Managing Farm Risk Using Big Data

A guide to understanding the opportunities and challenges of agricultural data for your farm

Dr. Shannon L. Ferrell
Professor, Agricultural Law
Oklahoma State University Department of Agricultural Economics

Dr. Terry W. Griffin
Assistant Professor, Precision Cropping Systems Economics
Kansas State University Department of Agricultural Economics
Project Funding
This project was funded by a grant from the United States Department of Agriculture National Institute of Food and Agriculture administered through the Southern Risk Management Education Center under Agreement Number 21667-19 in the amount of $49,776.

Acknowledgments
The authors gratefully acknowledge the contributions of Professor Paul Goeringer, Extension Legal Specialist for the University of Maryland’s Department of Agricultural and Resource Economics and Ms. Ashley Ellixson, Vice-president of Legal and Environmental Affairs for United Dairymen of Arizona for their contributions to the handbook manuscript. The authors would also like to thank Dr. Brian Arnall of the Oklahoma State University Department of Plant and Soil Sciences and Dr. Paul Weckler for their contributions and review of these materials, and Emily Carls and Jared Cullop, graduate research assistants in the Kansas State University Department of Agricultural Economics for their editorial work.

In Memoriam
This work, and indeed much of the technological infrastructure that makes much of the precision agriculture tools mentioned in it possible, would not be possible without the contributions of Dr. Marvin Stone, Oklahoma State University Department of Biosystems Engineering. Dr. Stone was killed in the Oklahoma State University Homecoming Parade accident of October 24, 2015. The authors hope this work furthers Dr. Stone’s legacy of contributions to the agriculture industry.
Introduction

Why do we care about agricultural data?

“Agricultural data”, “small data,” “big data” - they all sound like industry buzzwords that may get thrown around too much. To some people, these sound like important opportunities for farmers to profit. To others, they trigger concerns that information about the farm might fall into the wrong hands or be used to create profit for others at their own expense. With this handbook, you will be able to separate fact from myth and learn how you can determine the data tools that can help your operation.

The discussion of agricultural data has grown as the technology to collect agricultural data has expanded rapidly in the past few years. Some technologies that form the foundation of agricultural data collection, such as global positioning systems (GPS) have been used in agricultural equipment since the 1990s. However, an explosion of technologies in the 2000s and 2010s allow farmers (as well as other parties) to monitor an ever-growing number of parameters about their equipment, crop environment, and the crop itself and to instantaneously transmit that data to crop consultants and data aggregators. At the same time, new research and software tools allow farmers to collect and analyze unprecedented amounts of information about their own operations. Moving beyond the farm gate, an increasing number of “Big Data” tools allow consultants, equipment manufacturers, input providers, and others to aggregate information across hundreds and even thousands of operations to evaluate trends and benchmark performance.

Increasingly, data collection processes are automated by hardware and software embedded in agricultural equipment. Those systems may automatically transmit data to the equipment manufacturer. As a result, farmers may not even realize they are participating in data collection efforts since some equipment arrives at the farm “switched on” and transmitting to a party other than the equipment owner. This means not only learning about how to actively take advantage of the available farm data tools, but learning about the data collection efforts that may already be affecting farmers without them even knowing it.

In the end, a tool is neither good nor bad, and neither valuable nor worthless; what matters is how the tool is used. With this handbook, we hope that you can identify, select, and implement the agricultural data tools that can help you manage the risks associated with farming, improve profitability, and help make your operation the best it can be.
Chapter 1

How is agricultural data collected and analyzed?

Before exploring the tools to manage and analyze agricultural data, it is important to understand how agricultural data is collected. It also helps to understand the terms producers, engineers, and technology providers use in discussing these technologies.

1.1 The language of agricultural data and how it is collected

This handbook has already used the term “agricultural data” a number of times, but what does that term even mean? In this handbook, we will use the term “agricultural data” to discuss data that is collected about a farm or ranch. Agricultural data can include data that is manually recorded by farmers or ranchers, passively collected by sensors and other systems embedded in agricultural equipment, hand-held systems, or remote sensing systems such as aircraft or satellites. Agricultural data can also be created from the analysis of other data. Data can be collected intentionally by the farmer (yield monitor and soil sampling data), passively to the farmer (time-in-motion of equipment, fuel consumption), or about the farm (that is, information that is observable about the farm from an outside perspective - how frequently the principal operator visits the coffee shop, what color the operator’s truck is, etc.).

As you can see, the term “agricultural data” or “farm data” is a big umbrella even when focused on the production side of agriculture alone. Underneath that umbrella are a number of smaller categories of data based on which part of your farm they help describe.

“Small Data” consists of data isolated to the fields or farms where the data originated. Farmers who use information technology to conduct their own on-farm experiments, document yield penalties from poor drainage, or negotiate crop share agreements are making use of “Small Data.”

For data to be considered “Big Data,” the data must possess the following traits:

**Volume:** The amount of data is so large it cannot be stored on one physical volume such as a hard drive; put another way, the analytical software must “go to” the data rather than the data being “brought to” it.

**Velocity:** The data comes at higher rates than can be handled by a single computer.

**Variety:** The data comes from a broad array of sources including data automatically gathered by sensors and software as well as data manually entered by users. This also means the data may come in a variety of formats with a lack of structure or design to it.

**Veracity:** The data analytics are required to consider the accuracy and credibility of the data, especially considering the “variety” considerations mentioned above and the potential for mis-calibration of automated data collection systems or user error (and bias) coming from manually-entered data.

**Machine data:** You can think of “machine data” as data that comes from farm equipment looking “inside itself.” Consider your car or truck. Since 1996, all U.S. cars and trucks have been required to have an on-board diagnostic system that includes a port enabling
the owner or a technician to connect a computer to the system to collect data about a number of factors about the mechanical and electrical systems of the vehicle such as engine temperature, oil pressure, transmission RPMs, fuel use, and so on. Increasingly, agricultural equipment manufacturers have implemented similar systems on tractors, combines, and other power equipment.

When you are trying to keep your equipment operating as efficiently as possible and minimize downtime, information about the “health” of your equipment is very useful. Machine data can do more than provide information about when you need to change your oil; it can tell you much about how you could run your equipment more efficiently in the field while still accomplishing the same work (or even more).

In recent years, though, many equipment manufacturers have recognized the same systems that look inward to the machines could also collect data that looks “outward” into the crop environment. For example, you could consider a combine harvester yield monitor as collecting machine data about its threshing system, but it is also collecting information about how much grain is coming from a field – one of the most critical pieces in the entire life-cycle of a crop. The seed population information from a grain drill is machine data about the drill’s throughput, but it is also the foundation of the next year’s crop and represents a crucial management decision for you.

**Telematic data:** A textbook definition of “telematics data is “the branch of information technology which deals with the long-distance transmission of computerized information.” In plain English, though, telematics is the field that studies how computers can collect data and transmit it to a remote collection point. As a result, telematic data would include any data collected by computer and transmitted wirelessly. For example, data are collected on many modern different kinds of agricultural equipment by manufacturers who use cellular-enabled telematic systems to send that data back to the manufacturer. Put another way: a car sold as new in 1996 had the ability to collect data about its health that could be read with a handheld data reader. In more recent years, telematic systems (for example, OnStar on GM vehicles) have transmitted that data directly to the manufacturer via cellular modem.

**Agronomic data:** “Agronomic data” are data about crops, including information about the condition of the plants themselves and information about the crop’s soil environment. While some data are collected automatically without any intervention by the farmer, some farmers actively collect agronomic data regarding farming practices and productivity. For example, a substantial percentage of farmers have invested financial capital, and more importantly, human capital into collecting geospatial agricultural data (see the definition of “geospatial data” below) such as yield monitor, soil sampling, and as-applied data. However, agronomic data may also be collected by agents other than the farmer. Perhaps the most common example is grid soil sampling. Grid soil sampling can be performed by a dedicated soil-sampling service or provided by local agricultural retailer. Either way, when someone other than the farmer collects farm data it logically follows that the service provider has access to a copy of that data; that will be a point of discussion later in this handbook.

**Environmental data:** “Environmental data” refers to the data about the surroundings of the crop, most frequently about the subsurface (soil) environment and above-surface (atmospheric) environment. Soil and weather data are discussed separately below.

**Economic data:** Many farmers might not even think about “financial and economic information” as even being data. However, after agronomic data, it is probably the data type most often intentionally collected and analyzed by farmers. Economic data includes input cost data regarding items such as seed, fertilizer, machinery, and labor as well as the prices received for outputs. Economic data also includes data not only about individual crops at the field or enterprise level but also about the financial health and
performance of the operation at the whole-farm level. 

**Farmer data:**

The term “farmer data” might sound like data collected by the farmer, but in this discussion it is actually data about the farmer. Any time you sign up for an online service, you likely click on a button saying you agree to the “terms and conditions” of a user agreement that says, among other things, the service can collect data about you and how you use the service. Agricultural technology providers often do the same thing; in some cases, this is to help them better serve their clients, and in some cases it is to determine how to sell more products or services to the customer and/or to sell that data to third parties (and note, these activities are not necessarily mutually exclusive). Farmer data could include a wide variety of factors including:

- Their customer history: what products or services the farmer uses and patterns in their use, such as when the services are purchased or in what quantity
- Management patterns: what production practices the farmer uses, the mix of crops and other agricultural enterprises the farmer has on their operation
- The payment history of the farmer: are bills paid on time, what type of payment the farmer uses
- Personally Identifiable Information (PII): (PII is discussed separately below)
- Equipment compliment: what machinery does the farmer use, and which manufacturers made it; this could also include machine data about how the farmer uses the equipment and how often new or different equipment is acquired or disposed of through sale or trade.
- Other vendors: With what other groups does the farmer do business, such as input sellers, crop scouts, consultants, and universities?

**Geospatial data:**

“Geospatial data” (sometimes referred to as Geographic Information System or GIS data) is data that is somehow linked to a geographic location. As mentioned above, grid-sampled soil data consists of taking samples along a grid in a field, and adding location information to each soil sample. When the sample analysis is coupled with the location of each sample, it creates a soil map - that soil map is made from geospatial data. Using yield monitor information that is georeferenced with the location of each yield measurement by the combine’s GPS system creates geospatial data that is then used to create a yield map. Geospatial data is critical to a number of precision agriculture applications, such as variable-rate application of agricultural inputs, because it ties a crop environmental condition (such as plant-available nitrogen) to a location (a coordinate within a field) that allows for the location-specific application of an input (here, an increase or decrease in the amount of nitrogen applied).

**Personally identifiable information (PII)**

There are several definitions of PII, but one used by the federal government is “information that can be used to distinguish or trace an individual’s identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual.” PII can include information publicly available such as phone numbers and addresses. Some PII is necessary to companies to help them provide the services clients request. For example, it may be needed to help the company identify their accounts, i.e. to help them distinguish the account of Joan Smith from Sam Jones. However, some parties try to collect enough PII (such as birth dates, passwords, Social Security numbers, and so on) to assume the identity of another party (“identity theft”) so they can steal money directly from the person or use funds to purchase items on their accounts. How a company handles PII is often a critical concern to farmers before they enter into any data sharing agreement.
Soil data: Soil data can mean information about several traits of the soil environment (or subsurface) for a crop, such as soil texture, nutrient amount and availability, moisture, and others. Some soil properties do not change very quickly, and information about them can be collected well before or even well after planting, growth, and harvest of a crop such as soil texture, structure, depth of topsoil, and organic matter content. Other soil properties can vary substantially with time and the state of those properties at critical points in crop management (such as planting, germination, and other important growth stages) could influence how the farmer makes a number of decisions. As a result, more and more sensing systems are being developed to help provide instantaneous measurement of these traits.

Weather and climate data: Weather data could be regarded as information about the “above-surface” environment of a crop. This can include information about temperature, atmospheric moisture (humidity), winds, and sunlight. Weather data should also be distinguished from climate data. Weather data is generally considered to be information about the instantaneous condition of the atmosphere (for example, what is the temperature at this moment), while climate data is weather data measured over long periods of time (for example, what is the average daily high temperature for the month of May). Historically, weather and climate data have been one of the primary sources of information developed and distributed by public institutions, with significant effort made to analyze that information in a way that is useable by farmers. To date, many farmers have relied upon these publicly-available weather and climate data sources. However, farmers now have an increasing number of options to purchase “personal” weather stations that can be deployed at each field and use telematic systems to transmit instant weather conditions (and to accumulate very specific climate data). Weather and climate data are also archived by a number of public institutions, giving farmers many options for reviewing historical data.

1.2 How can agricultural data be collected?
You can think about agricultural data collection by asking who is collecting the data - the farmer, someone working on behalf of the farmer, or a third party who may not have a direct relationship with the farmer. Of course, each party collecting the data may have many different ways of collecting that data, but consider some of the examples included below.

1.2.1 Data collected by the farmer for the farmer
Farmers have been keeping farm records for themselves since they have had access to a rock and a chisel; some of the earliest human writings are records of crop and livestock production. Historically, the vast majority of farm records came from crop notebooks and financial ledgers as farmers kept track of plantings, input applications, harvest yields, and expenses. The dawn of the computer age enabled farmers to expand their record keeping processes and more quickly analyze data with the help of spreadsheets, searchable records, and downloadable information from consultants, government agencies, and financial institutions.

Even farmers who were extremely conscientious about keeping records sometimes had difficulty finding time to maintain their farm information when it had to be recorded manually, even with the aid of a computer. Today, though, more and more farm equipment automatically collects agricultural data with or without the deliberate intervention of the farmer. Understanding how farm equipment generates and transmits all this data is fundamental to understanding the “explosion” of data applications in agriculture.

The Society of Automotive Engineers (now called SAE international or “SAE”) created a “standard” (a set of specifications for an electrical or mechanical system) called SAE International Standard J1939 that governs how most telematics systems on tractors, combines, and other self-propelled agricultural equipment work. J1939 was originally designed as a standard for the operation of Controller Area Networks (CAN) used in on-road diesel trucks but has since extended to on-road and off-road diesel engines, including agricultural equipment. You may have heard the
term “CAN bus” as well - this refers to the system that governs how the CAN communicates with systems on the equipment. The J1939 standard describes how Electronic Control Units (“ECUs” – electronic devices that receive information from sensors embedded in the equipment and issue control commands to a system such as the engine or transmission – exchange information with other ECUs) throughout the equipment via a physically-connected network typically referred to as a “bus” - hence the term “CAN bus.” If you think about your body, your fingers have nerve endings that can tell you if something is hot or cold, rough or smooth. Those nerve sensations are carried to nerve nodes, which are then sent along your nervous system to your brain. Think of ECUs as collecting those nerve impulses and sending them along the CAN bus to a central computer (such as the engine control computer) for analysis and response.

In a piece of equipment like a tractor, the tractor’s bus relays both information (such as “current engine RPMs are 1,500”) and commands (“increase throttle to setting 17”) in real time by allowing multiple ECUs to use the same wiring system through a process called “multiplexing.” A CAN manages the use of the bus and defines a “common language” for its messages. Many equipment manufacturers are worried about third parties issuing commands to the system (which would permit a third party to “hack” the system), and as a result, many of the commands transmitted through the bus and the CAN use proprietary formats. This means the “common language” on the system is unique to that kind of equipment. This provides an important layer of security and safety for the equipment. The communications procedures of J1939 enable ECUs to automate many functions of the equipment, improving performance and reliability while also making more information available to the operator through dashboard displays that connect to the network.

Historically, telematics systems were intended to look “inward” to the machine itself to make the equipment work more efficiently and/or to give the operator more information about the machine (this is sometimes referred to as “machine health” information). That is changing though. Telematics systems continue to evaluate more and more information that is external to the machine that could impact its performance through “machine-vision tools” such as radar and image processing that tell the equipment what is around it and how it is moving relative to the vehicle as well as tools to sense the environment around the equipment such as air quality measures, temperature, humidity, and wind.

Perhaps just as important as the ability of a piece of equipment such as a tractor to sense both it’s “internal” and “external” environments is the ability of the equipment to automatically share the information they collect via wireless communications. A growing number of tractors, combines, and other self-propelled machines roll off the assembly line with cellular modems attached to their CAN bus. These modems continuously provide data to designated recipients such as the equipment owner and/or a dealer, enabling them to quickly diagnose machine health along with equipment failures or even prevent them. Generally, telematics systems do not accept outside commands via these cellular modems, but other industries have modified that functionality for a number of management purposes. For example, trucking companies (for whom the J1939 standard was initially created) have the ability to remotely limit horsepower to accommodate engine warranty specifications, to increase trucks’ horsepower output in mountainous regions, or reduce maximum speed when the vehicle is in a state with a lower speed limit. Agricultural operations could use the same equipment management procedures, although no major agricultural equipment manufacturers currently offer such remote management systems.

This brings us to how agronomic data can be generated and collected today. Farmers frequently generate agronomic data from implements pulled by their tractors or input from other sources (including the farmer’s own inputs for parameters such as seed variety and desired seed population rate or fertilizer type and application rate). This means that the tractor’s sensors must not only communicate with each other, but that the tractor also needs to be able to communicate with a planter, a fertilizer sprayer, and other implements. This means that all these systems must also have a common language. To this end, the International Organization for Standardization (ISO) created ISO Standard 11783, sometimes called the “ISOBUS” standard. ISO 11783 applies to self-propelled equipment (such as combines) and tractors, and defines two buses for its networks: a tractor bus and a “tractor/implement” bus. The tractor bus reflects the J1939 bus discussed above (indeed, the ISO 11783 standard builds upon the J1939 standard; in essence an ISO 11783 tractor bus is typically a J1939 bus on an agricultural vehicle), connecting the systems operating the tractor or the powertrain of the machine. The tractor/implement bus extends from front of the tractor or self-propelled implement back through the hitch via a standard connector and through the implements. Thus, the agronomic sensors installed on an implement (such as flow meters on a sprayer, plant population sensors on a seeder, or yield monitors on a combine) connect to the tractor/implement bus.

Operating the vehicle and implement combination (such as a tractor and sprayer) as an integrated unit requires a
connection between the tractor bus and tractor/implement bus, which is provided by a translator/gatekeeper called the “tractor ECU.” The tractor ECU might share information from the tractor with the implement such as PTO speed, ground speed, or engine RPM if the information is relevant to the implement’s operation. However, ISO 11783 also provides for a proprietary security mechanism within the tractor ECU to prevent the implement from giving the tractor commands unless it is authorized to do so.

1.2.2 Data collected by someone working on behalf of the farmer
While the technology discussed above gives farmers the ability to collect large amounts of data themselves, farmers may hire consultants or other service providers specifically to collect data about the farm to help the farmer’s analysis of the farm or to provide an analysis to the farmer. Such data can also be classified into two subcategories: data collected by service providers, and data collected by equipment manufacturers via telematics.

1.2.2.1 Data collected by service providers
Some outside parties, such as crop consultants, may be hired specifically to collect data. For example, a crop consultant may collect soil sample data and provide recommendations to the farmer. It should be noted that in this case the consultant would actually have the data before the farmer does, and might be able to make uses of the data that go beyond providing recommendations to the farmer, such as gaining insights into soil fertility levels across the land operated by all the consultant’s clients. In other cases, custom operators such as harvesters, planters, and chemical applicators may collect data (whether intentionally or unintentionally) about a farm simply as a result of providing their services. Custom operators may provide copies of that data to their clients, but might also retain the data for their own analysis.

You can see an example of how data collected by service provider could be used by the farmer in a variable rate crop fertilizer prescription. A crop consultant may examine soil and yield maps (that is, agronomic data) and create an application map called a “work order” specifying the rates of fertilizer application for the various areas of a particular field. The consultant likely creates this Work Order then sends the data file providing instructions to the variable-rate sprayer the farmer or custom applicator will use for the application. The data file contains the information needed for the machine to execute a “task.” Frequently, task data files use a Georeferenced Tagged Image File Format (GeoTIFF) to provide these instructions.

