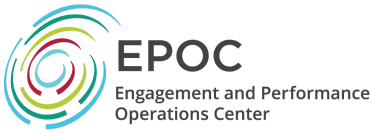

KANSAS STATE UNIVERSITY

Great Plains Network - Kansas State University Agronomy Application Deep Dive

May 20, 2019

U.S. DEPARTMENT OF ENERGY
Office of Science

ESnet
ENERGY SCIENCES NETWORK

Disclaimer

Great Plains Network - Kansas State University Agronomy Application Deep Dive

Final Report

Great Plains Network Annual Meeting
Kansas City, MO
May 20, 2019

The Engagement and Performance Operations Center (EPOC) is supported by the National Science Foundation under Grant No. 1826994.

ESnet is funded by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. Benjamin Brown is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231.

This is a University of California, Publication Management System report number LBNL-2001321[1].

[1]https://escholarship.org/uc/item/472814pk

Participants and Contributors

Kamuela Ahuna, Southwestern Oklahoma State University
Daniel Andresen, Kansas State University
Tom Boehmer, Nutanix
Christopher Bottoms, University of Missouri
Kevin Brandt, South Dakota State University
Brian Burkhart, OneNet
David Chaffin, University of Arkansas
Kevin Czaicki, CenturyLink
Ian Czarnezki, Kansas State University
Riley Epperson, University of Kansas
Derrick Erhart, Black Hills State University
Jeremy Evert, Southwestern Oklahoma State University
David Farmer, University of Minnesota
Allen Finne, University of Arkansas for Medical Sciences
Daniel Flippo, Kansas State University
Steven Fulkerson, University of Arkansas System
Jacob Gotberg, University of Missouri
Kyle Gruhn, University of South Dakota
Rick Haugerud, University of Nebraska
Ganga Hettiarachchi, Kansas State University
Neal Hodges, South Dakota School of Mines and Technology
Terry Homan, Nutanix
Kyle Hutson, Kansas State University
Jonathan Jakubcin, CenturyLink
Douglas Jennewein, University of South Dakota
Chad Julius, South Dakota State University
Branon Kane, Ciena
Mary Knapp, Kansas State University
Andrew Laubach, OneNet
Xiaomao Lin, Kansas State University
Asif Ahamed Magdoom Ali, University of Missouri
Brian Marxkors, University of Missouri
Dave Miller, Dakota State University
Kim Owen, North Dakota State University
Scott Pohlman, CenturyLink
Spencer Riley, Southwestern Oklahoma State University
Mike Rose, Ciena
Vonley Royal, OneNet
Jennifer Schopf, Indiana University
Garrett Stevens, Black Hills State University
Arnaud J. Temme, Kansas State University
Stephen Welch, Kansas State University
Ruth Welti, Kansas State University
Jason Zurawski, ESnet

Report Editors

Hans Addleman, Indiana University: addlema@iu.edu
Scott Chevalier, Indiana University: schevali@iu.edu
George Robb III, ESnet: grobb3@es.net
Jennifer M. Schopf, Indiana University: jmschopf@indiana.edu
Jason Zurawski, ESnet: zurawski@es.net

Contents

1 Executive Summary

In May 2019, staff members from the Engagement and Performance Operations Center (EPOC) and members of the Great Plains Network (GPN) attending their Annual Meeting meeting met with researchers at Kansas State University (KSU) for the purpose of an Application Deep Dive training session. The goal of this training session was to help characterize the requirements for an agronomy application, to enable cyberinfrastructure support staff to better understand the needs of the researchers they support, and to offer training to GPN members to be able to conduct these on their own. Material for this event includes both the written documentation from the agronomy application at KSU, details about the infrastructure setup at KSU, and also a writeup of the discussion that took place in person on May 20, 2019.

EPOC, GPN, and KSU recorded a set of action items for this use case, continuing the ongoing support and collaboration. These are a reflection of the case study report, and in person discussion.

1. KSU and the Welch team are exploring additional storage resources for data and model results.
2. KSU and Welch will explore the use of additional HPC resources, including TACC and XSEDE
3. KSU and Welch will explore additional network connectivity to fields that are remote. They are currently using microwave (or hand carrying discs), but as data volumes grow a fiber connection will be needed.
4. KSU and Welch are exploring the creation of portals to share data/metadata of work.
5. KSU and Welch will investigate better backup arrangements for data.
6. KSU and GPN will evaluate network connectivity to collaborators in industry and research partnerships, including GPN/KanREN peering arrangements.

2 Process Overview and Summary

2.1 Deep Dive Background

Over the last decade, the scientific community has experienced an unprecedented shift in the way research is performed and how discoveries are made. Highly sophisticated experimental instruments are creating massive datasets for diverse scientific communities and hold the potential for new insights that will have long-lasting impacts on society. However, scientists cannot make effective use of this data if they are unable to move, store, and analyze it.

The Engagement and Performance Operations Center (EPOC) uses Application Deep Dives as an essential tool as part of a holistic approach to understand end-to-end data use. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports five main activities

- Roadside Assistance via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- Application Deep Dives to work more closely with application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues;
- Provision of managed services via support through the IU GlobalNOC and our Regional Network Partners;
- Coordinated Training to ensure effective use of network tools and science support.

Whereas the Roadside Assistance portion of EPOC can be likened to calling someone for help when a car breaks down, Deep Dives offer an opportunity for broader understanding of the longer term needs of a researcher. Deep Dives aim to understand the full science pipeline for research teams and suggest alternative approaches for the scientists, local IT support, and national networking partners as relevant to achieve the long-term research goals via workflow analysis, storage/computational tuning, identification of network bottlenecks, etc.

The Deep Dive approach is based on an almost 10-year practice used by ESnet to understand the growth requirements of DOE facilities[2]. The EPOC team adapted this approach to work with individual science groups through a set of structured data-centric conversations and questionnaires.

[2] https://fasterdata.es.net/science-dmz/science-and-network-requirements-review

2.2 Deep Dive Structure

Deep Dives are basically structured conversations between a research group and relevant IT professionals to understand at a broad level the goals of the research team and how their infrastructure needs are changing over time.

The researcher team representatives are asked to communicate and document their requirements in a case-study format that includes a data-centric narrative describing the science, instruments, and facilities currently used or anticipated for future programs; the advanced technology services needed; and how they can be used. Participants considered three timescales on the topics enumerated below: the near-term (immediately and up to two years in the future); the medium-term (two to five years in the future); and the long-term (greater than five years in the future).

The Case Study document includes:

- ***Science Background***—an overview description of the site, facility, or collaboration described in the case study.
- ***Collaborators***—a list or description of key collaborators for the science or facility described in the case study (the list need not be exhaustive).
- ***Instruments and Facilities***—a description of the network, compute, instruments, and storage resources used for the science collaboration/program/project, or a description of the resources made available to the facility users, or resources that users deploy at the facility.
- ***Process of Science***—a description of the way the instruments and facilities are used for knowledge discovery. Examples might include workflows, data analysis, data reduction, integration of experimental data with simulation data, etc.
- ***Remote Science Activities***—a description of any remote instruments or collaborations, and how this work does or may have an impact on your network traffic.
- ***Software Infrastructure***—a discussion focused on the software used in daily activities of the scientific process including tools that are used to locally or remotely to manage data resources, facilitate the transfer of data sets from or to remote collaborators, or process the raw results into final and intermediate formats.
- ***Network and Data Architecture***—description of the network and/or data architecture for the science or facility. This is meant to understand how data moves in and out of the facility or laboratory focusing on local infrastructure configuration, bandwidth speed(s), hardware, etc.
- ***Cloud Services***—discussion around how cloud services may be used for data analysis, data storage, computing, or other purposes. The case studies included an open-ended section asking for any unresolved issues, comments or concerns to catch all remaining requirements that may be addressed by ESnet.

- ***Resource Constraints***—non-exhaustive list of factors (external or internal) that will constrain scientific progress. This can be related to funding, personnel, technology, or process.
- ***Parent Organization***—overview of the sources of funding and cooperation that facilitate the process of science and technology support.
- ***Outstanding Issues***—Final listing of problems, questions, concerns, or comments not addressed in the aforementioned sections.

At an in-person meeting, this document is walked through with the research team (and usually cyberinfrastructure or IT representatives for the organization or region), and an additional discussion takes place that may range beyond the scope of the original document. At the end of the interaction with the research team, the goal is to ensure that EPOC and the associated CI/IT staff have a solid understanding of the research, data movement, who's using what pieces, dependencies, and time frames involved in the case study, as well as additional related cyberinfrastructure needs and concerns at the organization.. This enables the teams to identify possible bottlenecks or areas that may not scale in the coming years, and to pair research teams with existing resources that can be leveraged to more effectively reach their goals.

2.3 Great Plains Network - Kansas State University Agronomy Application Deep Dive Background

In May 2019, EPOC and GPN organized a Deep Dive in collaboration with Stephen Welch from KSU to characterize the requirements for an agronomy application as part of a training session at the GPN Annual Meeting. The KSU representatives were asked to communicate and document their requirements in a case-study format (see Section 3). The use case for this deep dive was an Agronomy application supported by Welch. Several data sources for the application are given in Appendix A.

The face-to-face meeting took place at the Great Plains Network Annual Meeting in Kansas City, MO, on May 20 (see discussion in Section 4). We document next steps in Section 5.

2.4 Organizations Involved

The Engagement and Performance Operations Center (EPOC) was established in 2018 as a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

The Energy Sciences Network (ESnet) is the primary provider of network connectivity for the U.S. Department of Energy (DOE) Office of Science (SC), the single largest supporter of basic research in the physical sciences in the United States. In support of the Office of Science programs, ESnet regularly updates and refreshes its understanding of the networking requirements of the instruments, facilities, scientists, and science programs that it serves. This focus has helped ESnet to be a highly successful enabler of scientific discovery for over 25 years.

