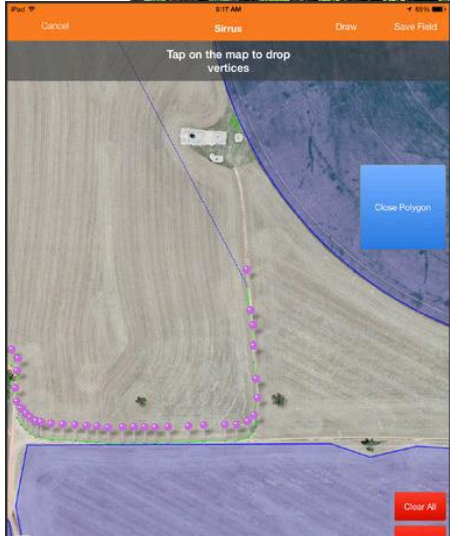
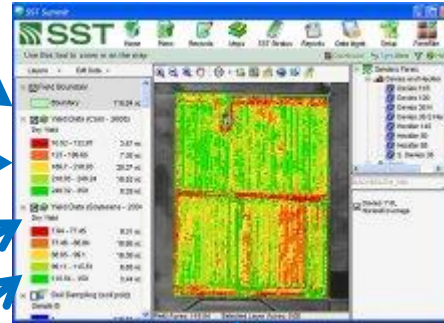
Varying sources, structure, aggregation and standards



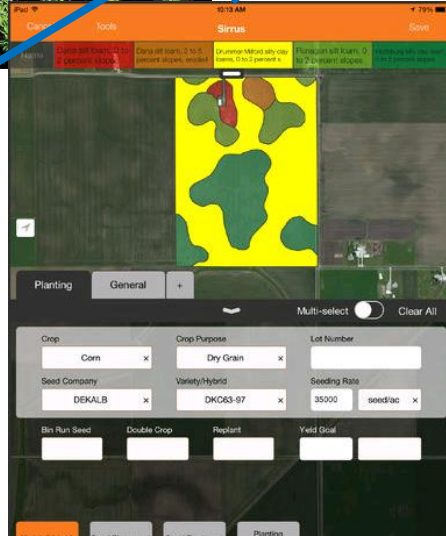
Field scouting



Farm management information system



Field digitizing



Management recommendations



shp, csv, ...

Scouting 2015					
4/11/15		Grower	Sample Grower		
		Farm	Sample Farm		
		Field	Sample Field		
		Area	9.91 ac		
Crop	Growth Stage Group	Growth Stage	Height	Crop Condition	
1.2	Wheat Hard Red Winter	Emerging/Small Grain	8+ Tiller	9 (0)	
Weeds	Density	Stage	Min./Max./Avg Count	Min./Max./Avg Height	
1	Panicgrass	Trace	All Stages	3, 7, 5 (plants)	2, 6, 4 (in)
1	Isome, Cheatgrass	Trace	Tillering	2, 6, 4 (plants)	2, 6, 4 (in)
Insects	Density	Stage	Min./Max./Avg Count	Location	
2	Culex, Army	Trace	All Stages	3, 7, 5 (sq ft)	Lower Leaves

Crop Protection				
4/12/15		Grower	Sample Grower	
		Farm	Sample Farm	
		Field	Sample Field	
		Area	9.91 ac	
Product Type	Product	Product Rate	Registration Code	
1	Herbicide	Roundup PowerMAX	32 (fluid oz/ac)	524-543
1	Insective	Amonium Sulfate	10 (lb/100 gal)	
Carrier	Total App Rate			
1	Water	15 (gfl/ac)		