The GeoTIFF comprises three sub-files: a SHP file containing a vector description of the graphical objects in the file (the shape of the field and the polygons that define the pieces into which the field is broken, for example), a DBF file containing a small database of instructions (“while the implement is in polygon 1, apply fertilizer at a rate of 50 pounds per acre, while the implement is in polygon 2, apply fertilizer at a rate of 45 pounds per acre, etc.”) and an index file allowing fast lookup of objects in the SHP file. The consultant or the farmer may load the task data file to the sprayer’s task controller.

As the farmer moves through the field, the task controller receives information from the CAN bus telling it the location, direction, and speed of the implement and enabling the task controller to implement the instructions contained in the task data file. Note that this information likely involves pieces of information from the tractor’s telematics system, as discussed above. With all of this information, the task controller positions each controllable element of the implement on the SHP map and reads the map data, sending a message to the implement to operate the element in a defined manner. The implement sensor feedback provides the actual rate it applied, and the task controller creates an “as-applied map” from that data. This record can be exported back to the farmer and/or consultant desktop as a task data file.

1.2.2.2 Data collected by equipment manufacturers
In many cases, the collection of data by equipment manufactures is automated and occurs in the background of many agricultural operations; indeed, many times farmers do not even realize it is happening. Since 2011, many tractors and combine harvesters have wirelessly transmitted farm equipment data to manufacturers of the equipment, including data points such as the equipment’s engine operation, fuel consumption, and location. Manufacturers can use this information to calculate equipment field efficiency and performance rates, such as how many acres can be planted per hour. While manufacturers might analyze this data for their own programs, they may also collect it for analysis as part of a customer service plan for the equipment owner, such as a preventative maintenance plan or a farm data analysis package.
1.2.3 Data collected by a third party with no direct relationship with the farmer

While farmers have a growing number of tools available to collect data, there are also a growing number of ways to collect information about a farm without having any relationship to the farmer. Publicly available sources, such as satellite and aerial photograph imagery, are just one example. Using a number of free tools, such as Google Earth, combined with readily available public-domain data, someone can derive or infer information about a farm's planted or harvested acreages, crop mix, livestock stocking rates, and many other items of interest. Historically, this kind of publicly available data been provided specifically to give farmers access to a number of tools (such as aerial imagery) that were far too expensive for them to acquire on their own.

Continuing with the example of aerial imagery, advancements in unmanned aerial systems (UAS, often called “drones”) have made aerial imagery much less expensive. While farmers can certainly purchase a UAS to capture images of their own farm, other parties could use a UAS to collect data about the farm without even flying over it. In such a case, the data was collected without the request or consent of the farmer, but potentially violated no law.

1.3 How can agricultural data be analyzed?

Collecting agricultural data is all well and good, but something should probably come of it, and that means the data must be analyzed and decisions made using the analysis. Here, again, we see some important differences between the use of “Small Data” to make decisions on the farm versus the increasing use of “Big Data” across hundreds if not thousands of farms.

1.3.1 Analyzing agricultural data “inside the farm gate”

When we talk about the use of data “inside the farm gate,” we refer to the use of data the farmer to make decisions that impact that farm specifically. This can also be referred to as “small data - where data is analyzed and applied only to the farm where the data originated.

For example, farmers could conduct their own on-farm experiments to see how different plant varieties responded to conditions on the farm. Yield monitor data from a harvester would provide the primary source of data to determine how varieties responded, holding everything else constant. However, what if the farmer had also used grid sampling of soils to get more specific information about how to apply fertilizer to the fields? This could provide information that adds more depth to the understanding of the yield data, and also helps with precision fertilizer applications. Now, what if one adds crop scouting data, whether that data was gathered from a consultant walking fields or a UAS (“drone”) to look for evidence of pests affecting crop health? This now also adds a layer of information and analysis for pesticide applications. And taken all together, this data could also help a farmer determine how he or she could comply with any environmental requirements, such as allowable runoff for agricultural chemicals in a watershed with a Total Maximum Daily Load (TMDL) that restricts total runoff loading to a creek, river, or lake.

In research circles, these kinds of uses are sometimes referred to as “G x E” analyses, where “G” stands for “genetics” (the particular crop type in question) and “E” stands for “environment” (such as soil type, soil moisture, soil nutrients, and so on). New data collection and analysis tools make it easier than ever for farmers to conduct more of these “experiments” and to have more data available to interpret the results. The increasing amount of financial information available to farmers also gives them increased abilities to estimate the economic impact of their farm management decisions.

All of these data uses are considered “primary uses,” meaning the farmer is using data about the farm to make decisions that directly impact that specific farm. Arguably, this is where the farmer extracts the highest value from his or her data through increased efficiency and/or productivity of the farm's resources.

1.3.2 Analyzing agricultural data “outside the farm gate”

There may be significant value to be had from the analysis of data outside the farm gate, however. To illustrate this, let’s step away from agriculture for a moment to look at the example of Amazon, the online retailer. You search on Amazon for a garden hose to water your vegetable patch. When you do, you see suggestions labeled as “customers also searched for” or “frequently bought together.” Those suggestions come from an analysis of data collected from thousands of other customers who searched for or bought an item similar to the one for which you were searching. In other words, you are seeing the results of analysis not only of your data, but your data compared with the data of many, many other Amazon customers. As a result of that analysis, you might discover an item that significantly improves your efficiency in using the item you originally sought, such as a spray wand to help you reach plants that were difficult
to access before, an irrigation timer that automates watering for you, or irrigation tape that increases your watering efficiency. Of course, you might also find and buy something you really didn’t need at all, like a garden gnome.

To bring things back to agriculture, the analysis of data for your benefit can provide a number of economic benefits to the operation as discussed in section 1.3.1 above. However, an analysis of your data coupled with the data of hundreds or thousands of other farms can provide benefits for you and the other operations as well. This is data analysis “outside the farm gate” and takes us into the realm of “Big Data.” Using the example above, if a group could analyze the planter and yield monitor data of hundreds of farms coupled with information about the management processes or conservation practices of those farms, they could conduct “G x E x M” (genetics x environment x management) analysis to see how many factors work together to influence crop performance. That information could help consultants provide even better recommendations to their clients about how to optimize their farms’ operations. Soil data collected across a region (as opposed to a single farm) could provide important information about the potential for nutrient runoff to pollute nearby waterbodies. Scouting information across a region (again as opposed to a single farm) could provide advanced warnings of pest or disease outbreaks that could prevent many farms from experiencing any productivity loss at all. And all of those factors together could be used to improve the environmental health of the region as a whole.

As you can see, analyzing the data of many farms can create products that provide value to the individual farms sharing data, but also provide additional value to the “community” of farms sharing the data as well. Of course, those aggregating and analyzing the data will likely want a share of that value, and may create completely separate value from that data as well; that will be discussed later.

1.4 How do agricultural technology providers aggregate data?

So how do companies “aggregate” farm data to create the value from Big Data analysis as discussed above. Some parties naturally accumulate data across multiple farms, as discussed above in the examples of crop consultants and custom operators. The agreements between these service providers and their customers may provide that the service providers can also use the data acquired from the customer in combination with data from other customers to develop new knowledge, improve their ability to serve customers, or in some cases, to sell the combined data to other parties for their analysis or to market products or services. With these agricultural service providers (sometimes also called “agricultural technology providers” or “ATPs”), the aggregation of data simply comes from the combined storage of their respective customers’ information.

You may use a “cloud storage” system such as iCloud, Dropbox, Google Drive, or OneDrive to store files on an Internet-based server. There are also companies that are specifically tailored to provide storage services for agricultural data. Some companies provide just that - data storage services - and their user agreements specify that customer data will not be combined with that of other users or analyzed in any way; the company is simply serving as a “data warehouse.” Other companies who provide data storage specifically offer to give the user access to information the company derives from analyzing the data of all the customers - in other words, one of the services provided in addition to data storage is community data analysis. These companies may provide farmers with Small Data reports about the customer’s own farm, but might also provide Big Data reports that provide the customer with insights applicable to their farm derived from the analysis of many other similar (or even different) farms. Those Big Data analyses come from the large-scale aggregation of data from many, many farms into a “pooled” dataset. This might involve the combination of data directly from the farms with “data about the data” (also called “meta-data,” which, using planting data as an example might include not just the plant variety but the location, time and speed of the planting). That data can also be re-combined with external data sources such as weather, climate, soils, and other crop environment factors to create “recombinant data” adding more detail to provide additional insights.

1.5 Why would companies be interested in the data from my farm?

Just as farmers collect and analyze their information because they see the potential for increased farm profitability, service providers of any kind provide their services because they see a potential for profit as well. That includes ATPs and data aggregators.

At the most basic level, a party who has access to agricultural data and who is not otherwise prohibited from doing so may simply sell the data to a third party. However, the data itself may not be as valuable as the insights to be gained from the analysis of the data. That said, some companies see their role as simply aggregating and selling data.

Other companies see the data from farms as an opportunity to conduct investigations that help them better
understand agricultural production and derive better insights about how to produce more food and fiber with fewer inputs. For example, a seed company using aggregated data could see how their seeds perform across thousands of farms, which also means thousands of soil types, fertilizer types, pesticide applications, rainfall and irrigation amounts, and so on. That means field trials that would have taken a decade may only take one or two years, meaning new varieties make it to market faster. Crop consultants can get more and better insights about what recommendations do and do not work for any number of crop conditions. Thus, the value that these companies derive from the data they access can also be shared with producers.

To be sure, there are “win-win” uses of data; there are also “win-lose” uses of data where one party profits at the expense of another, and then there are a spectrum of uses in between the two. With enough data, for example, an input company such as a fertilizer or seed company might know more about a farmer’s crop environment than the farmer does, and could use that abundance of information (sometimes called “asymmetric information” - a situation in which one party knows more about the other’s bargaining position than they themselves do) to target highly specific marketing efforts at the farmer. In some cases, this might benefit the farmer in helping them find products that are a good fit for their operation and about which they might not have known, but in other cases it might lead to pricing practices that put the farmer at a distinct disadvantage. Knowledge of how some products perform in a specific geographic region can help input producers streamline supply chain management of production and retail resources, meaning lower costs of production and thus lower prices for producers - but it could also simply mean more retained earnings for the input producer. With enough information, a data aggregator could derive insights about the commodity market before government reporting agencies or other private market participants do, giving them an advantage in commodities trading - when used for cost management, this can again mean lower input prices for everyone involved, but it could also mean increased profit capture to the exclusion of other parties, or at worst, market manipulation.

1.6 Chapter conclusions

Farmers have unprecedented opportunities to collect data about their own farm, which in turn gives them the chance to gain more insight into how their farm works. That creates the opportunity to capture increasing amounts of economic value for their farm. At the same time, other parties also recognize the value of agricultural data beyond the insights it can provide to the farm itself. As a result, more and more agricultural technology providers and other parties have stepped into the agricultural data arena, seeking to capture a share of that value.

As with any tool or resource, these tools and resources cannot be inherently “good” or “bad” – the moral value of it comes from how it is used. Similarly, there are almost infinite options as to how farmers and data aggregators can share the value to be derived from farm data, and defining whether a particular service, technology, or agreement is “good” or “bad” from the farmer’s perspective depends on the amount of value the farmer feels he or she should be able to capture. In the chapters that follow, we will discuss the value farmers can get from their data, and how they can protect that value.

1.7 Chapter Endnotes

5. “As-applied” data refers to the type and amount of agricultural inputs applied to a crop such as seed population, fertilizer, and pesticide applications
6. “Grid soil sampling” refers to taking soil samples from a field at regularly-spaced intervals to form a “grid” across the field. The samples are then assigned a geographic marker to note where they were taken in the field. When the samples are analyzed, they can create a more precise map of soil types and nutrients than traditional


13. SAE White Paper, supra note 9, at 5; Interview with Dr. Marvin Stone (June 10, 2015) (on file with author).


20. See Stone above.

21. Id.

22. Id.

23. Id.

24. Interview with Kevin Thedford, Precision Agriculture Specialist, P&K Equipment (June 30, 2015) (on file with author).


26. See ISO 11783, at §§ 6.3 – 6.4 (discussing connection of network architectures.)

27. See SAE WHITE PAPER; see also Stone.


29. Id.

30. See id.


32. See AGRIC. INDUS. ELEC. FOUND., AEF DRAFT INTERNATIONAL GUIDELINE: ISOBUS IMPLEMENTATION SPECIFICATION FUNCTIONALITY TIM 16 (June 18, 2015).


35. Id.


39. Id.

40. See generally Beck.

41. Id.
Chapter 2
How can I use agricultural data on my farm?
The case of “small” data

Almost every industry is abuzz with talk of Big Data and the potential it holds to unlock insights into how to make almost any venture more efficient and profitable. However, there is no Big Data without “small data” - data about the individual firm (or in the case of this handbook the individual farm). While Big Data models can provide significant economic value to a lot of parties, including farmers, there is a lot of value to be gained from fully understanding the small data that can be collected about your farm - indeed, as discussed later, your farm may already be collecting small data without you even realizing it. In addition, in order to participate beneficially in Big Data, farmers must do small data correctly.

Though they may have been called by other names, the concepts of small data have been a focus of many Extension programs for many decades. In the current Big Data world, many farmers are asking that Extension professionals return to small data so that they are comfortable with those skills before joining a community of Big Data.

2.1 Uses of agricultural data

One way to examine what could be done with your farm’s data is to examine what has been done with agricultural data on other farms.

One of the larger surveys of agricultural data uses is the USDA Agricultural Resource Management Survey (ARMS). Though not a direct survey about agricultural data technology use, ARMS focuses on agricultural production practices and captures some useful information about agricultural data practices. Each year a different crop is evaluated by the ARMS survey. Soybeans were the focus of the 2002 ARMS survey. In 2003, cotton, sorghum, and barley were examined. Spring wheat, winter wheat, and durum wheat were examined in 2004. In 2005, corn and oats were examined.

The ARMS data reveal much about the use of one of the predominant agricultural data technologies in crop production is the yield monitor - a sensor system mounted on harvesting equipment to estimate the amount of crop harvested over a given area. Data from the ARMS survey years referenced above indicate that the leading use of yield monitors by farmers has been to monitor crop moisture. Interestingly, though, the moisture sensor on yield monitors was initially intended more to accompany the mass flow sensor to correct for moisture when calculating yields; it wasn't necessarily provide moisture content as a primary data output. Nevertheless, the moisture reading on its own has been the most commonly used data from the technology. What you might expect to be the primary use of a yield monitor - documenting yields - was actually the second most common use of yield monitor technology (Griffin and Erickson, 2009).

Indeed, while yield monitors were initially commercialized for documenting yields although innovation by farmers has extended the frontier for how the data can be used to its potential. In addition to documenting yields, the most common uses of grain yield monitor data associated with a GPS are to monitor crop moisture (as discussed above), conduct on-farm experiments and tile drainage management (Griffin, 2010). Contrary to early anticipation from farm managers, yield monitor data have not been commonly used for apportioning crop share rents between farmers and landowners. Griffin (2010) reports that few respondents, based on United States Department of Agriculture data, suggest that yield monitor data are used for negotiating leases, dividing crop shares and managing irrigation.
2.1.1 Harvest management

It seems logical that farmers use the moisture sensor to determine if the crop is ready to be harvested and/or in deciding on which drying or storage facility to send the particular crop, and indeed, they do. The moisture content of a harvester-load can inform the farmer as to whether the grain can be loaded on a truck and taken directly to a point of sale such as an elevator without the need for further conditioning or whether the grain needs to be sent to on-farm storage for drying. The moisture-sensing capability of yield monitors can also help in in-harvest decisions for crops other than grains and beans as well. One example of harvest logistics is filling cotton module builders by tracking overall volume of cotton in the harvester. Seed cotton is temporarily stored in the field in a semi-compressed “module” that is constrained to hold a narrow range of cotton. Cotton yield monitor data gives the cotton picker operator information with which to decide to continue with the current module or to complete the module and begin another module.

2.1.2 Conducting on-farm experiments

Yield monitors and other site-specific sensors have allowed farmers to collect low-cost yield observations. With that information, farmers can see the effect of changes in their crop enterprises in a series of “on-farm experiments” or farm-scale trials. Farmers have used this information to compare crop varieties, tillage treatments, and other inputs or systems. Knowing how Variety X of wheat performs on a company or university research plot is nice, but what really matters to a farmer is how Variety X performs on his or her Field 1. For example, USDA data shows that, for cotton pickers equipped with GPS, conducting field experiments was the greatest use of the technology (and it should be noted that most cotton monitors have no moisture sensor, the “otherwise number 1 use of yield monitors).

Agricultural technology has reduced the cost of collecting farm data. Access to GPS enabled sensors has presented opportunities for farmers to conduct their on-farm experimental trial. For logistical reasons, relatively few farmers utilized strip trials or other on-farm designs derived from classical small plot experimental methodology. With the recent advances in spatial statistics and cloud computing, an alternative to traditional experimental commonly associated with classical agronomic trials became available. Rather than deliberate intervention style of experimentation, observational experiments with fewer replications using spatial analysis via cloud computing encouraged farmers to revisit on-farm research. Now that farmers are conducting more on-farm trails across wider regions, better farm management decisions with respect to those tested inputs and factors of production can be made across their farms.

2.1.3 Directed scouting

One common use of georeferenced (georeferencing is discussed below) yield monitor data - especially immediately after first using the technology - is to guide farmers and their advisors to specific locations within fields based on relative differences in yield, i.e. “directed scouting.” After examining yield maps, the first task farmers often perform is visiting specific locations to evaluate site-specific soil nutrient levels that correspond to relatively small and large yields. In practice, many areas with low levels of fertility are associated with larger yields than areas of high fertility levels where the yield limiting factor is something other than nutrient availability; thus there actually may be an inverse relationship between crop yield and fertility level. In other words, when fertilizer is applied at the same rate across the entire field, fertilizer applied to areas where other yield limiting factors are severe (for example, soil pH or drainage) accumulate nutrients because of limited nutrient uptake; thus these areas have large concentrations of crop nutrients and low yields. Directed scouting using data from your fields can help you and your advisers identify “trouble areas” and find the cause of the trouble.

2.1.4 Input decisions and variable-rate applications

Yield monitor results and directed scouting can reveal a number of conditions that justify treating some areas of a field much differently than others. At the whole-field level, yield data can help farmers determine if input applications have provided the intended results, or at least the expected results given the conditions present after the inputs were applied. In some areas, soil nutrient applications are based on crop removal using yield monitor data.

A field with fairly uniform conditions may perform well with uniform application of inputs, but fields with a high level of variability in soil nutrients, drainage, and other conditions might offer drastically different financial returns from treating each “subfield” differently. In such cases, the additional expense of using “variable rate” technologies might be justified by the reduced input costs and/or increased revenues that could be realized from the precise
application of inputs and practices tailored to small areas of the field. Today, variable rate technologies allow farmers to change the rate of all kinds of inputs, including seeding rate (and in some cases, even seed variety), fertilizer and pesticide applications, and irrigation water “on the go.” The added flexibility of variable-rate technologies does add cost, though, and thus an initial assessment of historical yields and soil conditions are likely a necessary foundation to determine whether the extra expense can give rise to profitable implementation of those technologies.

2.1.5 Justifying tile drainage
In areas of the U.S. that rely upon subterranean tile to drain soils, anecdotal evidence has suggested that yield monitors equipped with GPS have helped substantially to quantify the yield reduction due to poor drainage and the potential benefit from drainage improvements. The quantification of yield and profit losses due to poor drainage can be a factor in making land improvements where the farmer owns or leases the land. The ARMS data supports the notion that farmers are using yield monitors with GPS to make tile drainage decisions especially for soybeans, winter wheat, and corn with over 30% of farms with a GPS yield monitor. Coupled with other data layers such as local rainfall, evaporation potential, cost of tile installation, and revenues from crop sales, this data could help farmers determine whether the installation of tile drainage would justify its cost.