Indiana University (IU) was founded in 1820 and is one of the state's leading research and educational institutions. Indiana University includes two main research campuses and six regional (primarily teaching) campuses. The Indiana University Office of the Vice President for Information Technology (OVPIT) and University Information Technology Services (UITS) are responsible for delivery of core information technology and cyberinfrastructure services and support.

The Great Plains Network (GPN) is a non-profit consortium that aggregates networks through GigaPoP connections. They advocate for research on behalf of universities and community innovators across the Midwest and Great Plains who seek collaboration, cyberinfrastructure, and support for big data and big ideas, at the speed of the modern Internet.

Kansas State University (KSU) is a public research university with its main campus in Manhattan, Kansas. KSU was opened in 1863 as the nation's first operational land-grant university. The university is classified as one of 115 research universities with the highest research activity (R1) by the Carnegie Classification of Institutions of Higher Education.

3 Kansas State Agronomy Use Case Study

3.1 Science Background

If humanity is to avoid severe global food security disruptions in the coming few decades, crop production must double by 2050, which will require an increase on the order of 2.3% per year. All major grain crops, including wheat, are currently expanding at less than one half this rate and, in some cases, barely one quarter. In addition to needing more crops to support a larger population, we are also facing water and temperature changes that may adversely affect crop yields. For example, 20% of calories across human population comes from wheat, which is being significantly impacted by climate changes.

The rates of development and growth of individual plants comprising a crop canopy is quantitatively affected by many external factors, including air and soil temperatures, sunlight, availability of nutrients, and insect herbivory. Internal factors also affect growth and development, such as the leaf area, the available pools of stored starches and other carbohydrates, and the deleterious impacts of plant diseases. The most important internal influences are the actions of the gene networks, which function as cybernetic control systems whose aim is to secure plant survival and reproduction.

A deeper understanding is also needed to understand how to treat and maintain crops, which involves developing complex models of plant development and ecosystem interactions. For example, field temperature is generally treated as a single variable, but Figure 1 shows the actual temperature variance across a single field. The models used in this research enable a deeper understanding of these systems in part because where older models used constants, the current approach now uses additional submodels or even active measurement data for higher accuracy. Understanding the changes in day-by-day through more detailed modeling can result in more predictable and better outcomes in actual field trials and production crops.

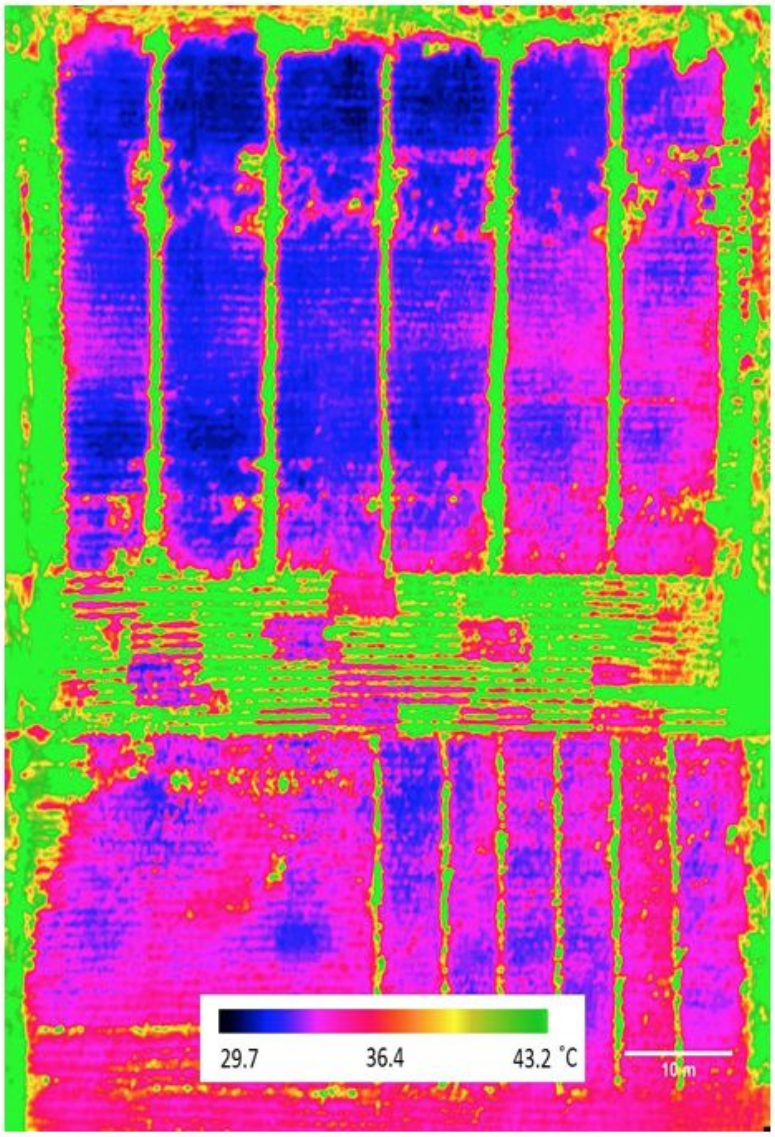

Figure 1 - The temperature map of a single field (Credit: Ciampitti and Prasad)

Central to increasing crop production rates is the ability to quantitatively predict the traits that a plant will exhibit in a natural or managed environment given a particular genetic constitution. Breeders need such information to select those particular genetic types, which, when combined, will yield the most advantageous result. In order to have effective production processes, a farmer must know the specific management actions to take (and when to take them) for each particular genetic type in a specific environment or field.

Standard crop improvement programs are long-term efforts.. In general, seven or more years of trials (or plant generations) might result in only ten to twelve new plant variants. The goal of this research to accelerate this cycle by using more advanced models and new technological approaches to gathering data. Overall, with funding from the National Science Foundation, the team is co-designing

mathematical models, sensor systems, and field devices, including robots, to make integrated quantitative predictions of the full plant behavior and to act on them appropriately.

3.2 Collaborators

In addition to the KSU collaborators who contribute data sets to the models (see Appendix A), the primary collaborators are at UC Davis, Kansas University, Langston University, and Oklahoma State University.

In addition, there are some industrial partners that are expanding into the research areas, including:

- Microsoft
- IBM
- BASF, which produces agricultural products such as fungicides, herbicides, insecticides, and seed treatment products
- Corteva, the agriculture division of DowDuPont, a major seed company
- Topcon Agriculture, which produces agricultural machinery positioning, sensor, and control devices/software
- Veris Technologies who design, build, and market sensors and controls for precision agriculture.

3.3 Instruments and Facilities

Much of the above research involves sensor data collected from various sources (see Appendix A). In addition to personal and/or lab-scale devices/facilities, computational processing take place on two resources, Beocat, which is part of the KSU Institutional computing, and resources at the Texas Advanced Computing Center (TACC). The KSU resource is the primary compute resource for the project. and is used to prototype work before the largest runs are sent to TACC.

3.4 Process of Science

There are three main team projects by the team that were described in Welch's presentation on May 20, 2019.

The first project involved a growth chamber experiment where photos of hundreds of *Arabidopsis thaliana* plants were taken 16 times per day to understand plant development, as shown in Figure 2. *A. thaliana* is a small flowering plant in the mustard family that is of great interest it is one of the primary model organisms used for studying plant biology and the first plant to have its entire genome sequenced.

The growth chamber was located at the University of California at Davis. Images were sent daily from Davis to TACC for intermediate storage and forwarding to KSU. A team of five graduate students have implemented an in-house developed software pipeline to process these images. The pipeline required over two weeks on the lab cluster to analyse the imagery from a single, 30-day experiment. In addition to

processing time, data storage is a major issue, particularly over the long term. While the raw data is only ca. 5 TB, intermediate checkpoints are ten times larger and absolutely required because of the extreme processing time. In addition, checkpoint files are kept so that re-working models can be done partially through the pipeline, without having to restart the pipeline from the beginning.

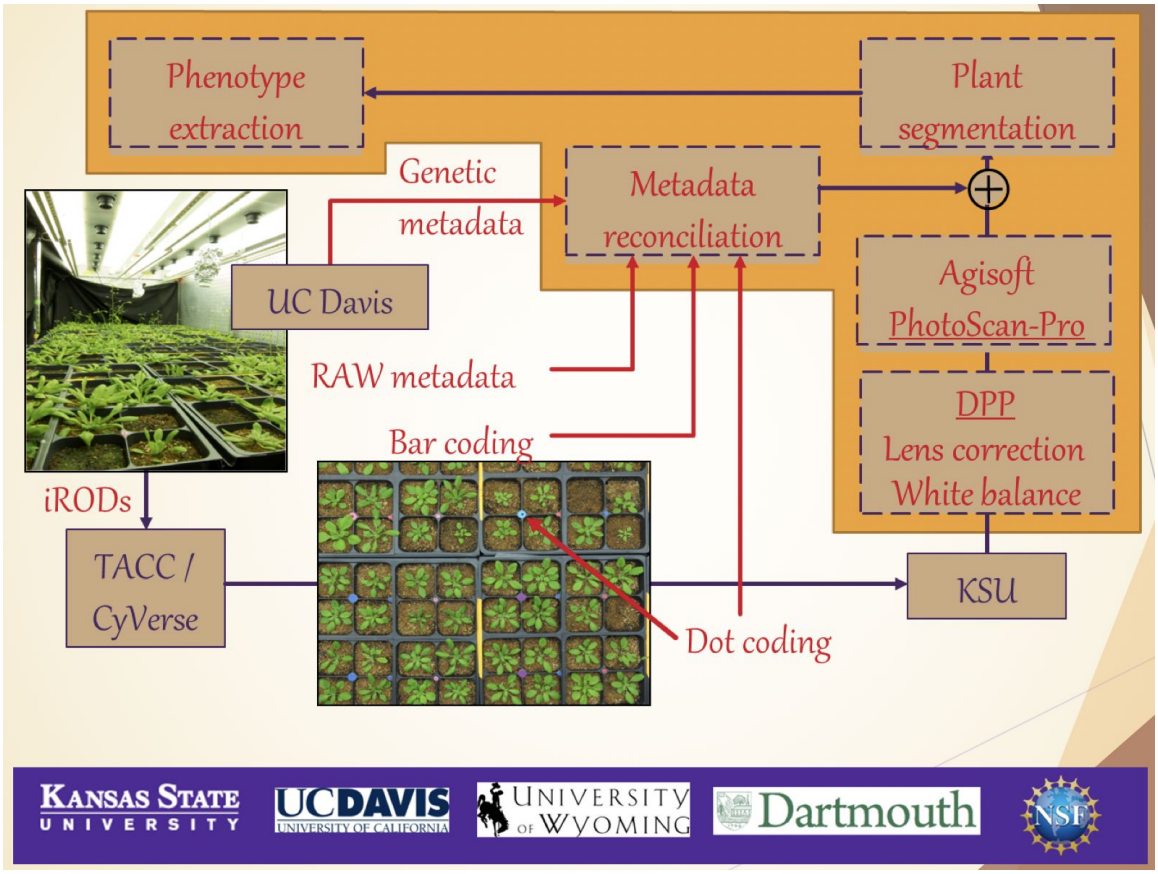

Figure 2 - Workflow for Project 1.