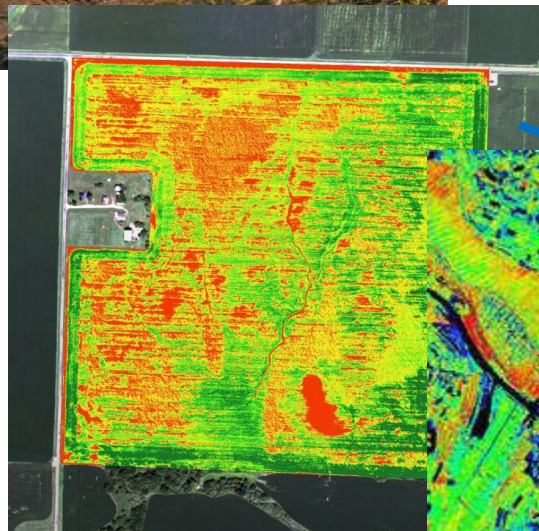
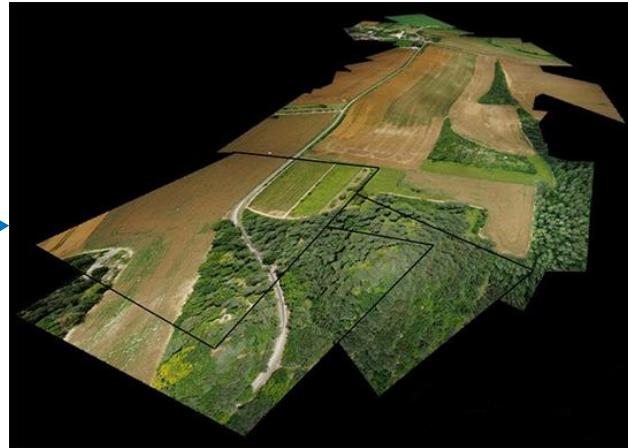
Report documents

Sensors and Multi-Source Imagery

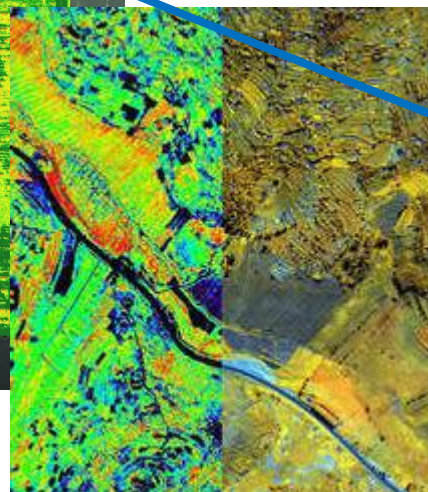
Covariate data capture and aggregation



Unmanned Air Systems

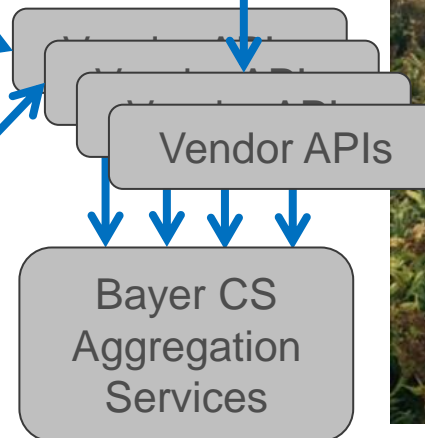


Airborne imagery



Satellite imagery

Weather and Moisture



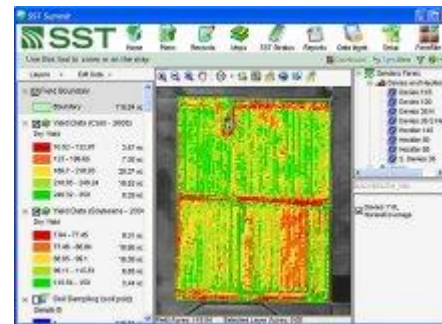
SST agX Cloud

Standards and AWS interface for field data



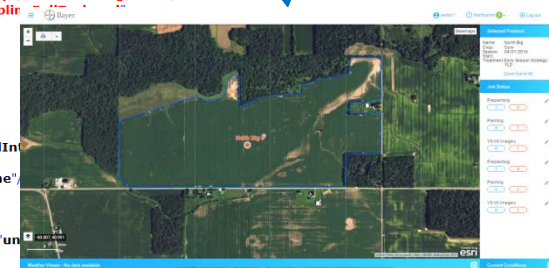
- **Standardized cloud interface for agriculture operations data**
 - Field-specific data payloads
 - Management data
 - Spatial data
- **XML encoding**
 - Robust schemas XSD encoded
 - WKT shape types for geometries
 - GeoTIFF for imagery
- **Practical standard supporting:**
 - Farm operations
 - System interoperability