2.1.6 Managing leases
With the again-notable exception of cotton, farmers have not widely used yield monitors in lease negotiations or splitting crop shares. Early in the use of yield monitors, it was expected that leasing arrangements would benefit from the technology; however, from the USDA ARMS data and anecdotal evidence, farmland lease arrangements have not been greatly influenced by precision technology especially for negotiating the lease. Farmers producing cotton, durum wheat, and sorghum have made at least some use of the technology for splitting crop shares. Chapters 6 and 7 discuss the potential uses of agricultural data in your farmland leasing arrangements.

2.1.7 Benchmarking
In the United States, the most complete farm-level data on production and financial performance are from Illinois through the Farm Business Farm Management Association (FBFM), and in Kansas through the Kansas Farm Management Association (KFMA). These two states are reasonably representative of the broader agricultural regions of the eastern corn belt and the great plains, which is where the majority of corn and soybean are produced. However, several other states similar albeit smaller programs. Many of these are members of the National Association of Farm Business Analysis Specialists http://www.nafbas.org/. These services often compare individual farm's production and financial ratios to “good, ok, and bad” levels (as those measures are determined by professionals in agricultural production and economics). In addition to farm level recordkeeping, tax reporting preparation, and changes over time, these associations provide comparative analytics across cohorts, i.e. benchmarking, to determine if the farm is performing similarly to their peers at any given time. If the farm is performing below these peer levels, it can be a signal that investigation is needed to see why that might be the case.

Participating farmers in these three farm records examples benefit in addition to compliance with the IRS; each participating farmer receives detailed summaries of the whole system months or even a year in advance of any public release of the summary. The farmer also receives detailed benchmarking reports of their physical production and financial performance compared to the peer group of like-sized farms. These benefits go further than just their yearly “report card;” they get to see why their farming operation differs from the average of their peer group and get to improve their operation.

2.1.8 Regulatory compliance
For nearly a century, farmers have used farm financial data to document the fact they are accurately paying their taxes. You might not have thought about it, but when you complete your tax return, you are indeed using financial data to document your expenses and revenues to validate that you have correctly calculated the amount of income tax you owe, or the refund to which you are entitled. Further, though, once farmers have compiled those records, they can review the records to determine exactly what the profitability of the farm was for the year. It should be noted that there can be (and very often is) a very big difference between whether a farm made money according to its tax return and the accounting measures used by financial professionals, but the “re-use” or “secondary use” of the financial data created for
regulatory compliance (i.e. “I correctly paid my taxes”) is financial analysis for the farm.

A secondary use of agricultural data (whose primary use may have been to help the farmer in making his or her management decisions) can be regulatory compliance, too. If your farm applies animal waste to fields under a nutrient management plan, it may be possible to use “as applied” maps to demonstrate your compliance with the plan. The same may be true of operations who operate in watersheds that are subject to a total maximum daily load (TMDL) regulation.

As mentioned numerous times throughout this handbook, the accuracy of data is crucial for its use to make sound farm management decisions. It may be even more critical for the use of data as a tool to demonstrate legal compliance. If a sensor problem shows you applied animal waste at twice the rate legally allowed, but the data was not validated to reveal the potential problem, a farmer could be in a lot of trouble for apparently committing a violation when, in fact, they did nothing wrong. As a result, before relying solely on agricultural data for your compliance reporting, be sure to constantly and consistently maintain your data systems, validate your data, and consult with your attorney before submitting compliance reporting materials.

2.2 The importance of georeferenced data on your farm

The “precision” in “precision agriculture” comes from using a GPS to geo-reference data to a specific field location. Doing so means, in the example of a yield monitor, that the monitor can identify the specific location in Field 1 where it was recording a yield of X bushels per acre. However, in the US, a substantial portion of yield monitors are not associated with a GPS (Griffin and Yeager, 2018; Schimmelpfennig, 2016). Although many farmers have adopted and continue to operate yield monitor equipped combine harvesters without georeferencing capability in an sequential adoption process, the usefulness of the yield monitor is dramatically diminished without the georeferenced data; however practical uses of yield monitors without GPS may exist but are drastically limited.

With the commercialization of instantaneous yield monitors and global positioning system (GPS) equipment, many georeferenced yield observations can be recorded inexpensively. Site-specific yield data provide farmers and their advisors with additional metrics to examine the factors that affect yield such as drainage, soil management and planned on-farm experiments. These methods of analysis have included both subjective visual observations (the “eyeball test”) as well as objective quantitative analysis with spatial statistical methods (sophisticated, software-assisted analysis).

Yield monitors associated with GPS provide intensive data over large areas. The data from GPS provide location in terms of coordinates (latitude and longitude), time, elevation and their derivatives such as speed and heading. All of this information can be important in helping define the specific site of issues to be addressed in fields, whether those issues are soil nutrients, drainage, pests, or other yield-influencing factors.

2.3 How long is agricultural data valuable?

Many farm records are usually destroyed after so many years (for example, tax records). The logical question is whether agricultural data has an “expiration date.” It is intuitive that more recent data is more valuable than older data, but at what age does precision agriculture data lose its value? The answer is, yet again “it depends.” With seed varieties or hybrids, the average market life is less than three seasons, so the relative value of varietal information that is four years old may be worth very little relative to two year old data. A compounding factor in agriculture is the yield trends from agronomic advances, including genetics. For instance, USDA data indicates soybean yields trended 0.42 bushels per year higher over the last 30 years; therefore soybean data must be de-trended in order to preserve its value for certain types of analyses where trends were not of interest. An example is when weather events are correlated to soybean yields.

The fact is that data value will depreciate over time, but time is not the only factor that diminishes the value of the data. A four year old hybrid may be irrelevant to the analysis and have no value but a ten year old nitrogen rate study may still have value. The figure below presents a hypothetical example of the potential value of fertility data and varietal data due to depreciation and usefulness. For varietal data, the value depreciates relatively quickly and drops off as the market life of the variety is exceeded; as opposed to fertility data that has some depreciation but even older data still have pertinent use in current analyses.

Hypothetical example of relative value of different types of data over time
Chapter 2 References


Risk comes in many forms for a farmer. Consider for a moment a car factory. If you put in the right amounts of parts, operate all of the machines correctly, and use the right procedures, it’s almost an absolute certainty that a car comes out the other end of the factory. Sure, there might be issues with the price for the components and labor used in your factory, and the interest rates you pay on the capital you borrowed to buy the factory or finance operations might change, to say nothing about the demand for the finished car, but hey, at least you have the cars, right?

Now consider the risks faced by a farmer. In addition to the variability in prices for your inputs, your product, and your capital, there is one other critical variable. You might do everything exactly right from an agricultural practices perspective, but factors completely outside your control like weather and disease might mean you have no crop to show for it. Thus, risk is a constant concern for farmers.

Fortunately, the data revolution continues to create new tools for farmers to gain more information not only about their own operations, but about more and more factors that influence the productivity and profitability of their farm as well. Actionable information on those risks means farmers have more and more options (and hopefully, ability) to manage the risks that influence their operations.

### 3.1 Production risk caused by weather issues

In grain producing regions, weather has been a source of uncontrollable production risk. However, weather risk can be managed to some extent by understanding the probabilities of weather events and using real-time application of local data to farm management decision making. Using an example specifically from 2018, extreme weather was observed across the midwestern USA. Lower than normal temperatures resulted in soil temperatures reaching critically cold levels later than typical for the region. Although some farmers may directly measure soil temperature at specific fields, additional peace of mind may occur when temperature measurements are confirmed with “mesonet” (mid-geographic scale) data. Knowledge of state-level and sub-state soil temperatures provides peace of mind knowing that other farmers are experiencing similar adversity, but can also help farmers make precise planting decisions. This knowledge of the farmers’ community of peers across larger regions relaxes the feeling of “being behind” on planting. This was pertinent in 2018 because as of April 15, 6% of corn acres were planted in Kansas relative to being 15% completed on average (USDA NASS, 2018).

Weather data has been increasingly utilized as a precision agriculture technology and becoming one of the more widely used forms of big data (Coble et al., 2018). Weather data in general and farm data from a network of weather stations in particular have been used in conjunction with precision agricultural technology for such farm activities as on-farm experimentation (McBratney et al., 2005). Weather data are available from NOAA and regional weather station networks known as mesonets. At east 28 non-federal, state-wide coverage climate focused mesonets exist (Mahmood et al., 2017). Although not all mesonets are intended for agricultural decision making, the majority provide data useful by agricultural applications. Early mesonet systems were influenced by agricultural climate applications (Hubbard et al., 1983).

Weather and climate data are important to agricultural decision making. With the advent of digital agriculture, traditional weather data has found new agricultural roles in the guise of big data. Weather data is a form of farm data that falls into the “information-intensive” category of precision agriculture technologies (Knapp et al, 2018). These services are typically offered to farmers on a subscription basis, as were the subscriptions at issue in this litigation.

Traditionally, however, weather data has been considered “free” to farmers. In the past, for example, many people...
obtained their weather information from the evening news and local newspaper. These traditional sources provided
data for the previous day as well as short-term forecasts.

Today, farmers have access to constantly updated and publicly available weather data from sources such as the
National Oceanic and Atmospheric Administration. In addition, high-quality local weather data is available from
several free sources. For example, the Midwest Regional Climate Center (MRCC) includes Iowa, Illinois, and several
other corn producing states, and provides data on temperature, precipitation, snow, growing degree days, pressure,
wind, humidity, present weather, visibility, wind chill, heat index and others. (https://mrcc.illinois.edu/data_serv/dataNetworks.jsp). These networks of weather stations are referred to as a Mesonet, “a combination of the words
“mesoscale” and “network”. In meteorology, “mesoscale” refers to weather events that range in size from about one mile
to about 150 miles. Mesoscale events last from several minutes to several hours. Therefore, mesoscale weather events
are phenomena that might go undetected without densely spaced weather observations. Thunderstorms, wind gusts,
heat bursts, and drylines, are examples of mesoscale events (http://mesonet.org/index.php/site/about). In Kansas, a
mesonet (a network of typically automated weather and environmental monitoring stations) comprised of nearly 60
professional-level stations offers real-time and historic weather data (Knapp et al., 2018; http://mesonet.k-state.edu/).
The Oklahoma Mesonet includes 120 weather stations with at least one in each country (http://mesonet.org/).

One can think of numerous uses of weather data in agricultural decision-making. For example, consider growing
degree days for corn, a metric comprised of temperature information used to determine crop physiological growth
stage for irrigation and harvest decisions. The Useful2Usable website is a free publicly available tool hosted by
University of Illinois (with collaboration from several Land Grant Universities and respective scientists) provide
growing degree day information for nearly all of the Midwest corn production area https://mrcc.illinois.edu/U2U/gdd/. And beyond readily available and free sources, weather data is also often offered with many other services or on
both paid and free websites commonly visited by farmers such as agricultural retailer or cooperative websites widgets
or links to other websites.

One may argue that data from public mesonets may not be a close or even near-perfect substitute of consumer
grade weather stations given that substantial amount of evaluation, oversight, and maintenance are given to each
weather station in the mesonet (Knapp et al, 2018). It is suspected that consumer grade weather stations have
little routine planned maintenance or evaluation for accuracy. In addition, mesonets in Kansas and Oklahoma have
instruments at 10 meters above the ground in addition to ground-based sensors to measure temperature inversion.
Therefore, using the logic of near-perfect substitutes may not be valid given the likely lower-quality data coming
customer-grade weather stations. Given that high-quality local weather information is publicly available from
university and state agencies for mobile and web-based platforms, it stands to reason that profit maximizing farmers
would not pay a fee for a near-perfect substitute. All of this is to say that with the continued improvement of mesonet
scaled weather information, it may not be cost-effective to purchase private weather stations unless those stations can
show a significant increase in the accuracy of information over the mesonet sensors.

Beyond the usefulness of immediate weather information, long-term trends in weather data (but perhaps still
short of reaching a “climate” scale) can also inform farm management decisions. One example of using long-term
weather data applied to farm management involves data associated with days suitable for fieldwork (DSFW). Days
suitable for fieldwork data have been collected weekly for several decades across most states by the U.S. Department
of Agriculture (USDA) National Agricultural Statistics Service. These data have been published in the weekly Crop
Weekly reports typically provides a single data point for the state and/or crop reporting district.

Analyses of long-term DSFW data have been reported for several states mostly by farm management economists
and applied agricultural engineers. These analyses reported how farmers and their advisors utilize fieldwork
probabilities to determine optimal machinery sizing and crop allocation. The “most active” time periods to conduct
fieldwork such as planting and harvest have been reported by USDA. Within these time periods, the number of days
suitable to enter the field to conduct fieldwork have been evaluated to determine long term trends and probabilities.
These probabilities are used in a risk management strategy when equipping an existing farm or deciding if additional
farm acreage necessitates additional machinery. These analyses are available in interactive graphs from both Kansas
(https://www.agmanager.info/machinery/kansas-days-suitable-fieldwork) and Oklahoma (http://agecon.okstate.edu/extension/fieldworkdays.asp). Days suitable data on their own have limited value but when recombined with other
data layers, these data empower farmers to make important farm management decisions about field operations.
3.2 Using data to manage price risk

When field level data are aggregated into a community, several possible risk mitigation opportunities become possible. One example could be the stochastic dominance seed selection tool that assesses relative rankings of seed varieties from publicly available data and privately held yield monitor data. The publicly available data were from Instituto Nacional de Tecnología Agropecuaria (INTA) but identically the same model could be used with university official variety trials (OVT) trials in the USA. At the time their study was published, very few if any community based tools for production decision making were available in practice.

They incorporated a variety of data (scrapped public data from the INTA website combined with a farmers own yield monitor data) that could be easily expanded in three important ways. First, rather than using a single observation from on-farm trial results the mean yields by soil or other zones could be utilized to increase the number of observations and broaden the range of applicable environments. Second, rather than using only the mean yields the full empirical distribution could be utilized by using all yield monitor data points. Finally, their model can be further expanded by taking into account neighboring farms’ data using a community approach. Having not only publicly available OVT data and a farmers own field data, but also pertinent data from across the community gives fuller perspective on the farm management question at hand.

3.3 Using “recombinant” data

Many individual data layers have been discussed here. Each of these have advantages on their own; however one of the tenets of Big Data is “recombining” publicly and privately held data into a new form. Recombinant data is nothing new to agricultural scientists. Assembling different data sources into framework suitable for analysis occurs frequently. A simple example is combining county level populations from US Census Bureau with county level agricultural acreage from USDA. An example in precision agriculture may be combining several site-years of field-scale yield and treatment data with weather data for analyses. Re-combining aggregated on-farm experiment data with other data that would not have substantial interpretation for a single farm such as weather, cropping history, input usage, historical yields increases the potential value of the analyses. In this example, weather could be assumed to be constant across the entire site-year or field; but when multiple fields across locations and/or years are assimilated weather becomes a variable because it cannot longer be assumed constant but rather expected to vary.

Another tenet of Big Data is using primary data again, i.e. a re-use or secondary use of data initially collected for another purpose or from axillary. In some geographies in the US, elevation and in particular changes in elevation have a substantial impact on crop productivity. On-farm experiment analysis is typically greatly improved by incorporating one or more variables that proxy for terrain attributes such as slope and aspect. The availability of high accuracy elevation has not been readily available although data can be acquired from publicly available LIDAR (Light Detection and Ranging, a light-based equivalent of radar) and within-field RTK (real-time kinematics) GPS. In most cases, combine harvesters and lightbar guidance utilize Wide Area Augmentation System (WAAS) differential correction which is sufficient for location but usually not adequate to create a digital elevation model (DEM). Today, many farms utilize RTK GPS which is sufficient to measure not just relative but also absolute elevation; and the more trips across a field that RTK GPS equipped machinery makes the higher spatial resolution acquired. In the event that a farm does not employ any RTK GPS on their own equipment, custom applicators may opt to use the highest accuracy GPS to collect elevation data for later re-use for that farm. Essentially, custom applicators who opt to use higher accuracy GPS than needed for the original task (parallel swathing) may be able to provide the particular farmer with value added services or simply sell elevation data to them. In addition to elevation, any additional data layers that can be collected on each pass through the field is generally lower cost relative to dedicated trips through that field.

Intelliair (http://intelliair.com/) is a grain storage facility company with the motto “intelligent solution to grain management”. In addition to their innovative grain storage technology, they implemented environmental sensors to complement their grain storage products. Knowledge of ambient temperature and humidity is important to decision-making process of operating grain bins. These environmental sensors evolved into their FieldData Manager for ‘comprehensive field and weather data’ including wind, rain, solar radiation, soil probe integration, and data tracking interface. Although weather has not been the core business of Intelliair, their innovation with weather sensors to initially support their grain storage products has found greater value in the re-use of that data especially when combined across larger geographies. Today, Intelliair is providing rainfall, growing degree day units (GDD), and real-time crop maturity prediction tool. Predictive analytics can easily be added to provide customers with information regarding expected weather based on sensors (presumably) west of the customer in question.
A note on recombinant data: recombinant data aggregation is not simply pooling data by stacking or adding rows or records into a data frame. Aggregators and repositories must provide indexing such that data are searchable; and not just for basic meta-data but for unique characteristics of individual observations from the field. For example, a simple study could be conducted on yield impacts of a pathogen (assuming said pathogen had been mapped, tagged, or otherwise searchable at the yield data level) and the efficacy of the fungicides applied to those fields. Such indexing can show the prevalence or intensity of disease over time by weather regime. In addition to intensities or presence of anomalies, site-specific meta-data should include where question were raised even if left unanswered. The lack of addressing a standing question is in itself data that may be useful to solve the problem once in the right format of volume and variety.

Adding searchable tags to digitized data such as aerial imagery, NDVI, yield data, soil sampling, soil maps, etc. improvise potential of Big Data in agriculture. In the early days of datafication of precision agriculture, I scanned the existing USDA Soil Conservation Service (SCS) soil maps (even at that time the agency was called NRCS but the soil maps from the 1960’s were still relevant), stretched the image onto existing GIS layers such as county roads (i.e. “rubber sheeting”), then traced soil series lines, homesteads, pastures, fence lines, ponds, and other manmade structures. Although the SCS soil map was scanned, it was not datafied but only digitized. Once soil lines were traced into a GIS polygon filed, it was datafied.

The essence of this task was simple; convert the data into machine readable form to remove the need for human interaction. The information would have multiple potential purposes and opportunity for further analysis by insightful scientists. I successfully datafied the SCS soils maps, but for only the fields I had immediate interests, something like 10 fields or 500 acres. Thankfully, USDA NRCS provides this data across the USA for free today.

**Chapter 3 References**

Chapter 4
How can I evaluate the economics of adopting agricultural data tools?

“Is precision agriculture profitable?” “Are Big Data tools worth it?” These are very common questions. And, as with most important questions, the answers are “it depends” and in this case it depends on “when, where, and who.”

For example, the profitability of automated guidance and automated section control depends on field geometry. For rectangular fields with relatively long passes, automated guidance is typically advantageous; for smaller irregularly shaped fields, automated section control has a comparative advantage. In these cases, “where” matters a lot. For the data technologies such as yield monitors, grid soil sampling, and variable rate applications, the profitability depends upon the field characteristics and also the decision maker, i.e. the “who.” Thus, data coming from precision agricultural technology and the associated data tools may be valuable, but only for the farm decision maker that is able to make better decisions with that information.

To make the best technology adoption decisions, farmers must take a close look at their own management abilities. First, they need to consider the amount of time and effort they are willing to put into enhancing their “human capital.” If farmers are not willing to invest in building the skill set needed to make use of technology, then they are likely better off to use only the automated technologies such as guidance (“auto-steer”) and section control. If the farm decision maker has the capacity and is willing to invest time and effort into making the data technologies work for their farm, then they may consider data tools arising from yield monitor data and associated technologies.

Beyond the investment in human capital needed to take advantage of agricultural data technologies, there are also direct benefits of agricultural technologies to your human capital assets as well. For example, air conditioned tractor cabs greatly improved the comfort of operators, letting them cover more acres in a day. Similarly, automated guidance has made many operators less fatigued meaning they can operate for longer periods during the day. Arguably, they have also let some farmers operate at greater ages than they could have without that technology.