The second project is being conducted in collaboration with Carl Leuschen at the University of Kansas (KU). A cart-mounted microwave radar, shown in Figure 3, generating 2-18 GHz chirps every 240 mu-sec is used to scan 1 x 5meter wheat plots in a breeding trial. Every two weeks, 315 plots are sampled, producing 122 MB of raw data per plot. The data was then physically transport from the field to a lab at KU, where reduction performed reduces each plot reading to a 40KB binary file. Last year, a set of 1,200 of these files were transmitted to KSU for further processing.

Figure 3 - Instrumentation used in Project 2.

The third project, outlined in Figure 4, involved parameter estimation and ran models on resources at TACC. There were over 384 million different variants to the model. The input to the computation was only 3MB, but the computation time was over 65,000 hours. The resulting spreadsheet of 50K was transferred back to KSU.

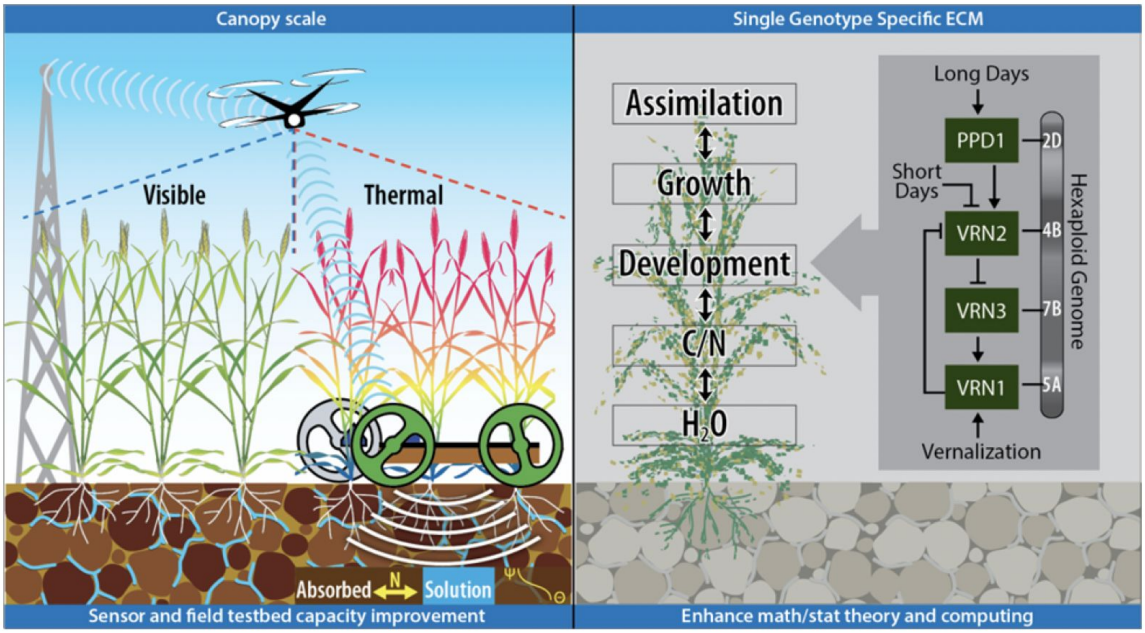

Figure 4 - Workflow for Project 3.

Currently this third project is the largest collaboration, and is supported by NSF EPSCoR[3]. Data flows from this project are expected to include aerial imagery, soil

[3] https://nsf.gov/awardsearch/showAward?AWD_ID=1826820

electromagnetic measurements, weather data, and time series of gene expression data. These datasets will largely be used to estimate the numerical constants in mathematical models and, independently of that, test model validity by comparing predictions with observation. This project is also developing new sensors and robotics approaches to measurement.

This research uses electromagnetic methods to determine several soil variables, including moisture levels and nitrogen content. A robot operates these sensors, and the collected data is transferred back to the lab via the internet. One of the challenges is understanding the current bounds on what the data limitations are, and when changes to the data sizes will affect other pieces of the computational pipeline.

3.5 Remote Science Activities

The third mentioned project (e.g. "***Building Field-Based Ecophysiological Genome-to-Phenome Prediction***") involves collaboration with 3 universities; KSU, Oklahoma State University (OSU), and Langston University in Oklahoma. Multiple test sites are present in both states, and one of the challenges is to create a uniform deployment for data compatibility.

3.6 Software Infrastructure

Currently, all soil samples are brought back to a lab environment since it isn't possible to do extensive chemical analysis in the field. Ideally, the team would like to be able to process a larger number of samples in a lab – but also automate the method to store and process the data for use in the models.

Currently, large crop models are mostly legacy Fortran codes that have been re-written verbatim into more recent versions of their native languages but not restructured. There are two major suites of models now in global use, but without any form of systematic funding to support them - primarily people have largely donated time and resources to maintaining them. There is no central planning, only ad hoc conversations when it has been needed. The models are in the public domain and researchers contribute as they can. That said, there are ca. 1,000 modelers worldwide currently working with these models routinely, and over 14,000 who have received training in their basics over the last ten years.

In the general case, research staff manipulate accumulated data into machine-readable formats (Microsoft Excel/CSV) that are processed through a variety of scripts written in Python or BASIC. The team is also working on developing a database and website for mass spectrometry information on plant lipids.

Currently, the team writes the majority of the software analysis packages, typically in either the Anaconda or Enthought Python. Sometimes, other codes such as ImageJ

[4] or one of the Decision Support System for Agrotechnology Transfer (DSSAT)[5] crop models may also be used.

The outputs of the models need to be post-processed into a human readable output, generally some kind of visualization. This is often in the form of a time series plot. For the EPSCoR project, new visualization methods are being developed. Current data analysis tools include R, Matlab, Python scripts, and various typical desktop products.

One ongoing issue the team deals with is that of data formats. Crop models from the 1980s used ASCII formats in text files, not databases, which means to incorporate these data sets, there need to be format converters and significant custom software. The EPSCoR project, which combines both models and sensor data, is trying to start over in terms of the data models, and is co-designing new ways to store and work with the data. For example, there may be up to thirty different models that are used to estimate wheat growth, and the team is trying to unify their formats for better comparisons.

All software is shared when requested, but no formal system such as github is being used at present.

Going forward, using GPUs is of interest especially for the visualization software. The bulk of the current models are not well structured for GPU systems at this time.

3.7 Network and Data Architecture

Present

KSU has a number of local HPC resources available for the research community. These include Beocat (a resource with ~8K cores, 3+PB raw storage running Ceph) which offers free access for KSU researchers and their external partners. This infrastructure is operated on a condo model, thus there is priority given to owned resources when needed.

Local HPC support (including Beocat) includes 3.5 FTE support staff, including director (0.5FTE), two sysadmins, and an application scientist available to help users integrate in research. There are introductory training sessions available online, as well as taught once per semester in person. Additionally, KSU offers access to a Big Data workshops via PSC and XSEDE. Dr. Dan Andresen is an XSEDE Campus Champion, and helps users that outgrow Beocat and other resources to move on to XSEDE-class systems.

Beocat is connected to internal KSU network at 40Gbps, and to KanREN/I2 at 100Gbps. The campus also hosts a FIONA device through the GPN RP program. Most buildings with heavy research activity are connected at 40Gbps to campus core,

[4] https://imagej.nih.gov/ij/
[5] https://dssat.net/

with 1Gbps to desktop (and 10 Gbps to a relatively small number of desktops). KSU traffic to KanREN/I2 (except for the Beocat DTN) is limited to 20Gbps with two 10G connections to KanREN.

For the most part, the agronomy research uses KSU compute resources such as Beocat to prototype the models locally and not waste external resources such as TACC.

When the model runs are complete, the data is stored on KSU resources. Ideally, final and intermediate results are shared with external collaborators as well. There is no portal being currently used, but there is an identified need to be able to publish data sets (with descriptive metadata) so that others can use it within the community. Data backups are also a concern due to lack of available storage space in alternative locations.

Research groups at KSU have access to a campus digital repository (K-Rex[6]) for storage, however, K-Rex is limited with respect to large data files (over 1GB). K-Rex was created to store research products such as research papers and thesis as opposed to output data sets from instrumentation. As a result of this, the project occasionally uses Beocat as a way to store their larger data sets for longer term. Data transfers with the Beocat infrastructure are typical done using Globus GridFTP or using desktop tools, such as WinSCP or rsync.

Beyond the K-Rex storage, most departments also have a local file server connected at 1-10Gbps. These resources are not centrally managed by KSU.

2-5 Years

Long-term image storage is an area of longer term concern. Image data from large studies needs to be kept purely as a matter of appropriate scientific record-keeping and the necessity that results be repeatable over time. Additionally, data needs to be kept so that new analysis algorithms can be applied to the same data as older methods to measure what, if any, improvements have been achieved. As one concrete example, a current limitation of deep learning machine vision studies is large data sets for training. Older data sets can be a valuable resource for this.