Farm management information system



Soil sampling XSD

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema targetNamespace="http://www.sstsoftware.com/EDS/SoilSamplingFullTask.xsd" element
xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:xsd="http://www.w3.org/2001/XMLSche
version="1.2" xmlns="http://www.sstsoftware.com/EDS/SoilSamplir
<!-- Define the guid datatype -->
+ <xsd:simpleType name="GUID">
  <!-- define the format of a version string -->
+ <xsd:simpleType name="VersionStringType">
<!-- Root Element definition -->
- <xsd:element name="SoilSampleFull">
  - <xsd:complexType>
    - <xsd:sequence>
      <xsd:element name="SummitID" type="xsd:unsignedIn
      <xsd:element name="ParentEventID" type="GUID"/>
      <xsd:element name="ModifiedOn" type="xsd:dateTime"/>
      <xsd:element name="Records">
        - <xsd:complexType>
          - <xsd:sequence>
            - <xsd:element name="Record" maxOccurs="un
              - <xsd:complexType>
                - <xsd:sequence>
                  <xsd:element name="RecordNum" type="xsd:unsignedInt"/>
                  <xsd:element name="ZoneID" minOccurs="0">
                  + <xsd:element name="SampleID">
                  + <xsd:element name="TopsoilSamplingDepth" minOccurs="0">
                  + <xsd:element name="Soil_pH" minOccurs="0">
                  + <xsd:element name="Buffer_pH" minOccurs="0">
                  + <xsd:element name="ExcessLime" minOccurs="0">
                  + <xsd:element name="PercentCaCO3" minOccurs="0">
                  + <xsd:element name="H_meq" minOccurs="0">
                  + <xsd:element name="ExchangeAcidity" minOccurs="0">
                  + <xsd:element name="OM" minOccurs="0">
                  + <xsd:element name="PercentOC" minOccurs="0">
                  + <xsd:element name="PercentHumicMatter" minOccurs="0">
                  + <xsd:element name="ENR" minOccurs="0">
                  + <xsd:element name="TotalN" minOccurs="0">
                  + <xsd:element name="NO3_N" minOccurs="0">
```



Soil Sampling Task v 2.0

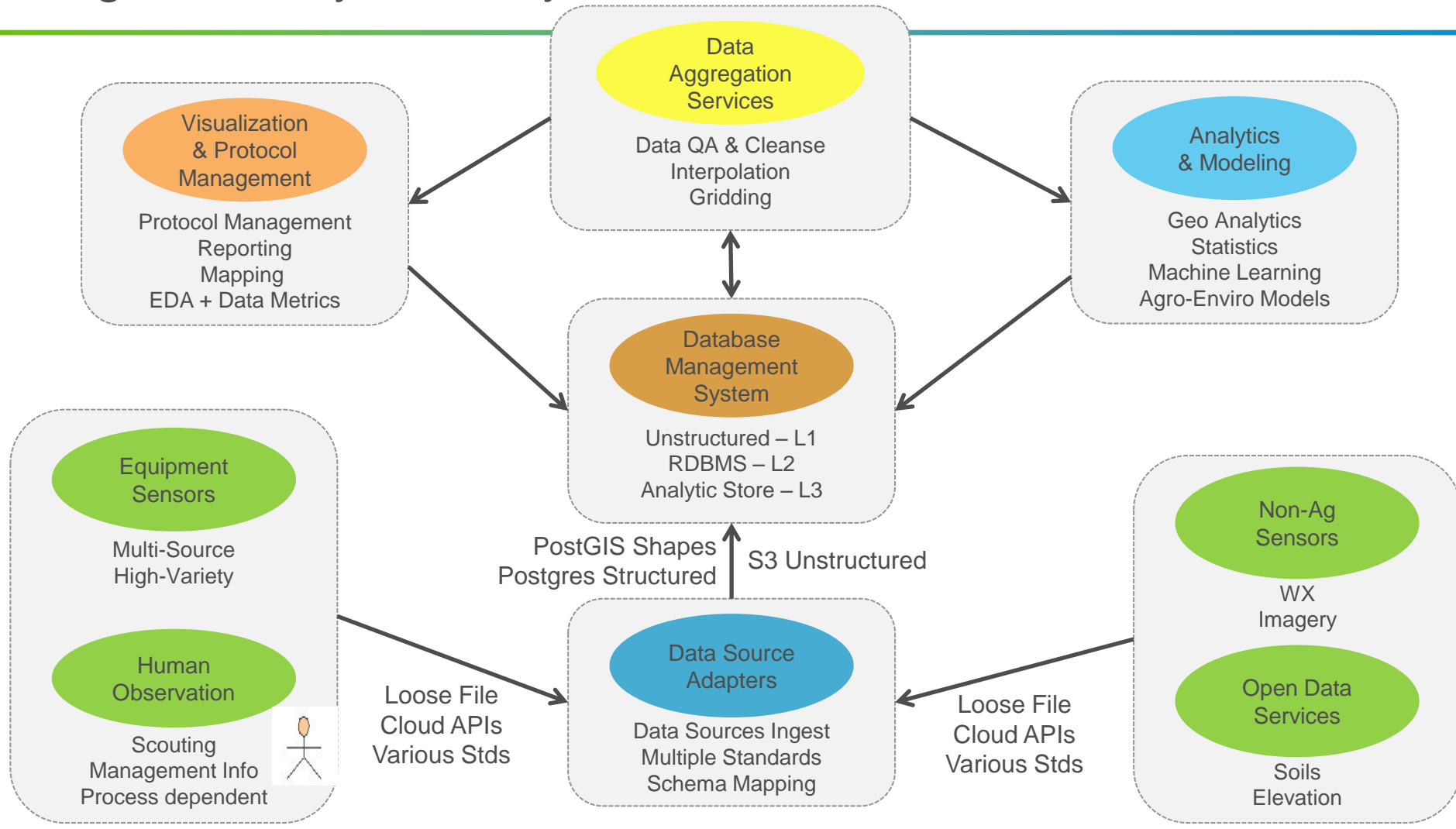
Bayer to evaluate this next quarter – global objective