Farmers more readily adopt technologies that are automated (like automated guidance) and tend to be slower to adopt the more analytical, data-intensive technologies like yield monitors. With automated guidance, little human capital is necessary to make the technology profitable for the farm. With yield monitor data, much more investment in learning and management are necessary to utilize that data. Research on the adoption rates for these agricultural technologies reflect these differences. For example, the “low-hanging fruit” of agricultural data utilization is using yield monitor data for on-farm experimentation (e.g. seeing how a different plant variety or fertilization treatment affects yields); this was the most frequent use of agricultural data farmers indicated on surveys. Farmers can get more local information much quicker for their farm by conducting their own on-farm research. Still, as shown below, yield monitor adoption rates are well behind those for automated guidance.

4.1. Is it economically worthwhile to adopt any of the precision technologies or Big Data tools?
As discussed above, the “when, where, and who” greatly influence the profitability of agricultural technologies. Given that the economics of a technology are a function of not only the specific grower’s fields but also the management ability of the grower, determining the profitability of specific technologies have been difficult (Griffin et al. 2018). Some farmers are not likely to be able to make profitable use of some agricultural technologies, especially those that are information-intensive such as the Big Data tools arising from agricultural data.

Precision agriculture technology has been separated into two distinct groups: “embodied knowledge” and
“information intensive.” Adoption rates of embodied knowledge technologies (such as automated guidance and automated section control) have been greater than information intensive technologies (such as yield monitors and grid soil sampling) as shown in the figure below.

Although farm data generally are collected via the information-intensive technologies, many data layers are collected from the automated, embodied-knowledge technologies. The role of the yield monitor and grid soil sampling is to generate data that farmers and their advisers can use in the decision making process. The role of automated technologies such as guidance are to conduct field operations; however, that technology also provides information in the form of data such as elevation, engine RPMs, fuel consumption, and as-applied data.

Nearly all farms researched showed a quick payback of automated technologies such as guidance and section control, especially where field shape and size were conducive to that technology. Returns on data intensive technologies have been less clear and rely upon more than field geometry. Economics of technologies such as yield monitors, soil sampling, and variable rate applications is a function of the management ability of the farmer, not simply the characteristics of the field.

Precision agricultural technologies have been criticized for benefiting larger acreage farms more than small acreage farms, mainly due to being able to spread the fixed costs of investment over more acres. Agricultural data tools have received some of the same criticisms in that, at first glance, large acreage farms benefit the most from farm data. That said, opposing arguments can be made such that small acreage farms benefit by “leveling the playing field.” Even small acreage farms may have access to insights that were previously only available to the largest acreage farms when participating in a community.

Adoption of precision agriculture technology in Kansas

4.2 What are some of the pitfalls to avoid when using agricultural data?

Farmers are intrigued but skeptical of services offered by farm data companies. More than once the authors have heard farmers share stories similar to the following.

“I wanted to try the farm data company’s service so I decided to do the free (or low cost) option. But to be safe, I sent only a couple fields of data but not my best fields. Also, these fields had seed varieties mislabeled but that’s alright since I know the correct varieties and can translate those in my head when I get the benchmark reports back.”

4.2.1 Assuming a database contains valid data

When working with computer technologies, have you ever heard the phrase “garbage in, garbage out?” Casual conversation with farmers provided insights that they have made the assumption that the majority of their cohort provided high-quality correct information to the data service. These incorrect assumptions lead the farmer to believe that they can reverse-calculate the results to get uniquely useful information for themselves. However, when more than one farmer attempts to skirt the system the overall quality of the results are suspect. Garbage in, garbage out!

Given that more than one farmer has shared similar stories, a farm data pitfall is evident. Assuming the benchmark reports provide yield by variety information for the farmers’ cohort, the results are meaningless even for large samples when a majority of respondents provide incorrect variety. Although exactitude is not required, wrong data is not useful (Mayer-Schönberger and Cukier, 2014).

4.2.2 Truth, lies, and statistics

For some reason, we give a lot of respect to numbers once they are called “data.” The existence of data does not indicate fact or truth, though. Likewise, access to farm data does not explicitly mean that the data are “true.” Documentation of a practice, rate, or yield response has not proven that the practice was conducted, the rate was applied, or the yield responded in a specific manner. An as-applied map collected from an applicator does not prove that the product was applied at those site-specific rates in part due to sensor errors and in part due to intentional misrepresentation. As an example, the sensor systems in nearly all herbicide applicators cannot distinguish applying water from herbicide mixture – they just knew they were applying “something.” The yield map can also be misleading for a variety of reasons, some of which can be intentional. The GPS coordinates of yield data can be altered such that the “field” can be moved to another location.

The intentional misrepresentation of farm data is not the greatest concern when it comes to big data, though. Historically, much of farm record keeping has been conducted for tax compliance purposes and only to a lesser extent as a form of good business practice, including the use of enterprise budgets and physical records for later use (and re-
Most agricultural fields have multiple equipment passes conducting tillage, planting, application, and ultimately harvesting; each field or subfield area receives a potentially different set of products, rates, timing, or other systems or bundles. To be useful in a big data context, these applications, in particular crop protection chemicals, need to be as accurate as possible and preferably digitized and turned into useful data. It is also common for the farm decision maker to not be the person making applications on the field; and the person actually mixing crop protection chemicals may not have incentive to mix the proper rates or record the actual product in a useable manner. Thus, the problem presents itself: how can products and rates be recorded in a manner such that accuracy is ensured?

One example of alleviating this problem is using a combination of direct injection system with sensors that recognize the product being inserted into the applicator (such as Raven and their Sidekick Pro). This can be as simple as the manufacturer of the an input placing a Radio Frequency Identification (RFID) tag inside the product container that is read by the applicator once the container is inserted or comes near the application equipment (such as an RFID tag in an ag chemical container that is read by the spray rig when the container is seated on the rig). As with most any other precision agriculture technology, if human intervention or interaction can be avoided with an automated process, the opportunity for error is reduced and the value of the data captured increases.

4.2.3 Ensuring data quality

To avoid some of the error problems discussed so far, farmers need to devote effort to making sure their data is well-maintained. Yield monitor data management workshops are typically offered by Extension specialists in several states each year. These workshops focus on data quality from collection of that data in the field to post-processing data cleaning.

Researchers have made substantial progress in automating yield data cleaning processes over the last decade. It should be noted that reliable processes do not remove high and low yield data points just because the yields are high or low. Instead, they evaluate machinery dynamics to determine if the yield monitor was able to make an accurate measurement under current conditions. A tool publicly available since 2004 is the USDA ARS Yield Editor software (Drummond, 2012) that is commonly used in workshops offered by Land Grant University Extension systems around the USA and the world.

The initial reasoning yield monitor data management and cleaning workshops were offered was so that participants would learn how to clean yield monitor data from their fields (or their customer’s fields). Even though that objective remains, the majority of participants will never actually perform these processes themselves because they felt the costs of cleaning the data were greater than the expected benefits of their time and effort. Farmers and crop consultants sometimes desire to learn what the process entails to increase their understanding (and confidence) in services that others were providing. Three secondary objectives eventually became the prime reasoning for offering these workshops.

• Learn how combine harvester operation affects yield data quality and how to improve overall data quality
• Understand that manually cleaning all fields takes substantial amount of time such that automated routines are likely the only feasible solution
• Learn how to tell when yield monitor data has not been properly post-processed by data service providers claiming to have cleaned and analyzed data

Discussing these objectives in more detail can help farmers recognize some important steps in making sure they are collecting the best data they can for their operations.

4.2.3.1 Learn how combine harvester operation affects yield data quality, and how to improve overall data quality

Combine harvesters were originally intended to reap and thresh grain. Today, one could argue that “collect data” could be the third intent. For decades, the combine operator was trained to operate the machine to harvest grain and unload to another machine. When the operator is concerned with collecting quality yield monitor data during harvest, the combine operator may also have “data acquisition” responsibilities as well. In general, it is advised that the harvester be operated under conditions similar to those under which the yield monitor was calibrated. This implies maintaining a steady flow of grain into the harvester. It could be argued that ground speed would need to be changed to maintain consistent volume of grain. And that ground speed changes must be gradual and not abrupt. If the operator abruptly changes ground speed, the yield monitor data may spike. Data cleaning software could examine both yield monitor data and vehicle speed to “flag” the data as being in error. Without sound data cleaning procedures, though, the data would “sneak” through and could lead to wrong conclusions about yields in the area. The moral of this particular
example is the behavior of the combine operator impacts the sample size and quality of the yield monitor data, meaning data cleaning is an important practice.

4.2.3.2 Understand that manually cleaning all fields take substantial amount of time, and although automated routines are not always perfect, are likely the only feasible solution to data cleaning for all fields.

Properly cleaning yield monitor data is somewhat tedious and time consuming. It's been estimated that it takes an experienced human 30 to 45 minutes to process each field. When there are only one or two fields this is not a major concern, especially if there are only a handful of on-farm experiments that the farmer is interested in analyzing (for example, the farmer was trying a new plant variety or fertilizer application and wanted to see its impacts on yields). However, when the desire is to clean all fields then the time requirements may be prohibitive. Although many people view automated processes less reliable than experienced humans at their best, research suggests that the costs of lower perceived accuracy with automation is outweighed by the time and human capital costs of manually cleaning data. In addition to desktop software commonly used for data cleaning, automated processes operating behind the scenes by some farm data companies are being introduced. For good reason, farmers and their advisors are becoming more comfortable with automated yield monitor data cleaning and processing.

4.2.3.3 Learn how to discern when yield monitor data has not been properly post-processed by data service providers claiming to have cleaned and analyzed data.

There are many companies providing precision agricultural services. Social media has a steady stream of yield and soil maps displaying the prowess of the services offered. It is unlikely that all precision agriculture data companies are providing the same quality of services, though. A few tell–tale signs of sufficient data quality can help a farmer note when has been properly post processed. These tell–tale signs, or “litmus tests” as they are called, include headland issues and flow delay settings. When flow delay has not been correctly set, it can be said that the yield data appears “out of focus” with jagged edges along regions of changing yield. Although flow delay is a relatively simple adjustment to make, many services do not adequately set flow delay. Learning these and other cues can help farmers determine when additional data cleaning is needed.

4.3. Considerations in “going off the grid” or joining the agricultural data community.

All the decisions and risks associated with farm data can be frustrating. When considering options that farmers have regarding farm data they can 1) farm with 1970s varieties and equipment, 2) not farm, or 3) participate in at least some level of farm data. To be competitive in the long run, though, the first two options fall out of the running fairly quickly.

The option of farming without equipment that contains embedded sensors and modems that transmit the data collected by the sensors will quickly become more and more challenging. As mentioned elsewhere in this handbook, a significant proportion of tractors and harvesters built since 2011 have been equipped with telematics that wirelessly transfer data between the equipment and the cloud. While this has affected lots of new equipment, over time, it will also affect more and more of the used equipment fleet as well. This means eventually the vast majority of equipment will be data and communications-enabled, and farmers wanting to farm “off the grid” will have to manually deactivate those systems. All of this means the first option, “going off the grid”, might not be a viable strategy as the agricultural big data systems mature. In other words, avoiding farm data capabilities may increase farmers’ risk and put them in a position of competitive disadvantage.

Consider an analogy from the industrial revolution (Griffin and Shanoyan, 2017). During the infancy of agricultural mechanization, tractors equipped with internal combustion engines became commercially available. Farmers were faced with farm management decisions of adopting the technology or continue farming with only animal power to pull their implements. Farmers utilizing only animal power gave up potential efficiency gains, and eventually were competed out of business by mechanized farmers. With animal power, farmers were largely self-sustaining by feeding their animals with grain produced probably on the same farm. With mechanized agriculture, farmers gave up some of their self-sufficiency and independence by becoming dependent upon a fuel supplier. Today, it is safe to speculate that a majority of farmers in developed nations would not trade their mechanized farm practices for self-sustaining animal power.

Looking back, the players that gained from the transition from animal power to mechanized power can be identified. The farmer saw gains in profitability and quality of life, although not by using the same farm practices as
before. The farmer was better off once modern technologies were adopted and utilized to their potential. Additionally, the equipment manufacturing industry came into existence. Fuel suppliers also gained by this transition. Modern farmers and farm data services are likely to follow similar patterns as mechanized versus animal powered farms. You could make the same arguments for the Big Data revolution as for the industrial revolution. Successful farms using sensor and data analysis technologies will look differently than they did before. New segments of the agricultural industry will emerge that never existed before. Farmers will give up more of their control and self-sufficiency to suppliers of products, services, and analysts. However, farms that are positioned to make the most of these opportunities afforded by big data will likely be the most successful. Not all farms today are expected to survive this transition much like not all farms successfully transitioned from animal power to mechanized power. The tradeoff for this may be a loss of some “independence” if independence is defined as the ability to do everything needed for the farm on the farm. Analogous to farms becoming dependent upon fuel suppliers (and thus getting their power from an off-farm source), more engagement with agricultural data service providers means farms are facing tradeoffs that may mean reduced farm data privacy.

Beyond the potential tradeoffs with respect to independence, there are certainly tradeoffs with respect to privacy issues as well. Farmers often wonder if they can maintain any sort of anonymity if they share their data with others. However, in the Big Data world, anonymity is no longer achievable, at least in the same manner as it once was. Even sanitized data can reveal the identity of individuals by combining one or two more layers of (probably publicly available) data (Mayer-Schönberger and Cukier, 2014). Given the prevalence of publicly available geospatial data, data from USDA, and plat maps, recombining these layers with sanitized version of community farm data are likely to disclose all the information that were intended to remain anonymous. Remaining “anonymous” is no longer a viable option in modern commodity crop production once agricultural Big Data systems reach maturity. This can become the cost of “joining the grid” and would presumably be avoided by going “off the grid”.

This discussion brings us back around to whether the costs of participating in a Big Data system are worth the costs. In the case of “joining the grid” (or adopting Big Data tools), the costs are the loss of anonymity and full control over farm data, but the benefits are still not clearly defined. There is a general expectation, based on observations of some current applications in agriculture and other industries, that by “joining the Big Data community/service farmers will gain access to previously unavailable insights which will enable better decision making and enhance their competitiveness. So to rephrase the question posed in the title: If a farmers decides not to incur the cost of joining the grid, what is the benefit that they will forego as a result?

### 4.4 Value of Farm Data in a Community

The value of the primary uses of farm data when kept in isolation to a single farm have been described in Chapter 2. But what is the value of that data when it is combined with the data from hundreds or even thousands of farms? That value is what comes from the “secondary uses” of farm data.

**Secondary (or “re-use”)** of farm data are typically shared such that the data becomes part of a community suitable for big data analytics. It is this community of farm data that, when appropriate analytics are applied, offers the greatest value for all the pertinent parties including farmers, aggregators, analysts, retailers, and society. These communities of farm data become networks once sufficient amount of data and farms have joined.

The value of networks is a function of the number of members, or here the number of farms providing data. The emphasis is on the number of farms rather than the number of acres. One thousand acres of 1,000 acres each are more representative of Big Data process than one 1,000,000 acre farm. One of the reasons for this is that with many farms, the farmers’ characteristics become a variable in the Big Data Analytics models.

Given that one of the defining characteristics of Big Data in agriculture is combining farm data from multiple farms into a community of data, each farmer becomes a variable in the analysis. If farm data were isolated to only a single farm, then no opportunity exists to evaluate the management practices specific to any given farmer. Farm data must be aggregated to perform these community analytics. Recall the example from Chapter 1 of G x E x M analysis. A leading example of community analysis is evaluating how a product (G for genetics from classic varietal or hybrid tests) in a given location (E for environment including soils, weather, and other uncontrolled factors) under the farm’s production practices (M for management including controlled factors such as planting dates, seeding rates, timing of operations, tillage practices and many others). When farm data are not aggregated across numerous farms, then the data remain ‘small’ and the value is limited since the M is not a variable but held constant. When data are aggregated such that M is a variable within the GxEzM analysis, greater insights can be discovered for all participants. Examples of previously unknown discoveries may include which product or bundle of products (seed, fungicides, planting dates)
maximize profitability for a given region under specific farm production practices. Other potential discoveries include which “different” seed varieties perform similarly enough to be considered “the same” variety such that the cheaper version would be preferred to other options.

Given Big Data is a network of not only farmers but other players in the agricultural industry, we need to consider how each of the different players benefit from the agricultural big data system. Understanding the economics of networks is important to fully understand the value gained from the Big Data community. The value of the data community depends on not only the quality of the data but on how many others participate in the system. Consider the example of the telephone. The value of the telephone was zero when there were only one telephone. As the number of telephones that were networked together increased, so did the value of the telephone system. The value to the 10,000th telephone customer would have been greater than the first customer. Data from multiple farms aggregated into a community are more valuable than data from any one individual farm. Once the data community has a critical mass of farms, i.e. the long run, farmers’ bargaining power with the data aggregator is greatly reduced. During the infancy of farm data, farmers have more negotiating power than at any other time over its life cycle. In the long run, the aggregator (the party collecting the data together) controlling the flow of data enjoys the majority of the value. Groups offering analytic services enjoy some value, especially in the short run.

It is clear that farmers are not the only participants that have a potential value capture from the Big Data system. Players such as input manufacturers, retailers, and advisors have their own levels of ability to capture value. It is important to realize 1) the farmer is not the only player at the Big Data table and 2) while the farmer may be able to capture some the value from agricultural data, other parties are likely to capture a share of that value that exceeds the farmer’s. From the farmers’ perspective, the overall question is how to make the most of the Big Data opportunity which also implies how to reduce risk of being taken advantage of.

There are legitimate concerns that farmers have about sharing data within a data community. Although going off the grid is not likely to be a sustainable farm management decision, the farmer must decide with which data services to interact and at what point in time to join. For the time being, it remains uncertain which if any of the currently available data service providers will be successful in the long run. Therefore, the optimal farm management decision regarding joining a data community (such as joining a data aggregation service) relies upon the individual farmer’s expected benefits relative to their costs. Once the benefits clearly outweigh the costs, then farmers are likely to join the data service provider. If the farmer is in doubt, though, the optimal decision is likely to wait rather than take immediate action.

4.5 Farm Management Implications of Farm Data on Farmland Leasing Example

In the US, most farmland is owned by the farmer; however, substantial percentages are owned by someone other than the farmer. In the most recent USDA Census of Agriculture, 62% of farmland was owned by the farmer–operator. The percentage of rented farmland has ranged from 35% in the 1960’s to nearly 43% in 1992. Rented farmland proportions are higher in the Mississippi Delta, Corn Belt, and Plains states than the rest of the country (USDA Census of Agriculture 2012). Therefore, a primary focus of farm management economists have been on acquiring and maintaining control of farmland. Precision agricultural technologies can be a useful tool for farmland management, especially considering how farm data can be utilized.

Historical yield, soil test results, and other farm data are commonly included in farmland sales or rental agreements, but these data have not been observed to influence the value of farmland directly. These data may be printed maps, georeferenced digital data, or even annual whole-field yield written on the back of an envelope. Regardless of the source, the data may have somehow helped prove historical productivity and soil amendments utilization but likely have not impacted the value of farmland other than reducing some transaction friction in the leasing of the property itself. This is analogous to providing oil change records for an automobile; the used car has the same value with or without those records but it may speed up the sale. However, when there was enough data to be considered “big” as in “Big Data,” the impact of farmland values must be reevaluated.

Once the agricultural big data industry is mature, farmland values and rental rates are expected to be a function of the presence of site-specific data and “meta-data.” Meta-data includes other data not commonly passed along with the land such as but not limited to: seeding depth, cultivar, machinery diagnostics, time and motion of the equipment, and the dates of tillage, planting, scouting, spraying, input application, and harvest. The management of the individual tract of land may depend upon this data along with data from nearby and potentially distant farm fields. Management of other farm fields may be influenced by presence of data from the given field such that data availability may impact the
whole farm system.