Furthermore, an emergent training method is the generation of synthetic image data via simulation. If this data has sufficient verisimilitude, it may be directly used in training. Alternatively, it might be the product of adversarial networks whose goal is to produce training data as realistic as possible as measured by its ability to fool the network being taught. Good training images and good adversarial images are resources that can easily be found to be reusable in other contexts and should be kept.

[6] https://krex.k-state.edu/dspace/

Finally, because image processing pipelines can contain multiple compute-intensive steps, there is merit not just keeping starting data and end products but also selected intermediate checkpoint dumps as well.

All of these reasons suggest a need for inexpensive, massive, long-term image storage archiving.

3.8 Cloud Services

Dropbox, Google Drive, and Microsoft Cloud commodity storage services have been used in the past, but currently data is more commonly stored on Beocat. Efforts to stand up a Microsoft cloud-based CUI infrastructure by KSU IT staff are being explored for production use in CY 2019.

3.9 Known Resource Constraints

Storage is the biggest constraint identified.

3.11 Outstanding Issues

The lack of a centralized research data repository can be limiting for research groups. At the current time there is not one widely available (beyond the K-Rex infrastructure), or being planned by either Central IT or the research computing groups. This effort could be strengthened if the respective research groups (e.g. domain experts, information technology, data scientists) worked to build better practices and infrastructure for the curation and management of research products.

4 Discussion Summary

On May 20, 2019, members of the EPOC team, members of the GPN attending the All Hands Meeting, and KSU IT staff met with Stephen Welch to walk through his application as part of a GPN Annual Meeting Training Session for Deep Dives.

During the discussion, the following points (outside of clarifications to the Case Study described in Section 3) were emphasized:

- Getting the data from the sensors on the field to the lab might be able to be improved. Researchers at Minnesota are using wireless to do this.
- Researchers are actively exploring including additional variables for the models, which is likely to significantly increase the computational time needed for the models. Many models from the 1980s had constants for values where it wasn't possible to pragmatically measure the values. New models are now adding in variables and submodels for these aspects.
- Questions were asked about the need for a library of models, especially in light of the added complexity. Some of archives already exist - there's a global repository of models that includes 100-150 different models. There are also two major suites of models, one from the US and one from Australia. The US version has 42 models that use interoperable data formats. The US suite has only a single wheat model and there are many other wheat models that aren't compatible.
- New technology, such as a handheld gene sequencing device, would contribute to the research approach. In general, there's an interest in anything researchers can do to increase the number of variables they can observe in the field and in adding new sensors to replace outdated parts of the models.
- One component of the EPSCoR project is looking at basic mathematical formalisms that are used in model construction.
- Software longevity is an issue as graduate students who have written the software eventually leave the lab. There isn't currently a formal process for maintaining scripts or student-developed software..
- Storage is an issue, and many institutions in addition to KSU are examining different approaches and trying to raise this as a significant research need to higher level administration. Many GPN sites are looking at cloud-based storage solutions, as it is not feasible to keep growing locally. It was stated that researchers can't always afford what they would prefer, for example, pulling data out of the cloud because it can be very expensive.
- Advanced file formats can be more compressive than legacy ones and might be considered, however, lossy formatting would need to be strongly managed.

5 Action Items

EPOC, GPN, and KSU recorded a set of action items for this use case, continuing the ongoing support and collaboration. These are a reflection of the case study report, and in person discussion.

1. KSU and the Welch team are exploring additional storage resources for data and model results.
2. KSU and Welch will explore the use of additional HPC resources, including TACC and XSEDE
3. KSU and Welch will explore additional network connectivity to fields that are remote. They are currently using microwave (or hand carrying discs), but as data volumes grow a fiber connection will be needed.
4. KSU and the Welch team are exploring the creation of portals to share data/metadata of work.
5. KSU and Welch will investigate better backup arrangements for data.
6. KSU and GPN will evaluate network connectivity to collaborators in industry and research partnerships, including GPN/KanREN peering arrangements.

Appendix A - Additional Data Sources

Theoretical plant modeling relies on a multiplicity of data sources that are researched, gathered, and curated by other groups, as listed here.

1. Ruth Welti, Kansas Lipidomics Research Center

The Kansas Lipidomics Research Center (KLRC)[7] has two primary goals:

1. Measure levels of many lipids
2. Use information on lipids to understand lipid metabolism and related genes in plants, particularly plants under stress.

1.1 Process of Science

A number of data products are produced as a result of this research. The formats, KLRC's current practices for archiving and providing access to data and materials, and future work to improve the process, are listed in the table below.

In the general case, KLRC handles samples and data from several types of users. Data and samples generated by KLRC lab members (PI lab), Co-PIs, and Senior Personnel will be made available through direct sharing, databases, and publications as described in the table. Collaborator data and data obtained on a fee-for-service basis will be returned to originating laboratories as indicated in the table. All samples and data will be handled with sensitivity (e.g. if related to human samples). If materials or data are derived from original data, they will be archived and shared in a manner similar to the original data.

Materials or data produced	Current practices at KLRC for this type of materials or data	Planned practices
Experimental protocols (document files)	Archived on hard drive and published on web at www.ksu.edu/lipid/analytical_laboratory/protocols_and_methodology/index.html	Continue current practice; expand protocol collection on the web and include video demonstrations
Data acquisition methods, i.e. files that direct MS acquisition. These are instrument-specific files in proprietary formats.	Archived on back-up drives and DVDs; shared on web at www.ksu.edu/lipid/analytical_laboratory/analysis_components/data_acquisition_methods/index.html Parameters used for acquisition are also published.	Continue current practice, but add additional methods including those from developed for Sciex 6500+ with DIM to KLRC's website.
Raw data output files, i.e. instrument-specific files of mass spectra in a proprietary format	Archived by date in duplicate on DVD. Three additional electronic copies archived on separate drives with other user- & analysis-specific data (identified by researcher and date of data acquisition).	Continue current practice, but adding remote archive site in 2016.
Processed data output files (lists of masses and signals in Excel format)	Three electronic copies archived on separate drives with other user- & analysis-specific data (identified by researcher and date information). Also stored in LipidomeDB (see Row G and footnote*). Users encouraged to make data available as supplemental data in publications. PI's and collaborators' data stored in PMR: Plant/Microbial and Eukaryotic Systems Resource at metnetdb.org/PMR/	Continue current practice; adding remote archive site in 2016; provide files to scientists who originated samples upon request. Provide users of KLRC with a statement informing them of their responsibility to share with other researchers, at no more than incremental cost and within a reasonable amount of time, the primary data and samples created or gathered in the course of work under

[7] http://www.ksu.edu/lipid

		NSF grants. Encourage them to utilize PMR or another public database.
Spectral images	Provided to users as pdfs. Three electronic copies of pdfs archived on separate drives with other user- & analysis-specific data (identified by researcher and date of data acquisition).	Continue current practice, but adding remote archive site in 2016.
Experimental samples for mass spectral analysis	Stored at -80°C; Returned to scientists when requested or discarded after approximately 1 year with permission of originating scientist.	Continue current practice.
LipidomeDB and LipidomeDB Data Calculation Environment (LipidomeDB DCE)	Available at the password-protected site 129.237.137.125:8080/Lipidomics/ Documented in Zhou et al. (2011) Lipids 46, 879-884 [reference 1]. Passwords are available to interested persons via form at site. Tools are available to those with passwords.	Continue current practice.
Other data processing strategies, including those for MRM data on the Sciex 6500+ with DIM	Housed in house and archived at "K-State Online".	Will provide relevant files via the KLRC's website.
Analytical target lists, i.e. lists of lipids, their chemical formulas, and fragments used for scanning	Some stored in LipidomeDB and available through LipidomeDB DCE. There are default target lists, available to all users, and user-generated lists, available only to the generating user. Others published as supplemental documents to papers.	Continue current practice with LipidomeDB and add other targets lists to KLRC's website.
Lipid profiles, i.e. identities and quantities of lipid compounds in experimental samples	Stored in LipidomeDB DCE; archived in Excel format in an electronic folder with other user- & analysis-specific data (identified by researcher and date information); these are stored on DVDs in duplicate plus on two hard drives.	Continue current practice plus provide users of KLRC with a statement, as described in Row D. Encourage users to provide complete lipid profile data as supplemental data at publication and to deposit data in PMR or other public database.
Metadata associated with experimental samples	Metadata, including researcher and contact information, organism, genotype, age/stage, growth conditions, tissue information, sampling conditions, other sample-specific information, tissue metrics, MS sample preparation, and analysis data are collected and stored in separate fields in Excel files with lipid profile data. For some experiments (data collected for the Arabidopsis Metabolomics Consortium, funded by NSF MCB 0520140 and MCB 0820823) and those of the PI and collaborators, data are stored in the PMR database.	Continue current practice. Encourage KLRC users to put their data and metadata in publicly available databases such as PMR.
Internal standard compounds	Mixtures are assembled at KLRC based on traditional quantification methods (GC analysis of fatty acids and phosphate analysis of phospholipids) from purchased and semi-synthesized compounds. Mixtures are aliquoted and stored at -80°C, and are available at cost to the scientific community:	Continue current practice.
Scientific results of collaborative projects	Publish in scientific journals.	Continue current practice.

KLRC performs Genome Wide Association Studies (GWAS) through the aide of the GWA Portal[8], specifically focusing on lipid levels (and other traits) as the phenotypes. This site is critical, as it stores around 1200 files, each about 300 MB, of research data (e.g. 360GB total). This information is periodically downloaded for local use. Many people upload from Kansas state but also download from this international resource; Also trying to shorten loop with industry.

[8] https://gwas.gmi.oeaw.ac.at/

1.2 Collaborators

There is extensive collaboration with the KU Molecular Graphics Lab in the form of the use of a shared server that houses curated research data.

The Plant/Eukaryotic and Microbial Systems Resource (PMR), located at Iowa State University, is used to publicly present and store processed lipid data that is available to the public.

These data products are used by many other researchers. The project also imports other data sources for their own use which are then shared out again.