```
<?xml version="1.0" encoding="UTF-8"?>
- <SoilSample xmlns="http://www.sstsoftware.com/EDS/SoilSamplingTask.xsd" SchemaVersion="2.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <SyncID>20</SyncID>
  <ParentEventID>dcf0831c-aa91-41bb-9a77-6ae2367e2e71</ParentEventID>
  <ModifiedOn>2016-07-14T16:19:33.248Z</ModifiedOn>
  - <Records>
    - <Record>
      <RecordNum>0</RecordNum>
      <SampleID AgXAttID="502">1</SampleID>
      - <TopsoilSamplingDepth>
        <Depth AgXAttID="503">12</Depth>
        - <Unit AgXAttID="504">
          <ID>52</ID>
          <Name>in</Name>
        </Unit>
      </TopsoilSamplingDepth>
      <Soil_pH AgXAttID="728">6.9</Soil_pH>
      - <OM>
        <Measure AgXAttID="731">3</Measure>
        - <Unit AgXAttID="1217">
          <ID>329</ID>
          <Name>Percent</Name>
        </Unit>
      </OM>
      - <P>
        - <Phosphorus>
          - <ExtractionMethod AgXAttID="739">
            <ID>2</ID>
            <Name>Bray 1</Name>
          </ExtractionMethod>
          - <ObservedP>
            <Measure AgXAttID="738">70</Measure>
            - <Unit AgXAttID="740">
              <ID>324</ID>
              <Name>ppm</Name>
            </Unit>
          </ObservedP>
        </Phosphorus>
      </P>
    </Record>
  </Records>
</SoilSample>
```

FTPPro Architecture

High-Level System Object Model



Principles of our Approach

Iterations on a Minimum Viable Product



- **Standardize the workflow UI and UX in software first**
 - Maximizes Product Owner participation at the start
 - *Then* iteratively re-engineer the back-end and interfaces as demanded by user stories
- **Maximize use of Open Systems and Open Standards in early platform development phases**
 - Core back-end systems are the hardest to change
 - Use COTS proprietary systems for fast build-out where expedient
- **Maximize use of loosely coupled web services for platform interfaces**
 - It is easier to use other's interfaces than design and build your own