It is likely that a farmer who does not have a history of true Big Data for their fields may opt to pay a premium to secure at least one field that includes these data. One reason is that some data services require farmers to submit data, and farmers desiring to participate in the big data system may be more likely to pay higher rental rates or purchase price for data-endowed land. In the long-run, though it is expected that a “data penalty” (an amount deducted from the market price of land) would be observed rather than a “data premium” (an amount added to the market price of land). Initially premiums may be observed until the agricultural Big Data system reaches maturity.

While we are discussing the link between data and farmland values, it may be useful to point to some other values of agricultural data when the data is not used in Big Data analysis but rather has little or no analysis. Take, for example, the printed yield map. Printed yield maps can have some value in helping visualize the relative productivity of various areas of a field, but has little value beyond that. The one exception is that some landowners appreciate printed yield maps, especially when presented in a format such as framed like a picture suitable for hanging or as kitchen table placemats. Even though not all landowners find value in receiving printed yield maps on coffee cups at the end of the year, many may cherish the linkage to their home farm and ultimately could make the difference between continuing to farm that tract or be outbid by a competing farmer. The overall farm management principle here is that farmers who get to know what makes their landowners happy can position themselves better to maintain and enhance that relationship (assuming some level of utility maximizing behavior). Some landowners view their investment just as that, an investment, and value the revenue stream only (i.e. profit maximizing). Other landowners may enjoy telling their friends about their asset, the history, and current events expressed through printed maps, either framed or imprinted on a coffee mug or perhaps some other creative expression.

4.6 Closing thoughts on data valuation
Farmers should weigh the costs and benefits of participating in any farm data service. As usual, the costs are easier to measure than estimating the potential benefits. At the very least acknowledge the roles of farmers are changing, as well as for the crop consultant, salesforce, and manufacturers of technology. However these changes have been the one constant over the history of production agriculture.

Chapter 4 references
As discussed in Chapter 4, the value of data to a farm depends on many variables. It seems obvious, though, that many companies think that data generated by farmers is quite valuable to them. Agribusiness and technology firms have invested billions of dollars in acquiring and developing farm data aggregation and analysis systems. This level of investment shows these companies’ belief that aggregated farm data holds significant value. Many farmers have taken note of this, and wonder not only how they can capture the value of the data they generate for themselves, but also what value they may be losing through the sharing of their data.

Farmers may be justified in wondering about what protections are available for the data they generate and data generated about their operations, particularly with respect to what legal protections exist. Courts often struggle in applying existing laws and previous rulings to modern technology, and the case of farm data is no exception. For example, current laws provide limited privacy protections from UAS. States have also struggled how to handle access to social media and other digital accounts (which include accounts that could be used to store and share farm data) upon a person’s death when the terms of service may limit access to only the original user. As of 2018, no existing laws address farm data ownership or implications of misappropriation of that data. However, in the long run, case law and/or federal legislation are likely to be the deciding factor in determining the rights and protections associated with farm data. Alternatively, state legislatures may help by clearly defining rules to guide courts in handling new technology.

If there are no current state or federal laws that clearly address what rights a farmer might have to agricultural data, the questions become “are there any laws that do apply to agricultural data” and “do those laws give farmers any rights with respect to agricultural data?”

### 5.1. The basics of intellectual property law

There is an old adage in property law that says “property isn’t stuff; it’s your legal relationship to stuff.” So when we ask what “rights” a farmer has with respect to agricultural data, we have to ask what kind of “stuff” agricultural data is, and what rights the law might allow an individual to have with that “stuff.”

First, we have to think for a moment about what it means if we do consider agricultural data to be “property.” An entire area of the law is devoted to studying the rights that come with various forms of property and how those rights are created, shared, or transferred among parties. The rights someone has with respect to property are usually defined in terms of what powers the law will recognize for them to do some or all of six things with the property:

- **Possess**
  
  With real estate, “possession” includes the right to be physically present on the property. With personal property (like a car or a computer), it includes the right to have physical control over the property. However, with intangible property — property that cannot be touched or held, like data — it refers more to the ability of a person to have access to property.

- **Use**
  
  “Use” implies the ability of a person to interact with or alter the property. A farmer who plows a field is altering the field in a way. A party downloading data, uploading it to a service, or editing it could be considered to be “using” it.

- **Enjoy**
  
  The right to “enjoy” property means something a bit different from the usual
understanding of that word. In property law, the right to enjoy property means to profit from its possession or use. Selling a crop you produced on land you own or rent is your “enjoyment” of the property. Deriving some form of profit from data would be enjoyment of the data.

Exclude

“Exclude” is a counterpoint to “possess” and “use” in that it is the right of the owner of property to keep others from possessing or using it. Full “exclusive possession” of property means an owner is the only person with the right to possess or use the property. As discussed elsewhere in this handbook, it can be quite difficult to exclude other people from data, especially when multiple copies of the data exist.

Transfer:

“Transfer” refers to the ability of a person to sell, lease, or give their property away while they are alive, but it also refers to the ability to leave binding instructions about who is to receive their property when they die (such as a will or trust).

Consume or destroy:

If you own a sandwich, it’s pretty easy to think about how you could consume that property; you eat it. But how do you consume land? You might be able to eat it, but if you were to mine something like coal from it, you would be using up one of its resources. What about data? You could argue data cannot be used up, but it could be destroyed if all copies of it were eliminated. If you had your crop records in one notebook, with no other recording of that information elsewhere, you could destroy the data by destroying the notebook. But if the information was digital and had been shared with a crop consultant or a cloud service, it could quickly become almost impossible to truly destroy the data, since you could never really know if all copies of the data everywhere had been destroyed.

Looking at these six factors with respect to property shows us that while it is easy to envision how they work with physical property like land or equipment, they become more challenging to understand when talking about property that doesn’t have a physical form. While agricultural data can be stored on a physical device like a computer or a memory stick, the data is not the device — it is simply information stored on the device. In and of itself, agricultural data is information and does not have a physical form. Since it can’t be touched or held, it is referred to as “intangible property.” Intangible property includes lots of things from ideas to financial assets. When we speak of intangible property that includes information or ideas, we categorize it further as “intellectual property.”

Since agricultural data could be considered a form of intellectual property, the intellectual property framework is our starting point to define what rights a farmer might have to agricultural data. Under American law, intellectual property can be divided into four categories: (1) trademark, (2) patent, (3) copyright, and (4) trade secret. The first three areas are governed by federal intellectual property law since the authority to pass laws regarding them is explicitly given to Congress by the U.S. Constitution. The fourth, trade secret, is governed by state law. Let’s examine each are to see if it could be used to define the rights a farmer might have to agricultural data.

5.1.1 Trademark Law

Trademark won’t provide a useable legal tool to define any rights to agricultural data because it was never meant to apply to information like data. Instead, the the Federal Trademark Act [1] (sometimes called the Lanham Act) defines “trademark” as “any word, name, symbol, or device, or any combination thereof … to identify and distinguish his or her goods, including a unique product, from those manufactured or sold by others and to indicate the source of the goods, even if that source is unknown.” Examples of trademark include product names, such as Coca-Cola®, but also the design of symbols of the company — in this case, the design of its contoured bottle or the white script writing of Coca-Cola on a red background. While you might be able to create a logo for your farm and get trademark protection for it (meaning other companies could not use the logo — only you could), trademark doesn’t fit at all for agricultural data.

5.1.2 Patent Law

manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor …” In other words, if you invent something, you can patent it. A patent gives you the exclusive right to make the thing you have patented. While most people think of patents in terms of physical items such as machines, patents can also apply to processes (as in the method used to make something or conduct a form of business).

Generally, for an invention to be patentable, it must be useful (capable of performing its intended purpose), novel (the invention is new and different from existing knowledge in the field), and “non-obvious.” Non-obvious is somewhat difficult to define, but as set forth in the Patent Act says: “a patent may not be obtained… if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.” Put another way, if just anyone could have come up with the invention, it is not patentable; some unique skill or creativity must have been involved.

Patent law is a poor fit for defining rights in agricultural data since it protects inventions and not information. Raw data, such as agricultural data, fails to satisfy the definition of invention under the Patent Act as discussed in Digitech Image Technologies, LLC v. Electronics for Imaging, Inc.: “Data in its ethereal, nonphysical form is simply information that does not fall under any of the categories of eligible subject matter under section 101 [of the Patent Act].”

It should be noted patentable inventions could come from the analysis of agricultural data. Examination of agricultural data could give someone an idea for a device or process they wouldn't have had without the data. While this does not mean the data itself is patentable, it does suggest that farmers may want to examine any data sharing agreements they have to see what the agreement says about inventions derived from the shared data.

5.1.3 Copyright law
The best place to start with a discussion of copyright law and agricultural data may be the federal Copyright Act itself:

Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device. Works of authorship include the following categories:

1. literary works;
2. musical works, including any accompanying words;
3. dramatic works, including any accompanying music;
4. pantomimes and choreographic works;
5. pictorial, graphic, and sculptural works;
6. motion pictures and other audiovisual works;
7. sound recordings; and
8. architectural works.

One critical thing to pull out of this section of the Copyright Act is that it protects “original works of authorship.” For intellectual property to be protected by the Copyright Act, it must have to have some creative element. The term “original works of authorship” has been interpreted to require some element of creative input by the author of the copyrighted material.

This requirement was highlighted in the case of Fiest Publ’ns, Inc. v. Rural Tel. Serv. Co., where the U.S. Supreme Court held the Copyright Act does not protect individual facts. In Fiest Publications, the question was whether a pure telephone directory (consisting solely of a list of telephone numbers, organized alphabetically by the holder’s last name) was copyrightable. Since the directory consisted solely of pure data and was organized in the only practical way to organize such data (alphabetically by last name, the Supreme Court held the work did not satisfy the creative requirements of the Copyright Act. In other words, raw facts and data, in and of themselves, are not copyrightable. However, an author can add creative components to facts and data such as illustrations, commentary, or alternative organization systems and can copyright the creative components even if they cannot copyright the underlying facts and data. To borrow from the Court’s opinion in Fiest Publications:

It is this bedrock principle of copyright that mandates the law’s seemingly disparate treatment of facts and factual compilations. “No one may claim originality as to facts.” This is because facts do not owe their origin to an act of authorship. The distinction is one between creation and discovery: The first person to find and report a particular
fact has not created the fact; he or she has merely discovered its existence. To borrow from Burrow-Giles, one who
disCOVERs a fact is not its “maker” or “originator.” “The discoverer merely finds and records.” Census takers, for example,
do not “create” the population figures that emerge from their efforts; in a sense, they copy these figures from the world
around them. Census data therefore do not trigger copyright because these data are not “original” in the constitutional
sense. The same is true of all facts -- scientific, historical, biographical, and news of the day. “They may not be
copyrighted and are part of the public domain available to every person.”

Factual compilations, on the other hand, may possess the requisite originality. The compilation author typically
chooses which facts to include, in what order to place them, and how to arrange the collected data so that they may
be used effectively by readers. These choices as to selection and arrangement, so long as they are made independently
by the compiler and entail a minimal degree of creativity, are sufficiently original that Congress may protect such
compilations through the copyright laws. Thus, even a directory that contains absolutely no protectible (sic) written
expression, only facts, meets the constitutional minimum for copyright protection if it features an original selection or
arrangement.

Put another way, the facts that hydrogen has an atomic number of 1 or that the number of ABC Plumbing is
555-1234 are not copyrightable, but an article about hydrogen in an encyclopedia with an original layout of the facts
about the element, or a Yellow Pages® ad with ABC Plumbing’s number along with a graphic and description of their
services are.

To bring the discussion back to agricultural data, raw agricultural data itself would not be copyrightable, because
it would simply represent facts without any creative component. That said, agricultural data can lead to copyrightable
works even if the underlying data is not protected itself. For example, the “as-applied” data from planting, fertilization,
and pesticide application may not be copyrightable, but a report summarizing the data and adding recommendations
for action might be.

There also continues to be much debate about the extent of copyright protection and ownership for “works”
created by copyrighted works such as software embedded in agricultural equipment — could a report that is
automatically generated by a computer program (when the program itself is a copyrighted piece of software) be
copyrighted? [12] Again, then, it is incumbent upon those disclosing farm data to include language in their
agreements with the receiving party to define the rights to such works derived from the data.

5.1.4 Trade secret law

While trademark, patent, and copyright do not appear to fit as models for farm data ownership and protection, trade
secret at least has the potential to fit. However, defining whether farm data (or any information, for that matter) can
be considered a “trade secret” presents a challenge, as courts routinely make observations such as “[t]he determination
of whether information constitutes a trade secret is a highly fact-specific inquiry” and “[the] same information that
qualifies as a trade secret under one set of facts may not be afforded protection under a different set of facts.”

For centuries, the common law of England (which was “imported” to the United States after its independence)
recognized trade secrets as having at least some property rights attached to them. For many years, states either drafted
their own statutes dealing with trade secrets or relied upon cases interpreting the common law to define trade secret
issues. Then, the National Conference of Commissioners on Uniform State Laws sought to codify and clarify much of
the common law of trade secret and authored the Uniform Trade Secret Act in its first form in 1979, later amending
the uniform law in 1985. As of this writing, all but two states have adopted the Uniform Trade Secrets Act (“UTSA”).

Under the UTSA, a “trade secret” is defined as:

… information, including a formula, pattern, compilation, program, device, method, technique,
or process, that: (i) derives independent economic value, actual or potential, from not being generally
known to, and not being readily ascertainable by proper means by, other persons who can obtain
economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the
circumstances to maintain its secrecy.

We have seen so far that neither trademark, patent, or copyright law would define the rights someone might have
with respect to agricultural data. Could trade secret define those rights?
5.2 Can trade secret law be applied to agricultural data?

As set out above, for a trade secret to give someone rights to exclude others from access to it, the UTSA requires the trade secret to meet the following three conditions:

1. It must consist of information, including a formula, pattern, compilation, program, device, method, technique, or process
2. It must derive independent economic value, actual or potential, from not being generally known to or readily ascertainable through appropriate means by other persons who might obtain economic value from its disclosure or use
3. It must be the subject of efforts that are fair, proper, or moderate under the circumstances to maintain its secrecy. What is fair, proper, or moderate under the circumstances to maintain data’s secrecy will depend on the circumstances in each case.

When applied to farm data, these three conditions help clarify how agricultural data can be considered a trade secret. Could farm data be considered a trade secret under these three conditions? Consider the UTSA definition application of these conditions to growing corn using farm data the farm data used to grow and harvest a corn crop.

5.2.1 Is the manner and strategy in planting, harvesting, and otherwise cultivating the crop a formula or pattern?

Arguably, yes, they are. The data collected over many years and fields of crop rotations, applied input products and rates, timing of applications, and equipment settings collectively itself becomes a holistic pattern that may not be readily observed without advanced analyses. Taken together, a strong argument can be made that these elements constitute a business process, and business processes are routinely recognized as protectable trade secrets.

5.2.2 Does growing corn in this manner derive economic value from not being generally known?

In good years, absolutely. It is generally accepted that a farmer’s specific process for growing and raising corn is based in part on experience (i.e. data) from previous years (soil conditions, fertilizer use, irrigation practices, etc.) that is “generally not known or readily ascertainable” to other people in or outside of the industry. This knowledge can increase production and/or improve efficiencies, thus providing economic value to the farmer although profits will approach $0 in the long run. However, showing that data has value is a necessary but insufficient condition for satisfying this element of the UTSA. Rather, one must show that the data “derives independent economic value, actual or potential, from not being generally known to or readily ascertainable through appropriate means by other persons who might obtain economic value from its disclosure or use.” Put another way: is a farmer’s farm data more valuable because other farmers don’t know it?

On one hand, if a farmer has discovered a truly unique way to grow corn with fewer inputs or to drive yields upward, he or she could capture additional profit in the short run that would not be available if all other producers knew it. Conversely, if other farmers could get access to data to show that the farmer was underperforming on a crop share lease, they might be able to bid leased land resources away from the farmer. Thus, one could argue the farm’s data does have value from being secret.

On the other hand, though, one could argue that anything done in corn production is a mere variation on farm processes that are subject to significant research and examination available to the public and thus are “readily ascertainable.” This argument relies upon the idea that the practices, including the type, rate, and timing of crop inputs are well-known and publically available from public research institutions such as the Land Grant University System and Cooperative Extension systems. The courts have recognized that where information is publicly available in other formats (e.g. an individual’s address and contact information in a phone book), the data is not protected by trade secret law because that data is “readily ascertainable.” Consider a public database that contains information on dairy cow genetics – this information would not be considered a trade secret, nor protected as such, because of the ease by which it can be ascertained. The increasing proliferation of data in different forms have become a focal point where current law needs improvement.

One could also argue that farm-specific data are just that: specific to the farm. The extension of this argument is that the data does not have economic value to another individual producer (though the value of aggregated data is discussed below) because it is essentially meaningless when applied to a farm with different soil types, hydrologic
conditions, micro-climate considerations, and so on. While it might provide some general information to another
individual producer, it would quickly lose most if not all value if the owner tried to apply the information to his or her
own farm. Put another way, the owner’s manual of an electric car is of limited value to the owner of a diesel pickup.
The basic techniques of steering and changing a flat tire are the same, but putting diesel or an extension cord into the
wrong one will fail to yield desirable results. Similarly, trying to apply prescriptions from one farm to another may
yield marginal improvements at best, or be damaging at worst. Given all this, there are significant questions around
whether the second condition of trade secret protection could be satisfied with the data from the example.

5.2.3 Does the farmer actively protect data like it is a secret?
The UTSA requires the putative owner of a trade secret to undertake reasonable efforts to keep the data secret. The
specific reasonable measures that a court would require to consider farm data as a trade secret is uncertain; however
it is known that courts look for active measures taken to ensure privacy. The following steps may potentially be
considered as “reasonable measures” to protect farm data as a trade secret while the data is “on the farm”:

- Properly screen prospective employees for attitudes on confidentiality and proprietary information.
- Ensure that employees understand the importance of maintaining confidentiality in data on the farm.
- When creating backup copies of data, make sure no other entities have access to these backups.
- Restrict employee access to sensitive information.
- Introduce password protections for electronic servers and files.
- Regulate visitor and employee access if possible to physical locations where sensitive data may be accessible.
- Conduct ongoing employee training on the measures used to protect farm data,
- Require a majority vote by farm operators before data are shared with a third party.

An additional issue arises when employees leave the farm operation or are otherwise no longer employed. The
farmer should ensure that any access that the employee had to farm data is denied, which may mean changing
passwords, access points, etc. At the employee’s departure, the farmer should consider conducting an exit interview
including reviewing the signed NDA and ensuring the employee knows that the obligations regarding data secrecy
remain in effect in the future.

All of this may sound like a significant amount of time and expense, and it is. Farmers would have to place a
high value on the confidentiality of the information to justify the expense of such measures. Further, while there are
measures farmers can take to protect data while it resides on the farm, in an era of automated telematics and wireless
communication, that protection becomes even more difficult.

One could argue that farmers, landowners, and their advisors have historically not employed reasonable efforts
to maintain the secrecy of farm data or their farming practices. At first blush, many would assume the production of
corn that occurs outdoors would not have a reasonable expectation of privacy. However, a farmer who has grown and
harvested corn on the same farmland for several years such that they obtained unique knowledge of how to profitably
produce corn on that particular tract of farmland has a good argument that the data is a trade secret, as long as
reasonable steps are taken to maintain the secrecy of that information.

So the question posed by the third criteria for trade secret protection has two components: Do farmers take steps
to actively protect their data from disclosure, and perhaps more importantly, can farmers take those steps?

First, do farmers treat their data like a trade secret? At least up until now, the answer would appear to be “no.”
While they may not have broadcast their data for all to see, farmers, landowners, and their advisors historically have
not employed efforts to maintain the secrecy of farm data or their farming practices. With the dramatic increase in
discussion of farm data and its prospective economic value since the early 2010s, however, many farmers have started
to ask questions about how they could keep their data private. As their awareness of data collection and transmission
technologies increases, some farmers may be more active in seeking to secure their data.