1.3 Software & Hardware Infrastructure

A shared server with the KU Molecular Graphics Lab handles storage and basic processing of many aspects of the research data. There is no excessive use of HPC at this time, as most analysis can be handled on workstations.

Software use is varied, and a mixture of public resources and self-developed infrastructure:

1. We use scripts written in BASIC to extract data from the proprietary mass spec programs into Microsoft Excel. We use LipidomeDB DCE[9] in JSP/Javascript to isotopically deconvolute the data and calculate lipid amounts
2. From GWA Portal we download CSV formatted files. We filter the data with a Python script, and export relatively small CSV files that can be viewed in Microsoft Excel.
3. The PMR database at Iowa State holds some of our published data. This site holds metabolomics and transcriptomics data and can be used to integrate these.

In the general case, research staff manipulate accumulated data into user-manipulatable formats (Microsoft Excel/CSV) that are processed through a variety of scripts written in Python or BASIC. The team is also working on developing a database/website for mass spectrometry information on plant lipids.

[9] http://129.237.137.125:8080/Lipidomics/

2. Arnaud J. Temme, Soils & Geomorphology

The Soils & Geomorphology research group at KSU maps and simulates the development of soils and landscapes over millennial to human timescales. To accomplish these goals, there is extensive reliance on field data. These base observations can be statistically related to existing maps, or can be used to create new maps to contrast with and inform computer model simulations of soil and landscape change.

2.1 Process of Science

Research takes on two main forms: going out and measuring things, and using computer models to test and verify hypothesis.

Soils support the growth of crops and natural vegetation, as well as carbon storage. The presence of different soils in different parts of our landscapes is an indicator of past use as well as future growth that shape our landscapes. The timescales involved in these processes are usually hundreds or thousands of years. Luminescence and carbon dating to help quantify rates.

Natural occurrences, such as glacial retreat, expose large areas of land that were previously covered. For soils developing on this new land, the clock is starting to tick from zero again. This allows us to measure rates of soil formation, right when soils start to form. In the last few decades it has been revealed that landscapes do not function according to simple rules. For instance, twice the amount of rainfall does not mean twice the erosion. In fact, under some circumstances, twice the amount of rainfall could even mean less erosion. In other cases, patterns in a landscape, such as the sorted circles in the pictures, form without external steering. This is called self-organization.

2.2 Collaborators

Collaborators include Professors Baartman, Schoorl, and Bartholomeus, from Wageningen University in the Netherlands. Collaboration involves data exchange which could be up to several GB per project. Earthcube work is consulted, but not actively participated in.

2.3 Software & Hardware Infrastructure

Local computation is utilized for any simulations. As the precision of data grows, it is expected that the data size will outpace local storage and ability to transfer data on network infrastructure. It is also expected that publication of results and sharing of data will move to more central locations.

Software consists of:

1. Microsoft Excel processing of results, and scripts that transfer 5-dimensional model data into R.

2. FTP or transport websites such as “WeTransfer”
3. Self-developed computer models
4. Agisoft Metashape for photogrammetry

3. Ganga Hettiarachchi, Soil and Environmental Chemistry

Research in Soil and Environmental Chemistry involves laboratory and field experiments on agricultural soils, contaminated urban soils, and mine-impacted soils/geomaterials in order to understand biogeochemical transformation of nutrients and potentially toxic elements and their role in controlling soil-plant transfer, mobility, and attenuation processes. Primary focus areas presently include:

1. "In situ" soil remediation involving the formation of stable solid phases, chemisorption, and phytostabilization to reduce soil-plant transfer of potentially toxic elements and/or reduce transportation of contaminated soils by air and water
2. Understanding complex redox transformations of potentially toxic trace elements and interactions between molecular level and macro-scale biotic and abiotic processes on the health of our soil/geo environments and water bodies
3. Determining reaction products of different P fertilizer sources in soils to understand their relationship to potential availability and plant uptake. The objective is to aid in the design of better and more efficient P fertilizers and P management practices
4. Evaluating the impacts of contaminants on food safety from urban gardens and other types of local farming activities on brownfield sites
5. Investigating the role of soil mineralogy and chemistry to aggregation and soil C sequestration in agroecosystems

3.1 Process of Science

Data collection, data analysis, data reduction, integration of experimental data obtained by various approaches (such as wet chemical, spectroscopy) or at different scales (such as micro-scale, macro-scale, field). Future work aims to impact & improve the integration of experimental data with modeling.

3.2 Software & Hardware Infrastructure

Local machines in the lab are used extensively to transfer data from multiple places (including national resources such as the Argonne National Lab Advanced Photon Source) to a single storage location. This departmental data storage/backup has a moderate capacity to store lab group data, but not enough to use effectively/extensively. There is currently no storage within the university.

Software Infrastructure includes:

1. Excel spreadsheets, Sigma plot spreadsheets, various other software compatible with Excel
2. FTP, Google, Microsoft
3. SAS, sometimes Genstat, Excel

4. Andres Patrignani, Soil Water Management

Our goal is to advance the science of multi-scale soil moisture monitoring and to find innovative applications of soil moisture information in agriculture and hydrology.

4.1 Process of Science

Soil science with an emphasis in applied soil physics. Better understand spatio-temporal patterns of soil water in the vadose zone of agricultural fields and small watersheds. Expand the applications of in-situ soil moisture information.

4.2 Collaborators

This research collaborates with:

- USDA-NRCS through nationwide environmental monitoring networks such as the US Climate Reference Network
- Kansas Mesonet
- Oklahoma Mesonet

4.3 Software & Hardware Infrastructure

Sensor networks form a critical component of this work:

- Set of 5 to 6 soil moisture and canopy cover monitoring networks in cropland fields across the state generating a total of 10 GB per station per year.
- Transects using a roving cosmic-ray neutron detector generating a total of 5 MB per year
- In-house maps of soil water for the state of Kansas at daily time steps and 1-km spatial resolution generating a total of 5 GB per year.

Computation is done using local to the lab resources (maintained by research staff), and some KSU resources which includes a virtual machine shared with the Department of Agronomy to host and store statewide maps of soil moisture for the state of Kansas. Additionally, there are several software infrastructure components:

- Matlab and Python for data analysis
- Data transfer is done within Matlab or terminal using scp command
- Javascript and NodeJS for web interfaces and server-side data processing.

The lack of a campus-wide Matlab license has been shown to be an ongoing source of complication for research. There are plans to Plan to explore Digital Ocean and Firebase for some projects.

5. Xiaomao Lin, Agricultural Climatology

This work involves agricultural climatology, climate science, and bio-atmospheric interactions. Goals include moving towards a real-time drought assessment and forecasting system for Kansas, including the Kansas weather library and mesonet program. Large-volume data sets are routine for this work.

5.1 Process of Science

The large-volume data sets (usually around a few GB to 50 TB depending on the coverage and data type/formats) include observation data and simulation data. The data sets are currently manageable but growing. Faster network and computational resources have the potential to impact output.

The simulation data are outputs of general circulation models (GCMs) that are used to forecast climate change. As an example, several of the 5-year assessments done by the Intergovernmental Panel on Climate Change (IPCC) have run DSSAT-type models using GCM-generated future weather as inputs to estimate the impact of future climates on cropping system yields around the world. In theory, one should be able to use such simulations to "design" what characteristics one would want to have in crop production. However, because the DSSAT-type models do not take any account of genetics, the resulting design can't be fully realized. Adding in this additional layer of information will help to generate the next generation of climate resistant crops.

5.2 Collaborators

This work relies heavily on outside data production that includes:

- NCEI (National Centers of Environmental Information)
- NCAR (National Center of Atmospheric Research)

5.3 Software & Hardware Infrastructure

Sensor networks are critical for data production and include:

- Trace gas analyzers including CO2, H2O and CH4
- Ultrasonic anemometers

Software infrastructure includes:

- WRF (Weather research and forecasting models);
- CLM (Community land models);
- Crop models (e.g. APSIM), not too much but students get into them
- Matlab, R, and Python

Currently most work is done internal to the university; some work is being put into learning about cloud computing options. As one of their users and/or collaborators, our computation and modeling operations from NCAR are not always smooth because of interruptions to the supercomputer's operation.

6. Daniel Flippo, Robotics

The goals of research include meshing the state-of-the-art robotic technology with conventional and non-conventional food production to move toward sustainably feeding the world past 2050. The problem of sustainably feeding the world has two constraints: producing enough yield to feed the population while doing it in a sustainable way to continue past 2050. Research thrusts range from conventional agricultural machines to autonomous vehicles and the tools and implements used by them.

6.1 Process of Science

Intelligently incorporated automation can provide a valid solution to deepen human-robot collaboration and meet our food, fuel, and fiber needs by better soil management, increased production, and responsible use of energy, water, and chemical products. There is much to learn in vehicle field dynamics, power requirements and alternative energy sources, networking, logistics, and autonomous precision farming. The paradigm in which we see the role for food production equipment can be drastically changed due to the opportunities and scale that these robotic vehicles allow. Small agricultural drones will affect the growing seasons due to their indefatigable nature and resilience to unfavorable weather conditions as well as bring a new precision to agriculture never realized before. These changes will impact and transform conventional food production with new possibilities in biodiversity, natural weeding, and pest management. Positive environmental impacts will be felt through use of hybrid power systems, better chemical and water management, and highly reduced soil compaction resulting in less erosion and chemical runoff.

6.2 Collaborators

Most collaborations are local to the University.

6.3 Software & Hardware Infrastructure

Local hardware resources are used for computation and design. Software consists of design packages and LabView for interaction with instrumentation. This telemetry data and soil sensing data is critical to the process of science.

7. Mary Knapp, Climatology

Research focuses on archiving, filtering, and making weather and climate information available to the university and public through the weather data library and the Kansas mesonet. Collaboration with CoCoRHS (Community Collaborative Rain, Hail and Snow) program, which is a cost-effective method to measure rainfall in the state.

7.1 Process of Science

The goals are to improve understanding of weather and climate, and increase the utility of weather and climate information for the citizens of Kansas, the region, and beyond.