 - De-couple interfaces later where proprietary interfaces or tight-coupling was used in early iterations, guided by technical debt and/or user stories
- **Be highly aware of accumulating technical debt**
 - Keep track in the product backlog
 - Standards are practical and essential but require thinking beyond the next 2 sprints

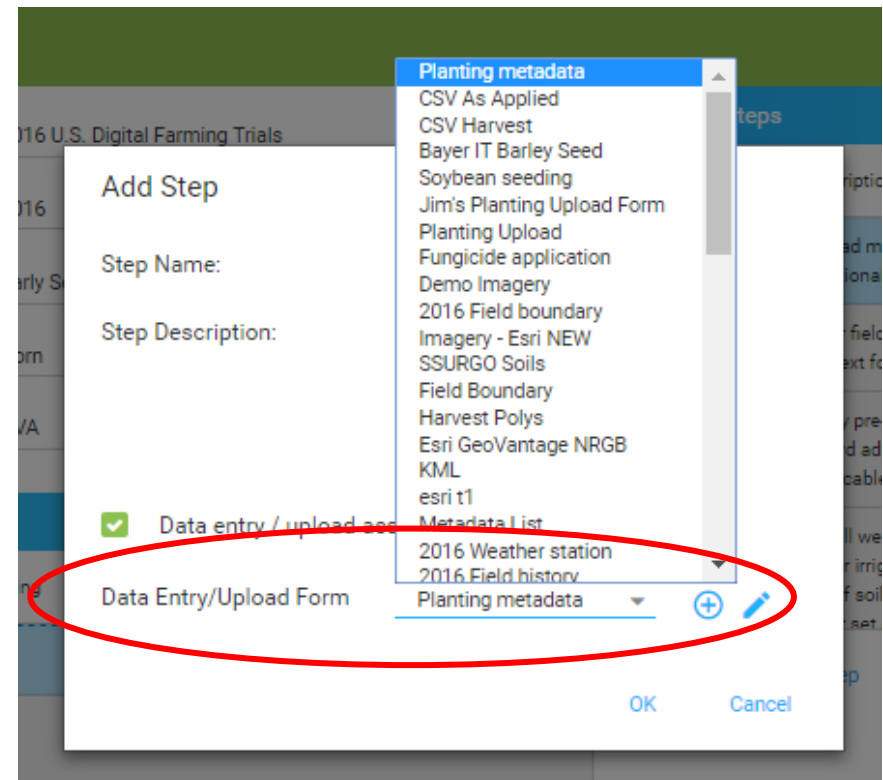
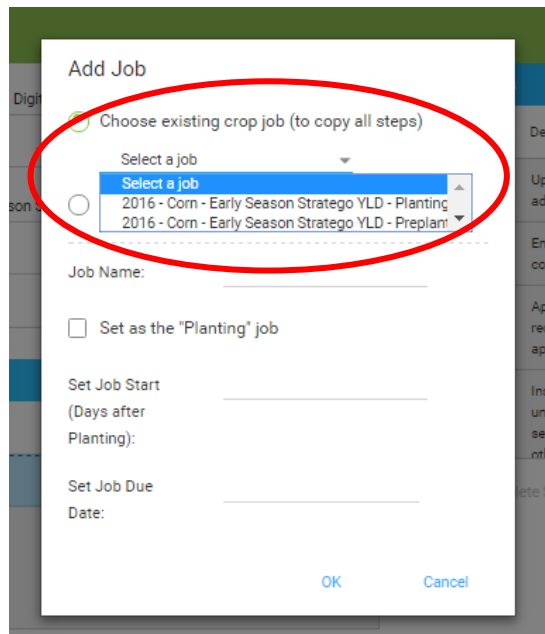
Standardizing the Protocol Contract

Template the protocol and its component parts



Each protocol and sub-elements are configurable templates

- The Job is a template
- The Job Step is a template



Allows flexibility for each experiment while governing inputs

Standardizing the Protocol Workflow

Adding Context – Geospatial Data is not Enough



Upload Data for 2016 Harvest

Step 1: Select a ZIP File for Upload

File Name	Type	Comments
<input type="button" value="Choose File"/> No file chosen	zip	No comment

Step 2: Complete Additional Data

Harvest date: 10/16/2015

Header width: 30 feet

Display monitor make & model: 2630

% Moisture: 17.5 %

Step 3: Add Step Comments

Add Comments

Upload Instructions: Upload zipped shapefile of harvest data containing .shp, .shx, .dbf, and .prj files, or zipped .csv file. Complete all additional data. Please name the uploaded file with the following nomenclature: GrowerFarms_Fieldname_Harvest

Upload Status

Upload Cancel

Metadata Capture Form

Contextual data varies depending on the job step

- Provides a flexible way to collect field and protocol management data without making changes to the data model
- The seed variety may be very important for understanding the harvest
- Equipment manufacturer will be important for application, but not for soil sampling

Standardizing the Analytic Pipeline

JSON Templates for Analytics



Authoring the Data Input Form

Each data set type has unique spatial processing sequences

- Spatial processing templates are easily configured and stored as JSON objects
- Gives users the ability to tailor analytics to the research objectives and source of data collection

JSON JavaScript Object Notation