The second question may be the most important: can farmers take steps to keep their data secret? At one level,
the answer to this question is tied to the fact that much of agricultural production occurs in the outdoors. A farmer
producing corn outdoors would not have a reasonable expectation of privacy, making it extraordinarily difficult to
argue that he or she is even capable of employing reasonable efforts to keep something secret when it is readily
observable from a public vantage point (which also goes to the arguments of the second condition as to whether the
secret is readily ascertainable). Still, though, there may be elements of the operation that are not easily observable or
determinable without direct access to the farm data that is not readily observable to the public but is rather collected
and stored within the farm’s telematics systems.

Importantly, though, data is not merely stored in telematics systems; it is often transmitted to equipment
manufacturers through automated processes requiring no operator intervention whatsoever under an “opt out” (as opposed to “opt in”) arrangement. Since 2011, many new tractors roll off the assembly line equipped with cellular modems that automatically transmit a broad range of machine parameters back to the manufacturer. While these arrangements can provide many benefits to equipment owners (such as enhancing the effectiveness of preventative maintenance and shortening equipment downtime), one could also argue that the farmer has no claim to a trade secret in farm data since a significant amount of the data composing the “secret” has already been shared with other parties through the equipment’s telematics systems.

The disclosure of data to another party is not necessarily the “death knell” for any trade secret claim, though. Trade secrets are routinely disclosed to other parties while maintaining their status as trade secrets, so long as an agreement to retain the secrecy of the data is maintained. The value of trade secrets would be diminished if they could not be shared with consultants and other advisers aiding the trade secret owner in determining how to maximize their use of the secret, and thus, trade secret law recognizes that trade secrets can be disclosed to others so long as appropriate measures are taken to make sure the party receiving the information also keeps it secret. “Information known by persons in addition to the trade secret owner can retain its status as a trade secret if it remains secret from others to whom it has potential economic value… the precautions required of the trade secret owner may increase with increasing dissemination.” (American Law Institute, 1995).

The question remains, though, as to whether the user agreements between equipment manufacturers and farmers would be sufficient to satisfy the “secrecy efforts” requirements of the UTSA and cases interpreting it. Granted, “the owner [of a trade secret] is not required to go to extraordinary lengths to maintain secrecy; all that is needed is that he or she takes reasonable steps to ensure that the information does not become generally known” (Smith, 2006). A non-disclosure agreement (NDA) between the farmer and the equipment manufacturer could be written to include sufficient restrictions on the uses made of the data as to satisfy the UTSA.

Conversely, though, most equipment agreements are written as by equipment manufacturers as “disclosure” agreements rather than non-disclosure agreements and define what manufacturers can do with farmer data and to whom they can disclose that data rather than defining what they cannot do with the data and what steps must be taken to safeguard it from disclosure to other parties. Some manufacturers have gone so far as to suggest the farmer doesn’t truly own the equipment itself (to say nothing of the data embedded within it) but rather operate under the theory that the tractor is simply subject to a “user agreement.” Determining whether an equipment’s user agreement would provide sufficient secrecy requirements to satisfy the third condition of trade secret protection is a case-by-case determination, but it is unlikely that many current user agreements would pass muster on this point. A farmer could still be able to execute a non-disclosure agreement with the manufacturer that would satisfy that condition.

Viewed together in this example of corn production, one can see it is unlikely that many farmers are actively taking the steps that would be necessary to claim trade secret protection for their data. However, it can also be seen that under the right circumstances, a farmer could take those steps if they felt the increased value they would enjoy from retaining the exclusivity of their data would outweigh the cost of maintaining that secrecy. If they take those steps to protect their ownership of the potential trade secret, what do they get?

5.2.4 What damages are available if “trade secret” agricultural data is taken or misused?

The consequence of securing trade secret status for agricultural data is that the farmer can pursue a “misappropriation claim” when farm data has been used in a way not allowed or has been acquired by an unauthorized party. To successfully bring a misappropriation claim, the farmer must first prove that the data was indeed a trade secret as discussed above. Next, the farmer must prove that the data was misappropriated, or wrongly obtained and used by another party. The court will look at whether there were reasonable measures in place to ensure the secrecy of the data for the farmer to prevail in a trade secret lawsuit. Although popular opinion seems to indicate farmers are expected to be endowed with ownership rights, the courts are likely to provide legal guidance unless laws are created that classify farm data.

If a farmer can prove a misappropriation claim, the court may allow them to recover one of the three types of damages:

1. Actual Damages: These damages include lost profits to the farmers, typically calculated as net profits (i.e. gross profits minus operating costs). For example, if the data service provider used the data to manipulate prices of agricultural inputs and extracted profits directly from farmers, they could then potentially recover actual damages from the service provider.
2. Reasonable Royalty Rate: These arise from the rate of return to the farmer that would have occurred had the data service provider properly negotiated a licensing agreement for the data at the point in time of misappropriation. This assumes that the farmer (who ordinarily may not have been willing to license this trade secret) did so willingly for a bargained-price.

3. Unjust Enrichment: These damages include all the benefits the data service provider gained through misappropriation of the farm data. An example of this would be the profit the service provider made in the sale of the yield estimates.

5.3. Can agricultural data be “owned?”

So far, our discussion has covered the basics of property and what it means for intangible, intellectual property such as agricultural data to be considered property. We’ve seen classification of something as property usually gives parties certain rights with respect to the property. So does all that mean a farmer “owns” their data?

The question of who owns farm data goes back at least to the advent of precision agriculture in the 1990s. Data ownership, privacy, and security have cyclically been hot topics since then but have recently peaked with the introduction of “Big Data.” There is a reason the ownership issue is particularly important in agriculture. A big part of growing and maintaining a farm is the acquisition and retention of assets like farmland and equipment. Farmers are taught from an early age (and often by hard experience) to clearly understand what rights they do and don’t have with respect to their property. So, naturally, when there was a sharp spike in the number of parties talking about how valuable agricultural data is, farmers instinctively started asking if they owned the data and what rights that ownership carried. When discussing ownership of physical goods such as commodities, machinery, and farmland, it is intuitive what “ownership” means, but Farm data does not fit our preconceived notion of “ownership” like these physical examples.

So, perhaps the best answer to the question of whether a farmer has ownership rights with respect to the data he or she generates on their farm, but depending on how the data is generated and shared, others may also have some ownership rights with respect to the data, too. If that answer sounds confusing, that’s OK, because it is — and it is confusing because the words “own” and “ownership” convey a lot of meanings that are sometimes unintended. As the character Inigo Montoya says in the film The Princess Bride, “You keep using that word. I do not think it means what you think it means.”

Let’s talk about the economic implications of agricultural data ownership for a bit. Just as the legal ownership concepts become a bit challenging to apply to intellectual property like agricultural data, economic principles applied to agricultural data have different characteristics than physical goods. For example, copies of digital data can be made at almost zero cost and are indistinguishable from the original data. Put another way, it may be expensive to produce agricultural data, but once you have it, it costs almost nothing to reproduce it. Given that copies are identical to each other and the original, very minimal control exists over what happens to that data once copies have been made available to another party. Multiple entities (e.g. farmers, landowners, input suppliers, soil sampling services, aggregators, lenders, etc.) may have partial access to viable copies of the same farm data.

What does this mean for the economic value of owning agricultural data? To explore that, we have to apply some economic principles about “public goods” versus “private goods” and “excludability” versus “non-excludability.” Ownership of “private goods” implies that the owner may exclude others from enjoying their good. “Public goods” are not privately owned and no one can exclude others from enjoying these goods. To fully understand this, we have to dive in a bit further and explore the concept of “rival” and “non-rival” goods need to be considered and applied to farm data. Private goods are typically “rival” in that if one person is using them, that means no one else can be using them. On the other hand, agricultural data are considered “non-rival” because the consumption or usage of data by one person does not alter another person’s ability to consume or use that same data. Classic examples of non-rival goods are books and movies; multiple people can watch Star Wars without any loss of value to any other viewers. Economic theory suggests that there is no loss in value to a person who has a copy of Star Wars on DVD or digital download triggered by someone else also having a copy of it. Examples of agricultural non-rivalrous goods include weather reports, commodity market information, and agricultural data — the value that the initial user receives from accessing data or information is not affected by another user accessing the same information. Multiple entities can consume farm data without diminishing the initial value enjoyed by the first or subsequent users of that data.

Since agricultural data are non-rival and digital (i.e. copies can be made at near-zero cost), that means we must accept that farm data are “non-excludable” once other users have access to it. Given that data copies are considered
identical to the original, the ability to exclude others from accessing data no longer exists once another entity has a copy (i.e. access). Most privately held goods typically are excludable; however the lack of excludability does not necessarily mean that farm data are a public good.

Since both the legal and economic principles make “ownership” of farm data difficult to define, maybe a better question regarding farm data may be “can I keep others from accessing data that only I control?” The answer is “yes, but only until the data are shared with a third party or aggregated into community.” Privately held farm data may be considered excludable if and only if it is in the possession of the party that generated the data. However, once the farm data has been shared with another party or aggregated into a Big Data community, it is no longer possible to exclude others from the data.

Therein lies the problem. To maximize the potential value of agricultural data it typically needs to be aggregated into a community for Big Data analysis to provide greater insights than it can provide when just analyzed in isolation on the farm. As a result, the notion of 1) excluding others from accessing data and 2) making the most valuable use of data are at competing ends.

If there is a potential that agricultural data is worth more to everyone (including the farmer that may have generated it) when it is shared, does that mean farmers should automatically share their data with everyone? Not necessarily. Farmers need to consider the costs and benefits of data sharing at each decision step. First, if a landowner asks for a copy of data generated from their farmland, it is likely most farmers would opt to maintain a positive working relationship with the landlord and would share the data. Second, data remaining in “data tombs” or “data silos” (for example, yield monitor data that is simply sitting in the combine’s computer and has never been accessed or analyzed) specific to an individual farm has essentially zero value until the data are converted to farm management information. When properly analyzed, farm data shared in a community has much greater value to not only the community but the individual farms.
Chapter 6
Protecting agricultural data

Chapter 5 discussed the rights farmers may have to agricultural data. To the extent farmers have those rights, part of protecting those rights is making an effort to restrict access to the data to parties that have an agreement in place with the farmer to access and use the data. Additionally, regardless of how a farmer may be sharing his or her data, there is value in making sure the data is not taken, lost, or corrupted. In this chapter, we will discuss three paths to securing agricultural data: maintaining physical, “cyber,” and legal protections.

6.1. Managing the physical security of data
When we discuss the “physical” security of data, we are referring to restricting the ability of an unauthorized party to get a physical copy of the data. In the past, when farm records were in notebooks or ledgers, that meant getting a copy of the notebook or ledger. In those days providing physical security could have been a simple as putting them in a safe to which only the farmer had the combination. Now, perhaps ironically, physical security might be the least important form of security for data, since the information is digital and may exist in many locations simultaneously.

Nevertheless, there are important things farmers can do to prevent unauthorized parties from getting physical access to agricultural data, and there are ways to use physical controls to restrict access to digital data. Putting aside for the moment the fact that many kinds of agricultural equipment transmit data electronically, they also store copies of the data in the equipment itself. Thus, one way to use physical controls to protect data is to restrict access to that equipment. Keeping equipment locked and in secured areas can prevent parties from getting access to equipment where data could be taken or corrupted. Removable devices that store sensitive data (such as flash drives, hard drives, and other storage devices) need to be kept in a secure, lockable location.

Some other tips for maintaining the physical security of data include the following:
- Store confidential information in areas not frequently used by personnel unauthorized to access the information.
- Keep confidential information in locked storage rooms and cabinets.
- Restrict access to such areas to only those with a “need to know” the information.
- Provide sign-out/sign-in registers for confidential information.
- Utilize camera/alarm systems to alert of unauthorized access.
- When disposing of confidential information, physically destroy hardcopies by shredding.

While we often worry about an unauthorized party sneaking onto our property and stealing, destroying, or corrupting data, we should also make sure the people that are on our property every day understand how to protect data, like our families and employees. As mentioned above, here are some steps you can take with the people who work on your farm to secure data:
- Properly screen prospective employees for attitudes on confidentiality and proprietary information.
- Ensure that employees understand the importance of maintaining confidentiality in data on the farm.
- When creating backup copies of data, make sure no other entities have access to these backups.
- Restrict employee access to sensitive information.
- Introduce password protections for electronic servers and files.
- Regulate visitor and employee access if possible to physical locations where sensitive data may be accessible.
- Conduct ongoing employee training on the measures used to protect farm data.
6.2 Managing the cyber-security of data

Physical security of data can still have some importance, but in modern times, far more threats come from unauthorized parties trying to take, destroy, or corrupt data. While threats from all kinds of sources make some farmers want to go “off the grid” and completely disconnect from any form of Internet or cellular data transmission, there are also significant advantages to being connected.

A farmer or consultant could upload a task data file or download an as-applied map by physically connecting a storage device (such as a USB drive) to the task controller through a connection on the tractor/implement bus. However, that same data transfer could be accomplished by adding a cellular modem to the tractor/implement bus or to the tractor ECU. With such a connection, not only could data be shared throughout the tractor-implement combination in real time; it could also be shared with a dashboard at the farm’s headquarters and/or with a consultant, who could make “on-the-fly” adjustments and upload prescription adjustments in real time as well. The system could also transmit the data to a cloud storage service for future use and analysis.

Many farmers select specific systems and service packages specifically to share their telematics and agronomic data with a third party, but others worry about how other parties could take such data without their consent. What pathways exist for such a taking? First, could someone “hack” the tractor/implement bus by physically connecting to the system? The answer is, “theoretically, yes.” A number of commercially-available technologies allow farmers to plug into an implement network and access CAN messages directly. For example, one could purchase a CAN message reader to read machine diagnostic codes for repairs. Someone wishing to “steal” data would likely not want to be present to read the data, though, and would likely prefer to use a CAN data logger coupled with a device to wirelessly transmit the data. Many data loggers are available to the public as well; for example, the “Snapshot®” device used by Progressive Insurance for some insurance programs is simply a CAN data logger plugged into a vehicle’s On-Board Diagnostic (OBD-II) port.

While such an approach would work for standard messages transmitted over the equipment’s bus, it would not work for “proprietary” messages (messages using a language or syntax specific to the equipment type or brand). To decode such messages, the prospective hacker would have to develop a system for decoding the information being provided from the task controller for the implement, and that task would take almost as much work (if not more) than the work in developing the task controller system in the first place. Note, though, that several companies now provide means for re-engineering proprietary CAN messages (such as those related to crop yield) so farmers can automatically transfer yield data to the cloud; such technology could also be used to decode other proprietary information.

What about the GPS system as a way for “hackers” to get information from equipment? In most tractors, harvesters, and other powered equipment, the GPS system connects directly to the task controller. As a result, a “bug” might receive information about the commands sent to the implement, but without the associated location data, the command information is rendered meaningless. The bug would require its own GPS receiver along with implement data (the configuration and dimensions of the implement), which today could be done for relatively modest equipment cost. Suffice it to say that obtaining agronomic data via a physical connection to an implement poses a task manageable for someone knowledgeable in ISO 11783 technology. However, building and deploying such a device poses a significant amount of effort (to say nothing of the potentially-criminal trespass involved in getting it onto a farmer’s equipment) in relation to the prospect of collecting data on only one farm.

Admittedly, most farmers rightly put little thought into their systems being physically hacked, but worry instead about their data being accessed through an intercepted cellular signal. Thankfully, some elements of the cellular communication system make that difficult as well. First, virtually all cellular signals are encrypted when transmitted and decrypted at the cellular tower; without the decryption key, interpreting any data transmitted would be quite difficult (although not impossible for a sophisticated hacker; recent news has highlighted the ability of some governmental agencies to do so). The use of data encryption through a secure socket layer (“SSL”) protocol by the farmer and his or her service provider adds another difficult-to-break security barrier to interception of the data.

Some producers also worry that another party could be feeding harmful commands to their equipment through a similar signal. Beyond the encryption issues discussed in the preceding paragraph, a number of built-in safety features make this difficult. First, as mentioned previously, the ISO 11783 standard provides a safety mechanism to prevent harmful commands being issued from the implement to the tractor. Second, the cellular modems connected to the tractor bus for sharing machine health information either cannot send commands to the ECUs in the system (because of the physical configuration of the system) or the ECUs are programmed not to accept commands from the modem. As mentioned above, though, systems can be configured to accept commands in the J1939 standard, such as
truck powertrain management systems. Again, though, most agricultural systems are not configured to accept such commands, although this may change in the near future.

While these basic conditions of the farm data transmission environment protect against most casual threats to farm data security, farmers can still take measures to increase data security. When using services that involve cellular data transmissions, farmers should ask the service provider what encryption methods are used by the network. Farmers should also ask what computer security measures are used by the receiving company. Data encryption occupies an ever-increasingly important role in securing information as computers become increasingly mobile and connected. For example, the receiving company should at least be using Secure Socket Layer protocols to ensure there are no “eavesdropping” systems obtaining data through the connection and/or that an “impostor” system is on the receiving end of the data transmission.

Even though there are several factors making it difficult to hack agricultural equipment to take, delete, or corrupt data, some farmers may still be uncomfortable with their equipment’s potential to “talk” to equipment manufacturers or other parties. If that is the case, consider the following steps:

• Ask equipment vendors for a detailed listing of the sensor and wireless transmission systems embedded in any equipment.
• Do not select any undesired sensor or wireless transmission systems when making option selections on the purchase of new equipment.
• Request the deactivation of any wireless transmission systems on the purchase of existing equipment (or new equipment where such systems are standard) and confirm the deactivation with a third party technician.

Beyond these considerations for how parties might access transmitted data from farm equipment, important agricultural data may be stored on computers at the farm as well. These steps can help manage cyber-security for those computers:

• Install and routinely update anti-virus and cyber-security software.
• Require password access to any computers storing confidential information and restrict password access to “need to know” employees.
• Immediately change passwords when an employee leaves or after any indication of a potential breach.
• Use computer software that alerts of any copying or downloading of confidential information.
• Monitor employees’ use of email or the Internet to detect any unauthorized transmissions of confidential information.
• Do not store confidential information on Internet-connected devices (admittedly this becomes increasingly difficult as more and more devices are Internet-connected).

While such procedures might provide some modicum of protection against the accidental or intentional disclosure of confidential information, they also pose a number of problems. Additional layers of security can hurt the efficiency of employees who do need to access the data in utilizing it, and blocking Internet access or deactivating wireless transmission systems means eliminating the myriad advantages in quickly sharing data with the vendors and consultants providing profit-enhancing services to the farm. Given this, the aforementioned practices may be of limited use in a highly-connected agricultural operation. Thus, the more economical application of security may lie in clearly establishing secure means of data transfers. To a large extent, determining the methods of data transmission used on a farm involves a significant amount of consumer research on the part of the farmer prior to purchasing equipment with embedded sensors and transmission equipment or separate sensing systems. It also involves monitoring the ongoing use of such systems.

6.3 Legal protections for data

As discussed in Chapter 5, the legal status of agricultural data can be difficult to define. Nevertheless, there are a number of legal tools available to help farmers manage the rights they and other parties may have with respect to agricultural data.

6.3.1 Evaluating user agreements

Whenever a farmer uses a service provider such as a crop consultant or data storage company, or the farmer purchases equipment that comes with some form of data services, they are often required to sign a “user agreement.” Typically, a
user agreement for a product or service involving the exchange of agricultural data will include provisions about what can and cannot be done with the data. However, there often much confusion about these agreements. For example, a recent survey by the American Farm Bureau Federation found 59 percent of the respondents were confused about whether the user agreements between them and a service provider to use the respondent’s agricultural data to market other services, equipment, or inputs back to them.