Data is collected at the remote sites, transferred via IP based modems to a data server. The data is QC'd, and archived in various SQL databases. There are several methods of dissemination, including web services and REST data pulls.

7.2 Collaborators

The goals are to improve understanding of weather and climate, and increase the utility of weather and climate information for the citizens of Kansas, the region, and beyond.

Data is collected at the remote sites, transferred via IP based modems to a data server. The data is QC'd, and archived in various SQL databases. There are several methods of dissemination, including web services and REST data pulls.

7.3 Software & Hardware Infrastructure

There are 65 remote data loggers with cell phone based IP connectivity. They record 1minute, 5-minute, hourly, and daily weather data in an array-based data format. Storage capacity at each is 4.19MB.

In addition to this, software infrastructure includes: Microsoft SQL, Filezilla, and R. Efforts are being made to explore the use of cloud services such as Digital Ocean, GitHub, Gaug.es, DropBox, and Google.

It is a priority to achieve server/site redundancy in the case of infrastructure failure (power outages, Internet interruptions, etc.).

Appendix B - KSU Cyberinfrastructure Plan

This appendix presents the Strategic Plans for Kansas State University's Network and Telecommunications (NTS) for the period 2014–2018. The plans represent the strategic priorities necessary to guide NTS in its support of the University's growth in the coming years. This five-year plan outlines the major technology initiatives that Kansas State University expects to undertake in support of the University's plan for 2014-2018. These initiatives will help promote student success, ensure efficient administrative operation, improve the quality of the undergraduate and graduate educational experience, advance research and creative activity, and help ensure the privacy and security of the University community.

Communication Infrastructure

Supply a highly reliable, effective, modern communications infrastructure over the next 5 years.

NTS has made major investments in the network infrastructure over the last several years to support the University's operational needs and research goals. We have made improvements to the physical environments, including the fiber plant, building wiring, HVAC and power. We have seen an explosion in the demand for wireless service, and have responded by redesigning our network infrastructure to support the latest wireless technologies, including early adoption of 802.11ac wireless in buildings that were constructed over the last year. We have undertaken both large and small building improvement projects to upgrade aging wiring and switching infrastructure. We have anticipated the campus community's increased reliance on the network, and are making significant improvements in the core infrastructure. In Summer 2014, we will be replacing the network core with 2 high-speed, high-availability routers. Initially, these routers will be connected at 40Gb, with the capability of going to 100Gb to meet future demands in bandwidth. At every juncture, we have worked to improve the security of our network. Efforts have ranged from physically securing network closets to employing a series of firewalls, IDS,VPN, and NAC solutions.

Our mission going forward is to build on the progress we have made in both the physical and logical aspects of the network. Over the next 5 years, we anticipate an ever-increasing demand for highly reliable, robust offerings in both wired and wireless services. We already have researchers requesting 10 Gb connections to their desktops and 40Gb aggregation at the core, to facilitate transfers of very large data files. These researchers are also utilizing lab equipment that requires secure, reliable wireless connectivity. In addition, professors are requesting robust wireless connectivity in the classroom, in order to facilitate presentation delivery and collaborative participation from their students.

As new technologies become available, the network infrastructure has to be both resilient and responsive. Our goal is to keep abreast of these technologies and evaluate how best to prepare for and incorporate these advances within our network.

Critical Success Factors (measures the degree of success over the next 5 years):

- Maintain network up time at 99.8%.
- Complete the update of the campus' core network infrastructure.
- Provision advanced high-speed research networking.
- Replace all shared Ethernet Technology with switched Ethernet Technology.
- Extend a 40Gbps core backbone with 1/10/40Gbps distribution to all buildings.
- Make available 1 or 10Gb desktop technology to requesting departments.
- Replace all aging wireless technologies and improve density deployments with 802.11ac.
- Construct Telecommunication closets in the buildings that do not meet the standards.
- Deploy production IPv6 services.
- Upgrade campus fiber backbone architecture and components.
- Offer HPC as needed to requesting researchers
- Research the convergence of voice, data and video technologies.
- Provision a unified communications messaging infrastructure.
- Execute a plan for disaster recovery and business continuity.
- Upgrade 800 MHz emergency communications system.
- Evaluate options for providing future voice services.
- Continue to upgrade NTS' HVAC, power, and security infrastructure.

Network History

2001-2002 Campus Core network consisted of A Catalyst 6509 and 7513 Router in the Power Plant, and Catalyst 5500 in the Hale Data Center, a Catalyst 5509 in Vet Med, and a Cisco 7000 in West Hall. The backbone was collapsed, where Vet Med and Hale connected back to the Power Plant 6509 by dual gigabit ethernet interfaces. The Power Plant 7513 and West Hall 7000 connected back to the Power Plant 6509 via a single fast ethernet interface. The Salina Campus was connected to the Manhattan Campus via an ATM DS-3 connection, with a T1 for backup. The University had a DS-3 which was shared for commodity Internet (15 Mbps) and Internet2 (roughly 20 Mbps). For building network connections, there were 5 at gigabit ethernet, 34 at fast ethernet (100 Mbps), and 34 at ethernet (10 Mbps). Finally, the Konza research site was connected by T1.

2002-2003 Added a second Catalyst 6509 to the Power Plant and built a mesh between 4 core locations, using gigabit ethernet links. K-State's Internet/Internet2 connection was changed from a DS-3 to an OC-3, through which K-State had 40

Mbps of commodity Internet and 95 Mbps of Internet2 access capacity. The IDEA Center moved off campus and was connected via a T1. Extension offices and Barton County Community College were added as T1- connected sites. Dole and College Court became the first two buildings dual-homed to core 6509s. Connections to buildings consisted of 6 via gigabit ethernet, 42 via fast ethernet (100 Mbps), 27 ethernet (10 Mbps), and 10 via a T1. 802.11b wireless was installed throughout Hale library and the Student Union, resulting in 17 campus buildings have at least partial wireless network service. Planning began to implement a pair of redundant border routers.

2003-2004 Added Cisco VPN3030 concentrator for securing remote access. Wireless networking had grown, and there were a total of 84 wireless access points installed across campus. There were 14,537 computers connected to the campus network. The Alumni Center was added to the campus network via a gigabit ethernet connection. A pair of locally built P2P systems were installed to enforce a ban of those types of applications. These systems made K-State's Internet connection usable again, as it had been saturated. The network backbone remained a gigabit ethernet mesh, with star topology from each of the 4 core locations. Building connectivity consisted of 9 gigabit ethernet, 60 fast ethernet, 11 ethernet, and 12 T1 connections. Notably, the residence halls were upgraded from ethernet to fast ethernet connections. Internet and Internet2 connectivity remained consistent, as did the connection to the Salina Campus. A Catalyst 4506 was installed in West Hall to replace the legacy router in conjunction with the release of an in-house developed Housing Network Registration system. There were support issues with that system, requiring it be removed while having issues corrected. In the data center, Cisco Local Directors were installed in front of some systems for load balancing capabilities.

2004-2005 Installed Intrusion Detection Systems at the border. There were 17,964 systems connected to the campus network. Wireless installations had grown to 172 access points across campus. Wireless coverage remained somewhat spotty, as installations were dictated by those departments that could afford to buy the necessary switches and access points. The number of devices on the campus network grew to near 20,000. Operation PC, an effort by IT to help get systems patched, antivirus software installed, and that systems had secure passwords. It was an intense 3 day initial effort by central IT, campus IT members, and Housing. Operation PC did pay back some dividends in that some of the compromises experienced earlier declined.

2005-2006 Evaluation of products for setting up an enterprise-wide NOC was conducted. Budget was not available to implement, though the computers and furniture for Operations was ordered, and nagios was used to give them visibility into any network problems. The GlobalFlyer flight out of Salina early in the year required some upgrades to the Salina router, including the addition of an additional BGP peering with Cox for redundancy. A Catalyst 6509E with Sup 720 was installed

in Hale, and the Catalyst 5500 in Hale was moved to Vet Med. In the data center, F5 Networks BigIP load balancers were installed in front of several core services.

2006-2007 Many changes to the network occurred. By this time, our connection to KanREN was via a pair of gigabit ethernet connections, and our ATM footprint was reduced to the Salina Campus connection. Commodity Internet capacity had ben increased to 83 Mbps and Internet2 capacity was the remainder of the ATM connection from the KanREN equipment back to Kansas City. With an exception of 4 buildings, all others were connected at fast ethernet or gigabit ethernet speeds. In the Residence Hall environment, a NAC, the Bradford Networks system, was implemented. This summer-long project was accompanied with firewall service module implementation in the Hale and Power Plant core 6500s, the addition of Cisco WiSM controllers, and the installation of 150+ wireless access points to provide wireless access in the Residence Halls. In addition, Telecommunications assumed management responsibility for the switches in the Residence Halls, primarily 3Com switches. The Core 6509 was replaced in the Power Plant with a Catalyst 6509E to boost performance and port density and capacity. The Anderson Hall network equipment was upgraded from hubs to switched 10/100 and in a few locations, switched 10/100/1000 connections, though much of the potential was limited by cabling, which was largely Cat 3 cabling. Many lightweight APs were installed across campus during the year. Client counts by this point were nearing 25,000.