```
object {2}
  point {4}
    method : empirical_bayesean_kriging
    aggregation : MEAN
    applicable_attributes [1]
      0 : yld_y01_dr
    parameters {3}
      transformation_type : NONE
      threshold_type : EXCEED
      semivariogram_model_type : LINEAR
  polygon {2}
    method : tabulate_intersection
    applicable_attributes [1]
```

Syntax

```
EmpiricalBayesianKriging_ga (in_features, z_field, {out_ga_layer}, {out_raster}, {cell_size}, {transformation_type}, {max_local_points}, {overlap_factor}, {number_semivariograms}, {search_neighborhood}, {output_type}, {quantile_value}, {threshold_type}, {probability_threshold}, {semivariogram_model_type})
```

Standardizing the Ingest

Templates for Schema Mapping and Visualization



Each internal geospatial data type has a standard schema

- Interactive schema mapping transforms data from any source to a common Level 2 data model for visualization, plus first-order analytics and modeling
- Deeper analytics go to the Level 1 data store
- Visualization symbology is also a user-modifiable template

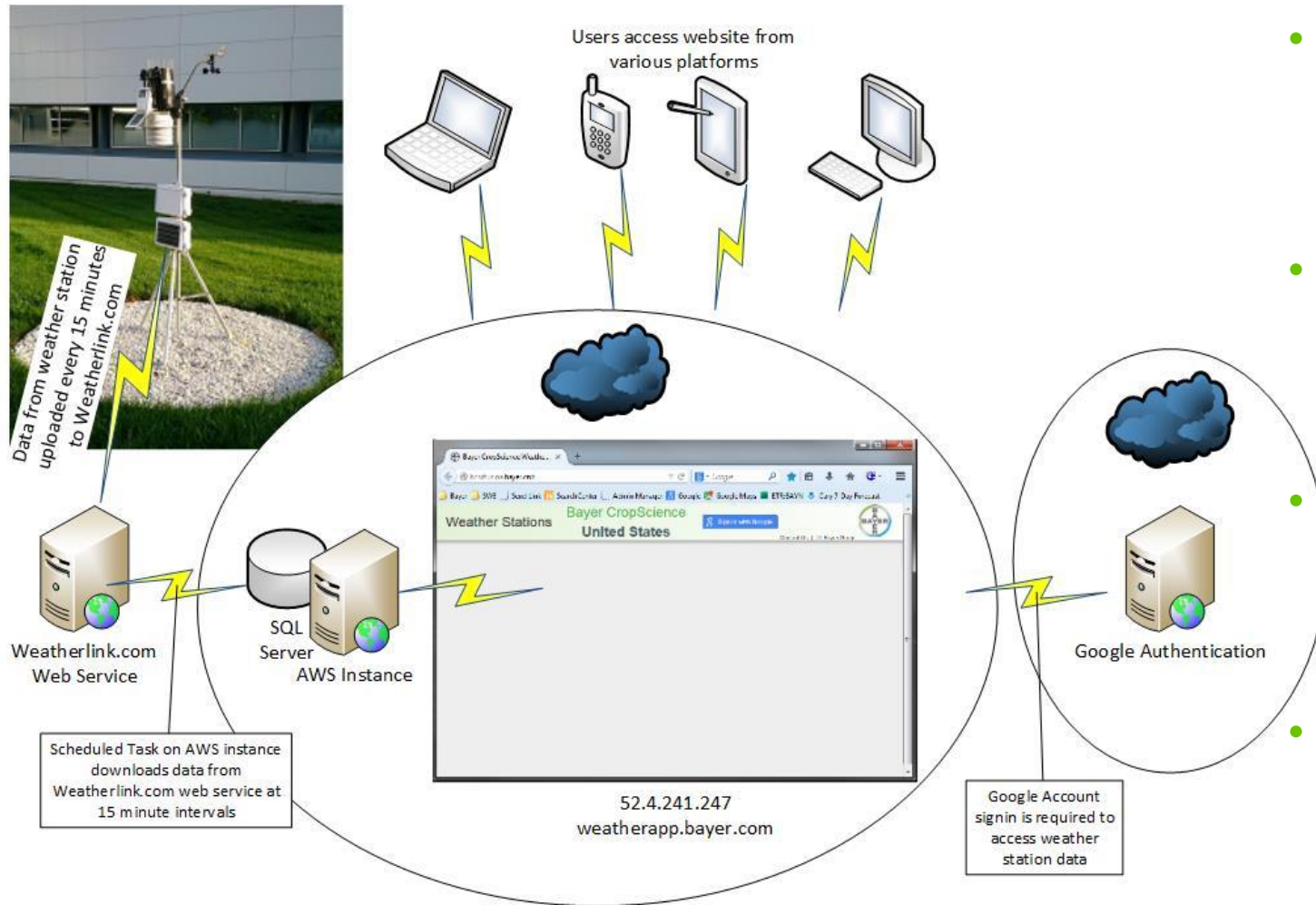
Schema mapping

The image shows a software interface for geospatial data processing. On the right, a dialog box titled "Processing Data for Harvest Corn" is open, showing a "Map the source fields to the target fields" section. This section contains a table with "Target Field" and "Source Field" columns. A