That agreement may also reference one or more company policies on data disclosure and privacy issues. Many companies offering consulting or data analysis services have company policies addressing various concerns such as confidentiality of the information, specifying to whom the data may be disclosed, and uses that may be made of the data. As an example of these policies, below is an excerpt from the John Deere Privacy and Data Statement:

John Deere understands that you may not want us to provide Personal Information and Machine Data to third parties for their own marketing purposes. We limit our sharing of Personal Information and Machine Data as follows:

We may share Personal Information and Machine Data with our affiliated companies, suppliers, authorized John Deere dealers and distributors, and business partners, which may use it for the Purposes listed above. We may also share Personal Information and Machine Data with our service providers to fulfill the Purposes on our behalf. Our service providers are bound by law or contract to protect the information and data, and to only use it in accordance with our instructions.

We may disclose Personal Information and Machine Data where needed to affect the sale or transfer of business assets, to enforce our rights, protect our property, or protect the rights, property or safety of others, or as needed to support external auditing, compliance and corporate governance functions. We will also disclose Personal Information and Machine Data when required to do so by law, such as in response to a subpoena, including to law enforcement agencies and courts in the United States and other countries where we operate.

Policy statements can have value, but they are only legally enforceable if their text is incorporated by reference into a binding agreement between the farmer and the service provider. This underscores the need for some form of NDA, as discussed below. However, the relative bargaining power between the farmer and the service provider will obviously vary. Negotiating the terms of “boilerplate” agreements large corporations will provide to their customers will likely require high-level collective discussions between industry groups and corporate service providers.

Fortunately, there are some resources available to help farmers evaluate the data management and privacy policies of many companies working in agricultural data. A coalition of agricultural industry groups, equipment manufacturers, and service providers, led by the American Farm Bureau Federation, created a statement called the “Privacy and Security Principles for Farm Data” in 2014. Those principles (also called the “Core Principles”) include the following:

Education: Grower education is valuable to ensure clarity between all parties and stakeholders. Grower organizations and industry should work to develop programs, which help to create educated customers who understand their rights and responsibilities. ATPs (Agricultural Technology Providers, such as equipment manufacturers or service providers) should strive to draft contracts using simple, easy to understand language.

Ownership: We believe farmers own information generated on their farming operations. However, it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders with an economic interest, such as the tenant, landowner, cooperative, owner of the precision agriculture system hardware, and/or ATP etc. The farmer contracting with the ATP is responsible for ensuring that only the data they own or have permission to use is included in the account with the ATP.

Collection, Access and Control: An ATP’s collection, access and use of farm data should be granted only with the affirmative and explicit consent of the farmer. This will be by contract agreements, whether signed or digital.

Notice: Farmers must be notified that their data is being collected and about how the farm data will be disclosed and used. This notice must be provided in an easily located and readily accessible format.

Transparency and Consistency: ATPs shall notify farmers about the purposes for which they collect and use farm data. They should provide information about how farmers can contact the ATP with any inquiries or complaints, the types of third parties to which they disclose the data and the choices the ATP offers for limiting its use and disclosure. An ATP’s principles, policies and practices should be transparent and fully consistent with the terms and conditions in their legal contracts. An ATP will not change the customer’s contract without his or her agreement.

Choice: ATPs should explain the effects and abilities of a farmer’s decision to opt in, opt out or disable the availability of services and features offered by the ATP. If multiple options are offered, farmers should be able to choose some, all, or none of the options offered. ATPs should provide farmers with a clear understanding of what services and features may or may not be enabled when they make certain choices.
Portability: Within the context of the agreement and retention policy, farmers should be able to retrieve their data for storage or use in other systems, with the exception of the data that has been made anonymous or aggregated and is no longer specifically identifiable. Non-anonymized or non-aggregated data should be easy for farmers to receive their data back at their discretion.

Terms and Definitions: Farmers should know with whom they are contracting if the ATP contract involves sharing with third parties, partners, business partners, ATP partners, or affiliates. ATPs should clearly explain the following definitions in a consistent manner in all of their respective agreements: (1) farm data; (2) third party; (3) partner; (4) business partner; (5) ATP partners; (6) affiliate; (7) data account holder; (8) original customer data. If these definitions are not used, ATPs should define each alternative term in the contract and privacy policy. ATPs should strive to use clear language for their terms, conditions and agreements.

Disclosure, Use and Sale Limitation: An ATP will not sell and/or disclose non-aggregated farm data to a third party without first securing a legally binding commitment to be bound by the same terms and conditions as the ATP has with the farmer. Farmers must be notified if such a sale is going to take place and have the option to opt out or have their data removed prior to that sale. An ATP will not share or disclose original farm data with a third party in any manner that is inconsistent with the contract with the farmer. If the agreement with the third party is not the same as the agreement with the ATP, farmers must be presented with the third party’s terms for agreement or rejection.

Data Retention and Availability: Each ATP should provide for the removal, secure destruction and return of original farm data from the farmer’s account upon the request of the farmer or after a pre-agreed period of time. The ATP should include a requirement that farmers have access to the data that an ATP holds during that data retention period. ATPs should document personally identifiable data retention and availability policies and disposal procedures, and specify requirements of data under policies and procedures.

Contract Termination: Farmers should be allowed to discontinue a service or halt the collection of data at any time subject to appropriate ongoing obligations. Procedures for termination of services should be clearly defined in the contract.

Unlawful or Anti-Competitive Activities: ATPs should not use the data for unlawful or anti-competitive activities, such as a prohibition on the use of farm data by the ATP to speculate in commodity markets.

Liability & Security Safeguards: The ATP should clearly define terms of liability. Farm data should be protected with reasonable security safeguards against risks such as loss or unauthorized access, destruction, use, modification or disclosure. Policies for notification and response in the event of a breach should be established.

Following the introduction of the Core Principles, a number of industry stakeholders created a tool called the Ag Data Transparency Evaluator to help verify whether the user agreements and other policies adopted by technology companies were consistent with the Core Principles. Companies wishing to participate in the Ag Data Transparency Evaluator must answer ten questions about how they store, use, and transfer agricultural data; their responses are then reviewed by an independent third party for transparency and completeness and if found acceptable, the company may use the “Ag Data Transparent” seal in their marketing efforts. The ten question are as follows:

1. What categories of data does the product or service collect from me (the farmer)?
2. Do the Ag Technology Provider’s (ATP’s) agreements address ownership of my data after my data is transferred to the ATP?
3. If the ATP contracts with other companies to provide data related services, does the ATP require these companies to adhere to the ATP’s privacy policies with me?
4. Will the ATP obtain my consent before providing other companies with access to my data?
5. After I upload data to the ATP, will it be possible to retrieve my original complete dataset in an original or equivalent format?
6. Will the ATP notify me when its agreements change?
7. Will the ATP notify me if a breach of data security occurs that causes disclosure of my data to an outside party?
8. Upon my request, can my original dataset be deleted when my contract with the ATP terminates?
9. Do the ATP’s agreements establish how long my original datasets will be retained?
10. Do the ATP’s agreements address what happens to my data if the ATP is sold to another company?

The Ag Data Transparency Evaluator can be an important tool to help farmers understand their user agreements with service providers. However, while the numbers of companies participating in it continue to grow, there are still
dozens of companies whose user agreements have not been evaluated. Thus, if you are presented with a user agreement that has not been evaluated, you can use the Core Principles as a starting point to help you evaluate the agreement yourself. Further, no matter whether the company or agreement have been put through the Ag Data Transparency Evaluator, there is no substitute for the independent review of a licensed attorney who is working directly for you if you are concerned with whether a user agreement adequately protects your interests.

6.3.2 Non-disclosure agreements
Perhaps the most important legal tool to protect against the disclosure of information (or to provide consequences if there is a disclosure) is a non-disclosure agreement (“NDA”). Farmers disclosing their data, and service providers receiving it, can enter an NDA in which both parties agree in advance to hold the information confidential and agree to what uses can and cannot be made of the data. Functionally, if the farmer is disclosing data through the use of a service provided by a vendor and has entered into a user agreement for the service, the user agreement may specify many of the things that would be included in a separate NDA, and those user agreements are discussed in the previous section. The following discussion addresses attempts to address some NDA issues by corporate policies, and the provisions that should be considered by farmers when negotiating an NDA with a party to whom they will be disclosing farm data. This discussion presumes at least some parity in bargaining power between the farmer and the service provider or other party receiving the farm data.

The following is a list of items the farmer and his or her attorney should consider in drafting an NDA for the disclosure of farm data to a service provider:

1. Execute the agreement prior to data disclosure: As mentioned in Chapter 5, trade secret law may be the area of law giving the best chance of legal protection to agricultural data. However, trade secret law will not protect information voluntarily disclosed or publicly available. Thus, it is critical the NDA be executed before the disclosure of any data.
2. Define who is disclosing and receiving the information: In most cases, the farmer will be the disclosing party, and the service provider will be the receiving party, though this is not necessarily always the case. In many cases, the obligations of the agreement will be defined the role of the party, so defining when those roles are triggered is important.
3. Define what information will be regarded as confidential: Blanket statements that all information disclosed by the farmer to the service provider may be ineffective as the protection of all information may be impractical or counterproductive to the services provided. As a result, the agreement should define what information is, and is not, to be kept confidential, whether by category of information or the channel by which such information is transmitted.
4. Exclude information that will not be regarded as confidential: By the same token, it may be useful to define what categories of information are not to be treated as confidential and may be disclosed without further consent from the parties. Other information may be disclosable, but only with the express written consent of the party providing the information.
5. Establish a duty to keep the information secret: Perhaps the most important portion of the agreement, an affirmative contractual duty should be established that the party receiving the information must keep it secret. On the other side of the same coin, this portion of the agreement should also explicitly prohibit the disclosure of the information, and should also define the measures the receiving party must take to maintain the secrecy of the information. This portion of the agreement may also be accompanied by a time limit on its enforceability, which is usually defined by an event (such as execution of a release by the party providing the information, or the public disclosure of the information by that party) rather than a period of time.
6. Specifically allowed/prohibited uses of information: This section of the agreement can spell out what uses of the information are specifically allowed, and which are specifically prohibited. The farmer and his or her attorney will wish to use care in making sure that the beneficial uses of the data motivating the farmer to seek the service provider’s services are not blocked by these terms.
7. Data destruction requirements: The farmer may wish to require the destruction of all data transmitted to the service provider in the event of a breach of the agreement by the service provider or some other event terminating the agreement. While there may be merit in such provisions, it should also be noted that data destruction in today’s highly-interconnected computing environment may be a practical impossibility. The most one may be able to achieve is the destruction of any hardcopies of the information and the complete
8. **Provision for injunctive relief**: Proving the case for injunctive relief (that is, an order from a court commanding an offending party to immediately cease a harmful activity such as releasing data, as opposed to the much more common remedy of ordering the offending party to pay monetary damages to the injured party) can be both costly and time-consuming, permitting the farmer to suffer continuing damages from data disclosure until it is stopped. A provision stating that the parties both agree that injunctive relief is appropriate in the specified circumstances can drastically shorten this process and limit the expenses in securing such relief.

9. **Indemnity clause**: The farmer may desire a clause stating the service provider will indemnify the farmer for any of his or her expenses (or the expenses of third parties asserting a claim against the farmer) caused by the wrongful disclosure of data.

10. **Integration clause**: An integration clause will state the entire agreement between the parties has been reduced to writing through the NDA. The effect of the integration clause is to exclude evidence of the parties’ discussions in the negotiation of the agreement and to limit the resolution of any disputes to the language in the agreement itself. If the parties agree to an integration clause, it is critical all of their concerns be addressed in the text of the agreement.

11. **Attorney’s fees**: The “American Rule” in most civil litigation is the parties pay for their own attorney’s fees, unless a statute or other legal rule overrides this presumption. Frequently, contracts override this rule and require the losing party pay the prevailing party’s costs; this is usually an attempt to minimize the chance of frivolous claims by one party. Farmers should use care in the inclusion of such language since it may result in the payment of significant legal fees if they should initiate what is eventually proven to be an unsuccessful claim against the service provider.

12. **Alternative Dispute Resolution (ADR) and venue provisions**: The parties may want to require any dispute among them be first submitted to ADR (arbitration or mediation) before the claim may be litigated. Large corporations often prefer arbitration as it may be faster and less expensive than litigation, but a growing body of research suggests arbitration may favor the corporation over other plaintiffs. The farmer may wish to specify mediation as a first line of ADR. At the same time, many large corporations fear they will be treated unfairly at the hands of local juries, where the opponent will have “home field advantage.” This may or may not be true; by the same token, if there is to be such an advantage, does the farmer wish to relinquish it?

13. **Disclosure under legal process**: One situation in which the receiving party may have little choice in disclosing information is when they are legally compelled to do so. However, there may be disagreement about when a party is “legally compelled” to disclose information. To provide the best possible opportunity for both parties to determine is such disclosure is indeed legally required, many attorneys recommend a fourfold approach: (a) disclosure of the information is prohibited unless the receiving party is subpoenaed or otherwise compelled by some form of legal process; (b) the disclosing party must be given as much notice as possible, allowing them to contest the legal process; (c) the receiving party must use best efforts to cooperate with the disclosing party; and (d) the receiving party may disclose only information which, in the written opinion of its legal counsel, it is required to disclose.

14. **Liquidated damages**: It may be difficult (or even impossible) to determine the amount of damages that the farmer has sustained from the disclosure of protected information. As a result, the farmer may wish to define an amount of liquidated damages in advance. Liquidated damages are simply an amount, agreed to in advance of a contractual breach, to be paid if a breach is proven to have occurred. The counterpoint to liquidated damages is that they serve as both a floor and ceiling to claimed damages; even if a farmer sustained greater damages than those negotiated in the liquidated damages provision, he or she will likely be deemed to have waived any claim to a greater damage amount.

### 6.3.3 Employee data agreements

While we have already discussed some of the physical and cyber-security issues to be addressed with farm employees, there remains the issue of how to handle the potential disclosure of agricultural data by employees who had perfectly legitimate access to the data as a result of their work on the farm.

Many human resources professionals hold investment in robust employee screening and interview processes as the best means of solving employee problems before they even start. With respect to issues of confidential information, consider the following in screening and interviewing potential employees:
• Avoid the targeted recruitment of employees of a competitor; doing so may increase the likelihood of a claim from the competitor that the hiring employer was attempting to gain access to confidential information or create an opportunity for espionage.

• Ask the prospective employee about their attitudes with respect to confidentiality or proprietary information (people “philosophically committed to universal freedom of information” may not be inclined to keep secrets well).

• Ask about the circumstances under which the prospective employee left previous employers who are potential competitors.

• Obtain copies of any documentation related to confidential information the prospective employee may have handled in prior employment, including any confidentiality agreements to which the prospective employee is still bound and any termination statements.

• Discuss any potential limitations on job duties related to confidential information.

Potential employers should be up-front about the potential to deal with any confidential information and should share their employee policies (discussed below) in advance to help “create a culture of confidentiality” even before hiring.

Once an employee has been selected, a thoughtful and clear set of employee policies, fairly and consistently enforced, and codified in an easy-to-read employee handbook, can further this culture of confidentiality. The handbook can provide an important reference for the employee if he or she has questions and cannot reach a supervisor for help, and can also be a reminder of the importance of specific issues (“the boss must think this is a big deal if there is a policy on it.”). Items to be included in an employee handbook regarding confidential information include:

• Advise the employee that some information they will encounter is confidential.

• Underscore the importance to the long-term profitability of the business in keeping such information secret.

• Describe the procedures in place to keep confidential information secret.

• Describe company policies regarding (or prohibiting) keeping any confidential information on the employee’s personal computer, mobile phone, or other electronic devices.

• Outline the disciplinary measures that may be implemented if an employee discloses confidential information or violates confidentiality procedures.

• Clearly identify who the employee can contact if they have a question about how to handle a situation involving potentially confidential information or if they suspect confidential information has been released.

A strong confidentiality agreement to be executed by the employee may be even more important than the policies embodied in the handbook. Such an agreement solidifies the obligation of the employee to maintain the confidentiality well beyond the implied duty of confidentiality and an employee handbook that may not rise to the level of an enforceable contract. Fortunately, a number of examples exist for such employee confidentiality agreements. Although valuable, confidentiality agreements should be carefully crafted to avoid potential legal claims arising out of anti-competitive behavior. Similarly, the confidentiality agreement should not be over-broad; “a confidentiality agreement cannot make a trade secret out of something that is generally known and, for the most part, courts will not restrain competition for alleged, but unproven, trade secrets.” On a related note, non-compete agreements or other restrictive covenants may also be in order to prevent confidential information from being used against the employer, but such agreements are strictly construed and often found unenforceable if too broad.

When an employee with access to confidential information leaves the enterprise, a number of precautions should be taken:

• Interview the employee to determine the reasons for which he or she is leaving (is the departure on good terms?).

• Secure all equipment containing confidential information including company-owned computers, mobile phones, or other electronic storage devices.

• Delete all confidential information from employee’s personally-owned computers, mobile phones, and other electronic storage devices.

• Review the confidentiality agreement with the employee and emphasize the employee’s continuing obligations of confidentiality.

• Obtain contact information for the employee and his or her new employer.
• Change all passwords for systems to which the employee had access; secure any physical keys held by the employee and consider re-keying particularly sensitive areas.

Although many farmers spend most of their concern on external threats to their information security, the most serious threats to confidentiality may lie in their own employees. Even though their breaches may be completely unintentional, employees routinely have access to information that would take extraordinary efforts by an outside party to access. Thus, perhaps nowhere is the adage “an ounce of prevention is worth a pound of cure” more true than in a farmer’s handling of employee confidentiality issues.

6.3.4 Data and farmland leases

A farmer justifiably feels like he or she has some claim to the data generated by equipment he or she owns running over his or her own land. But let’s consider an arrangement affecting at least part of almost every American farm – who has what rights with respect to the data regarding leased agricultural land? The tenant operating the land could easily assert the data belongs to him or her, since he or she expended funds to operate the equipment collecting the data about the agricultural practices he or she implemented on the land. However, the landowner might also assert the data describes attributes of his or her property and thus should belong to him or her. A share-lease arrangement would further complicate the issue (or perhaps clarify it in creating ownership rights in both landlord and tenant).

Whatever the doctrinally-correct answer may be to whether the landlord or tenant owns agricultural data, the most prudent course appears to be crafting an explicit agreement about farm data rights as part of the farmland lease. To this end, Janzen proposes a number of considerations in such agreements. In cases where the tenant is regarded as the owner of the data, such provisions might include:

• Landlord and tenant recognize that tenant’s farming of the leased farmland during the term of the lease will generate agronomic data, including information related to soil, water, seed variety, crop health, crop maturity, disease, nutrients, fertilizer, herbicides, pesticides, yield etc., in various digital forms, including files, imagery, records, video, photos, etc. (“Farm Data”).
• Landlord assigns all rights and interest to Farm Data to tenant and relinquishes landlord’s rights in the same. Tenant is the exclusive owner of all Farm Data generated on the leased farmland during the lease term. Tenant shall have all rights associated with Farm Data ownership, including deletion, transfer, sale, and disclosure rights.
• At the conclusion of the lease, tenant shall assign and transfer all Farm Data from the prior crop year to landlord, or at landlord’s election, the subsequent tenant.

If the landlord is deemed to have “ownership” of the farm data generated, the alternative form would be as follows:

• Landlord and tenant recognize that tenant’s farming of the leased farmland during the term of the lease will generate agronomic data, including information related to soil, water, seed variety, crop health, crop maturity, disease, nutrients, fertilizer, herbicides, pesticides, yield etc., in various digital forms, including files, imagery, records, video, photos, etc. (“Farm Data”).
• Tenant assigns all rights to Farm Data to landlord and relinquishes tenant’s rights in the same. Tenant shall cause all Farm Data to be transferred to landlord on or before December 31st each year by a mutually acceptable method of data transfer. Landlord is the exclusive owner of all Farm Data generated on the leased farmland during the lease term. Landlord shall have all rights associated with Farm Data ownership, including deletion, transfer, sale, and disclosure rights.
• At the conclusion of the lease, landlord shall retain ownership of all Farm Data. Tenant shall delete any copies of Farm Data under tenant’s possession, custody, or control.