2007-2008 Installed Extreme switches in Nichols Hall, rewired West side (Computer Science) on Nichols Hall. There were issues with this installation, and it was discovered that there were multicast-related issued between 3Com and Extreme switches that would result in the 3Com switches flooding those packets sent by the Extreme switches occasionally, disrupting various locations on the network. Extreme equipment was installed in the newly built and remaining Jardine Complex buildings, as well as the Residence Halls. At that point, wired connections were combinations of switch fast ethernet and gigabit ethernet drops. In August, 4 Extreme 12K switches were installed in the core, between the Cisco core devices, in a ring topology. Numerous software issues were run into, and several occasions if traffic causing the Cisco core switches to disconnect, effectively taking down the campus network. This pattern continued into the fall, culminating with the request for an external audit to be conducted early in 2008. Wireless was added to the Jardine complex, and many locations across campus. In addition, K-State purchased a connection from AT&T through which we could retain connectivity if KanREN or their upstream ISPs had outages, after some significant service disruptions, primarily KanREN's upstream ISPs.. KanREN's backbone design had changed to a dual ring, with a pair of gigabit ethernet connections between their core nodes. In addition, a decision was made by the KanREN board to open up access to the excess bandwidth they had been purchasing for redundancy. They effectively removed rate limits at our peering point, which allowed consortium members to access the full amount of commodity Internet capacity. Regarding wireless, 802.11n access

points began to be installed, with Nichols and the Union being the first locations. The P2P filtering at the border was changed to a Packeteer appliance, rather than the homegrown implementation used previously, mainly due to capacity and support issues. Roberts Hall was connected to the network, and several firewalls installed within the building.

2008-2009 The Calence audit was wrapped up in January, with the recommendation to remove the Extreme equipment to reduce complexity of the network and increase stability. Around March, the Extreme equipment was removed from the core, effectively placing the core back to its state before the Extreme installations. In addition, and pair of 6509E switches were installed in the data center in Hale. They included firewall service modules, and allowed the data center traffic to be self-contained in the data center, where it had previously hit the core since there were only layer 2 devices in the data center. In addition, the connections were all capable of gigabit ethernet. The Nichols Hall network was changed out to a Cisco infrastructure, with the remaining sections, primarily the east side, also being rewired and upgraded to switched connections. Calvin Hall received a network upgrade to gigabit ethernet, primarily to help them work towards implementing a virtual desktop environment. Most of the previously installed 802.11b access points were upgraded to support 802.11g, and the non-upgradable APs were replaced. Most access points on campus were now lightweight, with 4 WiSMs being in place. The connection to Konza was also upgraded to support 802.11g.

2009-2010 The connection to Salina was changed to ethernet, so the old, unsupported ATM infrastructure could be removed from service. The K-State Parking Garage and the Leadership Studies buildings were connected to the campus network. KanREN installed a second Foundry on campus, in Hale, to remove one single point of failure for our Internet connection. We worked with the Police to implement a mobile VPN environment so they could perform checks from laptops in their cars. Wireless use had grown, and the Residence Hall wireless offering was broken up into 4 "zones", many behind the firewall, to allow for the growth rate in wireless clients. Moore Hall was the first Residence Hall to be upgraded to 802.11n access points. Several switches were installed in the Vet Med complex to increase capacity to gigabit ethernet. Many building upgrades were completed, with switch replacements and/or installations. Building connections were upgraded from fast ethernet to gigabit ethernet in several locations. As the Internet service through KanREN was very stable, the the 70 Mbps AT&T service was not a feasible backup anymore, it was removed from service. In its history, it had carried us through what otherwise would have been some service disruptions.

2010-2011 The last of the ethernet-connected buildings were upgraded to either fast ethernet or gigabit ethernet. Performance problems with Residence Hall wireless surfaced, and after a great deal of troubleshooting, was determined to be due to the Packeteer installed at the border. Evaluations were conducted which

resulted in the decision to install the Procera for P2P filtering. The Bradford NAC appliance continued to be a time sink to support, and we began evaluations of other solutions, and are planning an installation in the spring of 2011. Wireless use has grown and now includes 1,300 access points across campus. An Aruba pilot was conducted in the fall, with installation in Willard being completed, installation in Olathe at this time, and installations in Hale, Fairchild, Eisenhower, and Umberger planned. The core remains a gigabit ethernet mesh, with 40+ buildings attached to the campus network via gigabit ethernet, nearly 40 connected via fast ethernet, and 8 locations connected via T1.

2011-2012 In the Summer of 2011, the West Hall router, which serves the Jardine complex and the Residence Halls, was upgraded to a 6509E chassis with redundant Sup2Ts. A new NAC solution – SafeConnect – was also installed at West Hall. The Jardine and Residence Halls subnets were moved to private 10.132.x.x space. The Aruba wireless rollout continued, increasing density and coverage in Ahearn, Calvin, Chem/BioChem, Call Hall, Cardwell, Fairchild, King, Parking, Military Science, and Umberger. By the end of the Spring semester, 1,500 access points were operational across the Manhattan and Olathe campuses. KSU Wireless and KSU Guest were divided in North, Central, and South regions, in order to reduce the broadcast domains. Multiple high tech classroom installations were completed during the year. Fairchild, Eisenhower, Willard, and Cardwell re-wiring and network infrastructure refresh was completed. PCI firewalls were added in both Hale and Power Plant. The border routers were upgraded to 6509E chassis with Sup2Ts, and border firewalls were added. 10Gb line cards were added to Hale and Power Plant, in order to establish a 10Gb backbone infrastructure between KanRen, the border, and core locations. West Hall was also connected to both Hale and Power Plant at 10Gb.

2012-2013 Several service initiatives were launched during this time, including support for an OpenFlow lab in the College of Engineering, Apple iPad/Apple TV initiative in the College of Education, EDUROAM membership, High Tech Classroom installations, and the initial phases of desktop virtualization in the University Computing Labs. Also at the beginning of 2012, 26 buildings were connected to the core at 100FX. Traffic in many of these buildings was bursting over 100Mbps during peak periods. Single mode fiber runs were completed, and 7 of these buildings have been moved to 1Gb. The remaining 19 will be connected at 1Gb by Summer 2013. In order to meet current demand for 10Gb connectivity between specific research collaborations on campus, we plan on adding 10Gb line cards in the Hale and Power Plant core 6509s. Once the 100Mb building connections are moved to 1Gb, the 100FX cards will be replaced with 10Gb line cards. We continue to improve wireless densities, to support ever-increasing demand. Additionally, we have installed Clear Pass to facilitate wireless guest management and on-boarding campus users/devices.

2013-2014 Sought funding for the replacement of two core routers, capable of 10/40 Gb backbone and building aggregation. The replacement will be completed Summer, 2014. All the campus buildings that were connected at 100FX were upgraded to 1Gb connections. Additional 10Gb line cards were purchased and installed in the core routers. The Data Center, Throckmorton, Durand, and Nichols connections were upgraded to 10Gb. Completed network design and implementation for new construction and renovation projects, including the Feed Tech Mill, Stanley Stout Center, Human Ecology Research Center, English Language Program at Wildcat Landing, Danforth Chapel, Mosier 2nd floor addition, Moore Hall network expansion, Rowing Center, West Stadium Center (WSC) and Honors House. Installed 802.11ac technology in the WSC and Goodnow Hall, and continued the upgrade of aging wireless technologies and improving density deployments across multiple campus buildings. Upgraded necessary infrastructure to support the VDI initiative, and Housing's thermostat control system.

Appendix C - Kansas State University Networking Diagram

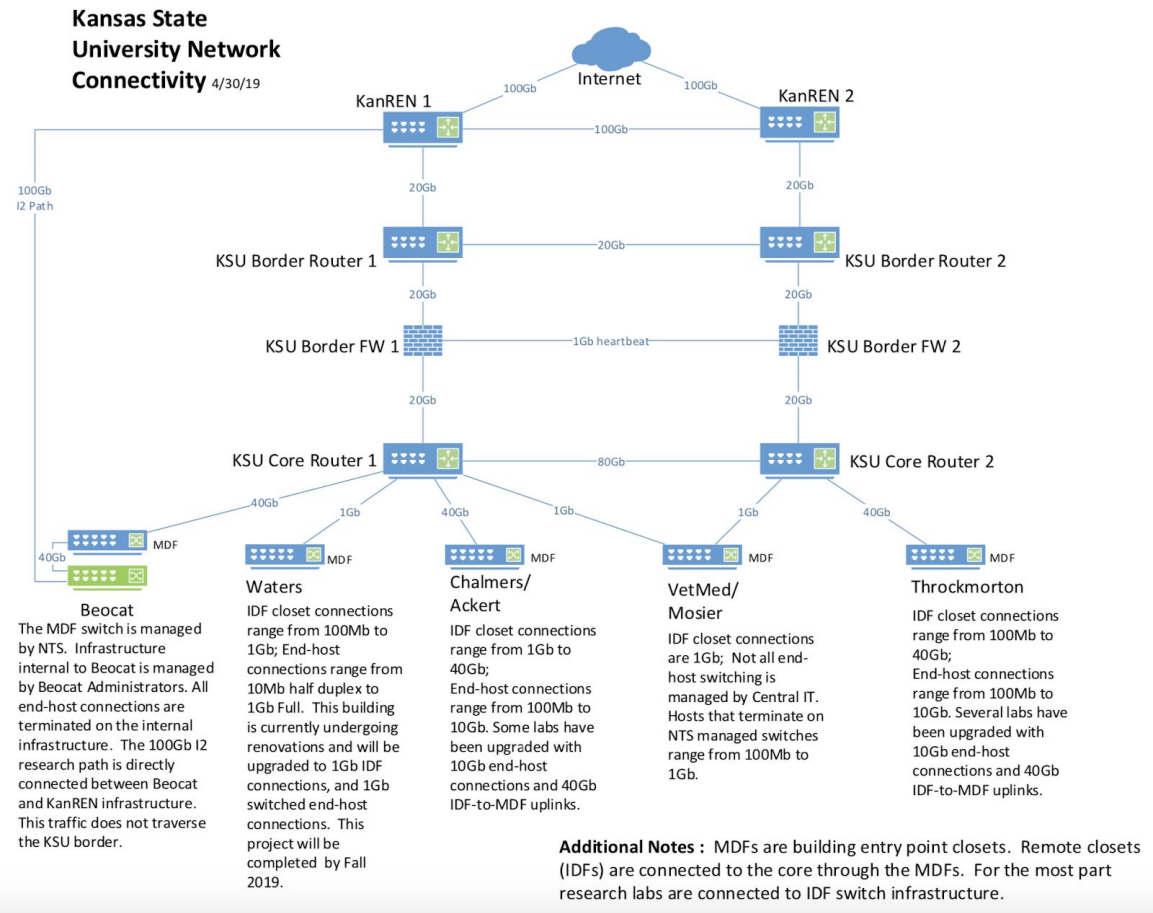

Figure 5 - Network map for Kansas State University.