dropdown menu is open over the "Source Field" column, showing a list of options including "(None)", "Adj_Mois", "Cycles", "Distance", "Dry_Yield", "Elevation", "Fld_Mois", "Flow", "Heading", "Latitude", "Longitude", "Pass", "RecNum", "Speed", "Swath", and "Wet Yield". Below the table is a "Status" field and "Proceed" and "Cancel" buttons.

On the left, a map interface is visible, showing a satellite view of a field with various data layers overlaid. A legend on the left side of the map lists several data types, including "DU_SoilChem", "DU_TissueSample", "DU_PesticideApplication_POINTS", "DU_YieldMap_POINTS", "DU_Seeding", "DU_PesticideApplication_POLYGONS", "DU_YieldMap_POLYGONS", and "DU_ExperimentalDesign". The map also shows a "Harvest - Corn Harvest" layer with a color-coded legend and a "Planting - Corn Seeding" layer. A red circle on the map is labeled "Esri Index 1 Feb 1".

Weather Service for On-Field Stations

Private weather station aggregation architecture v.1



- Each station xmits raw to Davis weatherlink cloud, 15 min intervals
- Weatherlink aggregates, publishes JSON messages
- BCS AWS service subscribes and aggregates JSON in RDBMS
- BCS publishes REST interfaces for consuming apps



Science For A Better Life



Thank you!

Kris Matson -- Geo-Analytics, Data Visualization, Software Development

Kris.Matson@bayer.com / kmatson@lifescleanalytics.com / m: 919-810-1839