Appendix D – KanREN Network Map

Figure 6 - Network Map for KanREN[10].

[10] https://www.k-state.edu/media/newsreleases/jun16/hypercore61516.html

Appendix E - Great Plains Network Maps

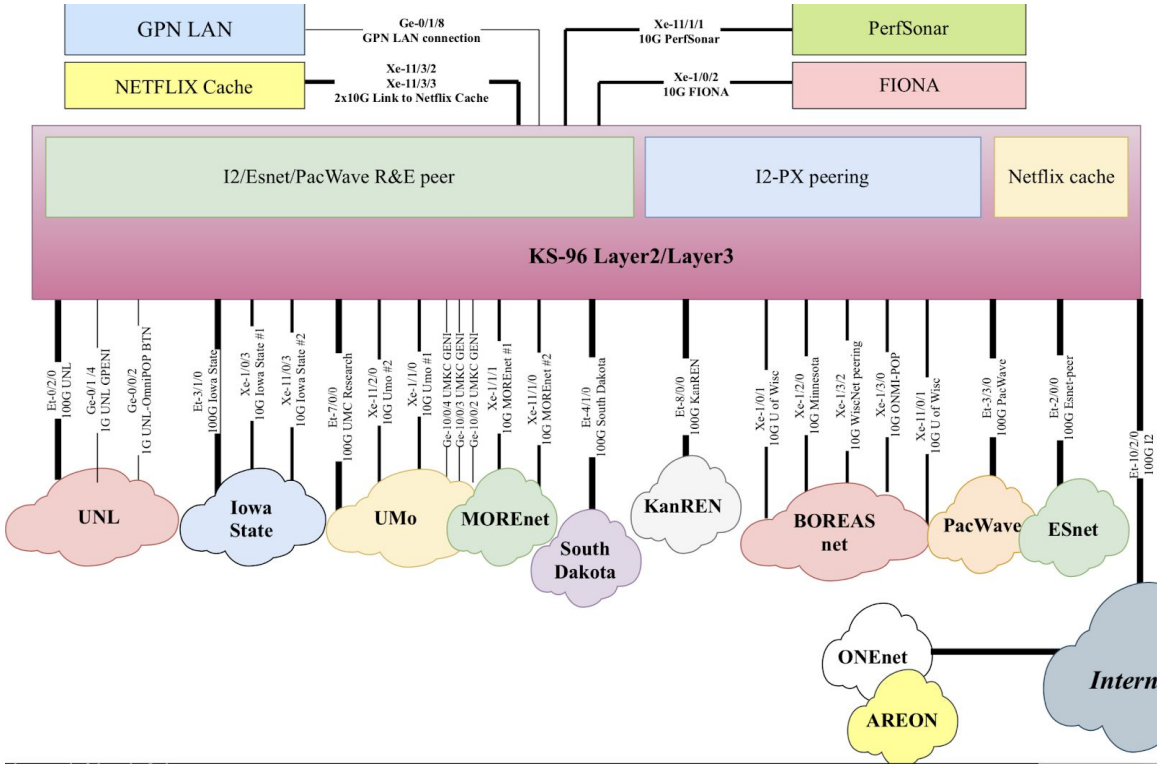

Figure 7 - Logical connection diagram for the Great Plains Network.

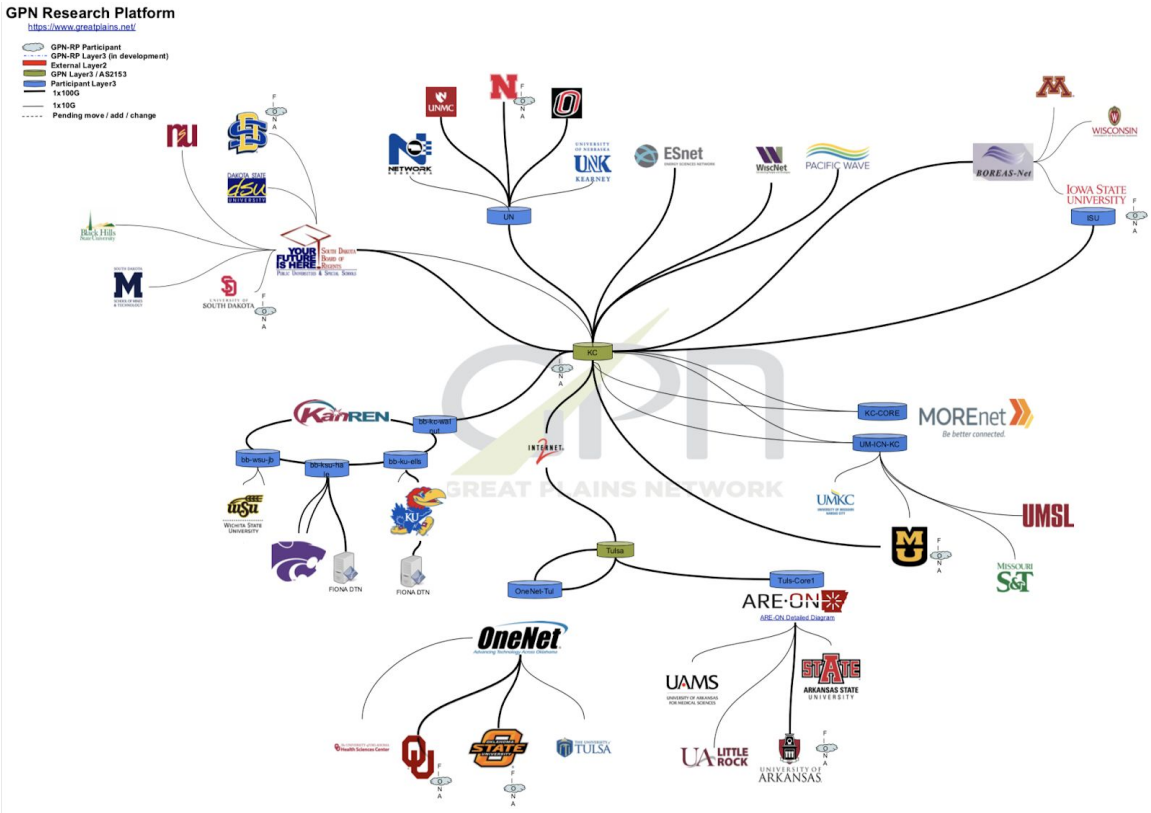

Figure 8 - The GPN Research Platform.

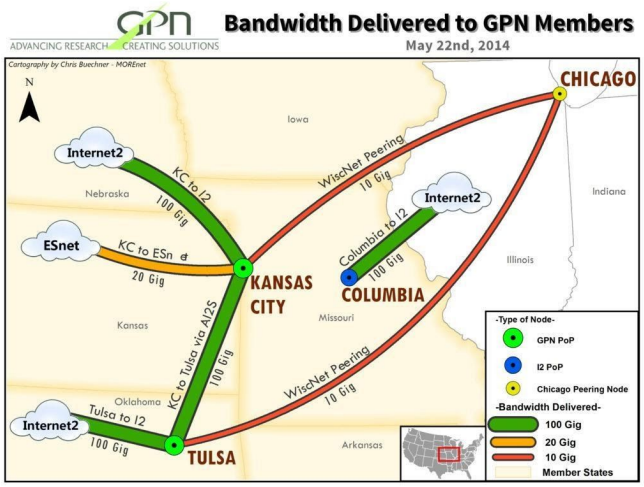

Figure 9 - The GPN bandwidth allocation.

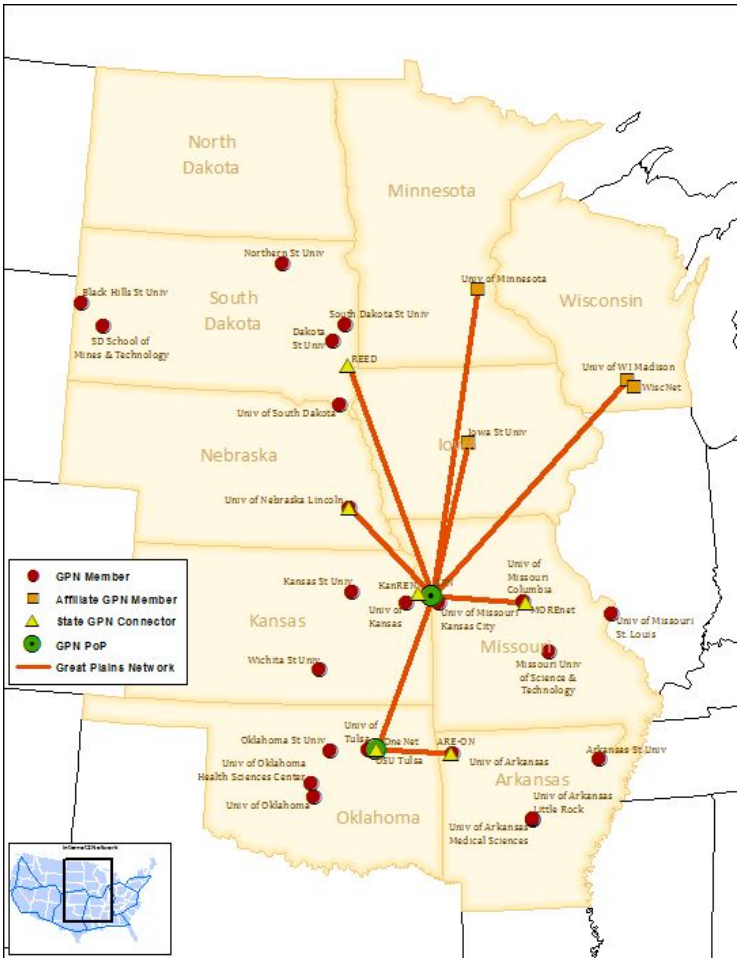

Figure 10 - A map of the GPN members.