As time goes on, data will be an increasingly important part of the “bargain” between the landlord and tenant. Indeed, we may someday see the point when the amount and quality of the data associated with a farm influences the value the landlord and tenant are willing to pay or receive for land. In the meantime, landlords and tenants should keep data issues in mind as part of maintaining a solid, transparent relationship between them.
Chapter 7
Managing data from your farm

Now that we have discussed a number of the opportunities and challenges associated with agricultural data, you’ve got some decisions to make about how your farm will approach data collection and management. If you’ve decided that the benefits outweigh the costs, having a “whole farm data plan” for how you will collect, store, analyze, and retain data will help you make the most of the data you generate on your farm (as well as the data your farm generates without you even realizing it).

Assuming you want to take full advantage of the data tools available for making crop decisions on your farm, you may want to consider the following checklist of data points for your operations.

General considerations
- What is the “field unit” designation you will use for your crop data analysis? That is to say, what name or number will you give to a field so that you can consistently compare data from the same field over time?
- What equipment or methods will be used to record each type of data?
- How will personnel be trained to consistently observe, collect, or record the data?
- If custom applicators or other parties will be involved in planting, input application, or harvesting:
  - Is their equipment capable of providing the data you need?
  - Is the format in which that equipment records data compatible with your analysis needs?
  - How will you obtain the data from them (and what happens if you can’t)?

Planting data
- Variety planted (with as much specificity as possible)
- Seed population (seeds planted per acre)
- Seed spacing (linear distance between seeds)
- Date and time of planting
- All available machine parameters coinciding with planting (see points below regarding machine data)

Soil information - at least soil map, but better is grid sampling
- Publicly available soil map
- Consider grid sampling for higher resolution of soil parameters
- Soil organic matter content
- Soil moisture
- Soil pH
- Atmospheric and water-based nutrients
  - Carbon (C)
  - Hydrogen (H)
  - Oxygen (O)
- Plant macronutrient values
  - Nitrogen (N)
- Phosphorus (P)
- Potassium (K)

• Plant secondary nutrients
  - Calcium (Ca)
  - Magnesium (Mg)
  - Sulfur (S)

• Plant micronutrient values
  - Boron (B)
  - Chlorine (Cl)
  - Copper (Cu)
  - Iron (Fe)
  - Manganese (Mn)
  - Molybdenum (Mo)
  - Zinc (Zn)

• Electrical conductivity (EC) measurements
• Electromagnetic (EM) measurements
• Gamma ray measurements
• Slope and aspect measurements to determine water flow
  - These measures may be available from machine data / real-time kinematics (RTK) data
• Drainage
  - Type of field drainage available
  - Tile drainage installed in the field, if available

Input applications
• For all inputs
  - “As applied” maps
  - Date and time of application
  - Per unit cost (for example, $/ton)
• Fertilizer: specific type and composition (such as N-P-K percentages)
• Soil amendment (pH adjustment such as lime, application of organic matter)
• Pesticide and fungicide applications
• Application of any other materials (for example, animal waste, oilfield waste, municipal waste water)
  - For any such applications, consult a professional for best practices in how to obtain a representative and accurate sample of the material to be applied for analysis of all relevant data points.
  - Be sure to keep a copy of applications with any nutrient management plans (NMPs) or other legally-required application plans.

Plant data
• Crop scouting information
  - Note: increasingly, apps and other tools are available to tie imagery and scouting reports to a geo-reference so that information can be made a data layer coinciding with other geo-referenced information
• Visible spectrum imagery
  - Aerial photography
    » Manned vehicle imagery
    » UAS (“drone”) imagery
• Normalized Difference Vegetation Index (NDVI) sensor data
• Normalized Difference Red Edge (NDRE) sensor data
• Handheld plant health sensor data (for example, Green Seeker data)

Irrigation maps
• As applied maps for all irrigation cycles
• Soil moisture before and after irrigation
• Cost of power for irrigation system by application
  – Include time of use costs for areas with different costs based on the time of electricity use
• Co-applied fertilizer / pesticide (“chemigation”) type, concentration, and composition
• Animal waste application via irrigation system
  – See note above regarding sampling needs for each application of animal waste.

Yield
• Quantity of crop harvested (mass or volume per unit of area, such as bushels per acre)
• Test weight
• Moisture
• Protein
• Foreign material content

Weather data for the crop
• Note that unless you have a remote weather station located at the field, publicly available weather data may be the best information available; depending on the source, this may take some additional data input to correlate the data to your specific field. For an example of these weather data resources, visit the Oklahoma Mesonet website at http://www.mesonet.org/index.php/site/about.
• Note that for weather data, the date and time of the weather event is important. Knowing that you received 1.25 inches of rain in July is important, but knowing whether the 1.25 inches were received over a half hour or a week is also important.
  • Rainfall data
  • Relative humidity / dew point temperature
  • Evaporation potential measures (combination of temperature, humidity, and wind conditions)
  • Temperature
    – Heat units (for example, degree days)
    – Cold stress measures (for example, hours below freezing)
  • Wind speed and direction
  • Severe weather impacts
    – Wind gusts
    – Hail events
    – Tornado paths
    – Flood areas

Machine data
• In general, make use of all data available from every use of machinery in the field, whether it is a tractor, sprayer, combine, self-propelled harvester or combine.
• As applied maps for all field uses including information on the machine’s path (showing overlaps and gaps in passes)
• Engine load
• Ground speed
  – Note that this can be an important factor in planting and input applications, and also requires calibration to take into account slippage; some equipment comes with radar sensors to manage the difference in tire speed and true ground speed.
• Fuel consumption
• Fuel cost
• Routine machine servicing and preventative maintenance (including cost of materials and labor)
• Machine repairs (including cost of materials and labor)
• Machine downtime

Financial data
• Some farm data systems allow the inclusion of financial information such as the cost of the input being
applied, but others do not. Thus, in many cases, financial data will be a separate data “layer” in your records that will have to be manually connected in your analysis to the underlying task (for example, how much labor expense was involved in harvesting a field). This takes some additional effort. However, assuming that your goal in collecting and analyzing farm data is to make your farm more profitable, it’s an effort that is much needed.

- Just as a crop consultant can provide important insights into the data that can drive production decisions, your accountant and tax preparer can provide you with important data points to be collected for your financial decisions as well.
- To get you started on financial measures that may need tracking, a good start is the Oklahoma Farm and Ranch Account Book, available at http://agecon.okstate.edu/farmbook/ . The Cooperative Extension Service for your state will also have a number of resources that may be specific to the crop enterprises common to your state.
- Financial data do not exist in a vacuum; rather, they come from the expenses and revenues of the farm operation. Thus, for each of the data points listed in the previous items in this list, the key is tracking the expenses and revenues connected to them. Maintain records of the costs of each input, and remember all the inputs associated with an activity. Planting seed involves the cost of the seed itself, obviously, but also fuel cost, machinery time, labor, transportation costs to get the seed from the provider to the field, and so on. Planting operations may be just one layer of cost data as you look at the crop, though - other layers would include seedbed preparation, cultivation and post-planting input application, irrigation, and harvesting.
- Interestingly, some farmers are quite detailed in keeping cost information, but spend little effort on the payoff - the revenues received for the crop. It is good to keep records of the payment received for the crop as a whole, but it is better to keep records at higher level of detail to see what profitability looks like at the field level. That means to the extent possible, keeping records of the price received for the grain from field A as distinct from field B. Consider the example of a wheat crop: Field A may have had very different test weights, foreign matter content, and protein than Field B, resulting in Field A receiving the standard cash price (or perhaps even a premium) while Field B may have received a dockage for lower quality. Without keeping field-level revenue information, those differences could be lost as they are aggregated to the whole farm.
- Financial data is perhaps the crucial data layer to the whole analysis, because ultimately it is the piece that enables you to make the farm management decisions to improve profitability on the farm. Consider the example in the previous bullet. You may have assumed that you really want to examine this year’s agricultural data because Field A is clearly superior to Field B. But what if, once you add a financial data layer, you find out that the returns above costs for Field B were actually higher for Field B - even though there was a dockage for Field B grain, so much was spent on inputs for Field A that it was actually less profitable?

Now, let’s discuss some “best practices” in collecting and maintaining that data.

### 7.1 Setting data collection and management procedures for your farm

You probably don’t start planting your crops by just throwing seed in the planter, hoping that the seeding rate just happens to be what you need. With that approach, you run the risk of wasting a lot of seed or running out well before reaching the end of the field. Rather, you likely spend some time thinking about the seeding rate you want for your targeted yield, make sure the planter is calibrated to that rate, and do calculations to make sure you’ve secured enough seed to fit the needs of the field. The same holds true for managing your data. To make the most of it, you need to make sure you are collecting enough data to support the decision tools you will be using, while at the same time not wasting time and resources to collect information that will never be reviewed.

### 7.1.1 Identifying the data sources on your farm - data collected automatically

One of the interesting turning points in the evolution of agricultural data is the fact that instead of having to ask “what data do I need to collect about my farm?” first, the better question is “what data is my farm already collecting about itself?”

As discussed above, many power units, self-propelled implements, and towed implements come from the manufacturer with sensor networks, on-board data storage units, and cellular modems that can transmit data wirelessly
to a number of receivers. Thus, your first step in a data plan may be to identify all of the data points that your equipment is already recording automatically. If you still have the operator manuals for the equipment, those resources can be a good start to help you see exactly what data points are being collected; if you don't have those manuals, a call to your equipment dealer or sales representative can probably get you a demonstration of what data is being collected, along with how to view and/or retrieve that data.

Once you have a grasp of the data points that can be automatically collected by your equipment, be sure to compare that list with the list of data you've determined you need for your farm. Some data will still have to be collected manually as discussed in the next subsection.

Since your equipment may have been recording data about your farm since you started using it, it may have a significant amount of historical data stored either on-board or as part of a data service provided by the manufacturer. As mentioned previously in this handbook, that data can provide an opportunity for further analysis, but only if the data came from properly-calibrated sensor systems or if the data can be “cleaned” to provide some assurance of its accuracy.

While the continuing advancements in data technologies mean more and more data points are automatically collected for us, that doesn't mean we don't still have a job with respect to data collection. We have to make sure sensor systems are properly calibrated (and kept calibrated - maintaining sensor systems is not a “one and done” job).

A final but important point to include in your data plan is what format your data needs. History suggests over time, the many different data formats used in the agriculture industry will settle down into a relatively small number that can be used interchangeably between a large segment of software and hardware providers. In the meantime, though, lots of technology providers have their own data file formats. This means you need to think about the format you need for your own data analysis. If you use consultants or other service providers to help in your data analysis, you need to ask them what data formats they need as well.

An important illustration of this comes from the analysis of harvester yield monitor data. One difficulty that applied researchers and consultants have with yield monitor data is that they often deal with data from a spectrum of yield monitor types, manufacturers and file formats. Put another way, identical fields would yield different data types depending on the brand of combine that harvested them. This is especially true when the analyst requires access to the original raw, unprocessed data, e.g. in “Advanced” export format, in order to have more control over what observations are excluded in the data set (such as when the user anticipates using a protocol with Yield Editor). For many years, working with different yield monitor file formats meant that the user had to maintain several software packages to import, view and export the raw data. The use of Field Operation Data Models (FODM) using FOViewer (Field Operation Viewer, MapShots, Cummins, GA, USA) allows users to handle yield data from different file formats with one software package that integrates with precision agriculture mapping software such as EASiSuites (MapShots, Cummins, GA, USA). Although the original intent of FOViewer was as a “test platform for companies that are implementing support for field operations data models” (http://www.mapshots.com/FODM/fodd.asp), it has proved useful for analysts and technicians working with a host of hardware and software. So, as you consider your data plan, also consider what data tools you will be using for analysis, and what formats those tools will require.

### 7.1.2 Identifying the data sources on your farm - data collected manually

The good news is that more and more data points can be automatically collected by our farm equipment. The bad news is there are still a lot of data points that must be kept manually. That fact makes having a data plan all the more important. Good decisions require consistent, accurate data. When data collection is manual, it’s a ball that can be dropped in the hustle and bustle of planting, cultivation, and harvest operations. Having a data management plan to which you frequently refer keeps your data needs in front of you, rather than “out of sight, out of mind.”

In the preceding subsection, we talked about the need to keep sensor systems properly calibrated to make sure data is accurate and consistent. The same goes true for the “human sensor systems” you use to collect data. Just as sensor systems can have glitches, we’ve all heard too many times the phrase “human error.” Have procedures for how information is to be observed and recorded so the data will be the same no matter who records it. Those procedures should include training for how the data should be observed and collected. For example, if you have on-farm storage of harvested grain, you are probably the first elevator your crop will see, so the people who work on your farm need training to properly record data points like test weight, protein, and foreign material. Since taking those measurements will also involve several pieces of equipment, the people recording the data will likely need to be trained in how to keep the equipment calibrated and properly maintained, too.
Another point to remember is that automatic data collection systems sometimes require a little human help. Make sure that everyone on your farm is trained to recognize the importance of accuracy in data, even if that means an extra few seconds here and there versus going with the “default” settings. For example, your planter can only provide accurate “as planted” data if the person running it makes sure they accurately specify what specific variety of seed was put into it. It’s not surprising that a review of data collected by planters revealed that the data indicated the most frequently planted variety was either the default variety for the computer or the very first variety listed on the computer system’s drop-down manual. Similarly, many sprayers rely on the operator to accurately specify what they are spraying – as mentioned above, they don’t know if they are spraying pure water or full-concentration pesticide without a user input.

7.2 Managing the security of your data
Chapter 6 discussed a number of ways to protect your data from unauthorized access, but it can be equally important to have a security plan for your data to protect it from damage, alteration, or destruction by physical and network-based hazards as well.

7.2.1 Securing your data against physical hazards
Data may be physically stored in a memory drive on the equipment generating the data, such as a hard drive on the tractor, implement, or harvester. In many cases, those drives have the ability to download their stored data to a flash drive or other memory device that you can connect to the equipment. Thus, it is a good idea to backup the data from your equipment after each operation (for example, after a field is planted, after an input application, after harvesting the field, etc.). If the equipment is connected to a cellular modem that transmits the data to a data collection or management service, it may be worthwhile to use the service to maintain a server-based backup, and also to download the data after each operation as well.

Once you have the data from each operation backed up to a physical storage device like a laptop or flash drive, that drive needs to be kept in a physically safe place. Electronic storage devices are best kept in a space with low humidity, relatively cool temperatures, and low levels of dust. That may be a place like your farm office, and is probably not a place like the floorboard of your pickup. Drives storing important data should also have restricted accesses (kept in a locked storage area) and / or be clearly labeled to someone does not accidentally delete or overwrite the data.

All farm records - whether electronic or not - are susceptible to hazards such as severe winds, tornadoes, floods, and fires. Important data should thus be kept in the part of your home or other building that is most hardened against those threats. The good news about electronic records, though, is they can be easily replicated and another copy kept elsewhere. Thus, another good practice is to duplicate records at regular intervals (at least quarterly if not monthly, and after major events such as planting and harvest as well) and to keep a copy at a separate site that is some distance from the farm - for example, in your safe deposit box at a bank. The odds of a tornado hitting your farm house are low, but the odds of a tornado hitting your farm house and your bank 12 miles away are much lower, and the bank vault may also be able to withstand a direct hit from a tornado. Multiple copies of data in multiple locations increases the odds that it can survive a hazard.

7.2.2 Securing your data against cyber-threats
Of course, one way to eliminate the threat of physical hazards to your data being stored on the farm is to store it somewhere else, and potentially dozens of other places simultaneously by storing it with an online server service, often called “cloud” storage. Of course, online server services rely on just that – a physical server that has to be located somewhere, and nowhere is completely physically safe. Thus, before selecting a cloud server, ask how many server locations the service maintains and what efforts they take to physically secure their servers. Most cloud services will maintain multiple server locations and keep copies of customer data on multiple services specifically to avoid damage to one server destroying all copies of a customer’s data.

In addition to the steps cloud services take to manage physical security of data, many cloud services also use sophisticated anti-virus systems and network security tools to make sure only authorized parties get access to data. Be sure to ask about the security protocols used by any cloud service you are considering.

Cyber security is something you must take seriously at home as well. Even if you store your data on a cloud service, a computer virus on the farm computer can mean the virus finds its way into your files before they are uploaded to the cloud service. Some cloud service systems will detect such viruses, but some will not. Make sure
you install anti-virus software on your farm computer and keep it updated. Only allow authorized personnel to use computers storing important data, and make sure they are trained in responsible computer usage. Such training should include how to recognize threats such as email fraud, “phishing” (attempts to get a computer user to voluntarily provide personal information that can be used for fraud), and warning signs that a website may be illicit or that a potential download could carry a virus.

7.2 Managing data generated on leased land
As discussed in previous chapters, some landlords will want access to the agricultural data generated on the land they lease to a farm operator. The major differences in the data management for these lease arrangements versus data management for land you own likely come from the data the landowner wants (or perhaps the summaries of the data instead) and arrangements for the sharing of the data.

7.2.1 What data is to be shared?
The type and amount of agricultural data a landlord may want will likely vary with the type of leasing arrangement and the sophistication of the landlord when it comes to crop production. For example, in a share-crop lease arrangement with an experienced agricultural producer, the landlord and tenant are functionally in a kind of joint venture, and will make many management decisions together. If the landlord in this kind of arrangement wants to be forward-leaning in using the full range of data to aid in decision-making, he or she may want access to all the data generated with respect to the land he or she owns. Data for everything from planting to harvest can help provide information about the proper calculation shares for both cost and revenue items and - if properly handled - can provide objective and transparent information for both parties about the arrangement, which should reduce the potential for conflicts in the lease relationship. As mentioned earlier in this handbook, the lease should carefully spell out what data are to be shared. If a written lease is already in place, new data-sharing provisions can be included at the next renewal or included as an amendment to the lease if agreed to by both landlord and tenant.

Another example may be an heir who received the farmland from an experienced operator with a share-crop lease, but the heir has little or no knowledge of agriculture. There may be a potential for confusion on the part of such heirs as to how shares are calculated or why the share fractions are what they are. In such cases, agricultural data on inputs, costs, and revenues can be a useful tool to demonstrate the reasoning behind the shares and to verify the amounts of crop delivered for payment of the lease. The far end of the spectrum could be a cash rent lease arrangement with a landowner who inherited the farmland but has little to no knowledge of crop production. Such landowners may not know what data they need, or “don’t know what they don’t know.” The most important data to them may be the numbers on the annual rent check. Even in such cases, agricultural data can be used to help educate the landowner about agricultural practices and to demonstrate the stewardship of the land by the tenant.

7.2.2 How is the data to be shared?
As with the kinds of data to be shared, the way in which data is to be shared is probably a function of the data sophistication of the landlord. Landowners with an advanced data management capability may request access to a shared cloud storage service or an equipment manufacturer data service so they can directly access data, perhaps even as it is generated. Such systems can do much to automate the data-sharing process. At the same time, though, if all of the tenant’s data is stored on such a service, sharing access to the service might mean accidentally sharing all of the tenant’s data with the landlord. Unless the service used provides a way to “partition” data and only allow sharing with respect to certain field units, it may be necessary to manually select and download the data to be shared with the landowner and turn it over via a separate file-sharing service or by providing a physical drive (such as a flash drive) to the landowner.

Other landowners may want reports generated from the data in hard copy. If so, the software used to collect and analyze your data may be able to automatically generate the reports needed if you have done a good job of collecting and managing the data needed for the report. Still other landowners may just want a yield map, which is what they view as the ultimate demonstration of the productivity of the land. If your landlord would like a yield map, certainly, provide one, but you have a “teachable moment” to use all of your agricultural data to show the work and analysis you have put into making the land as profitable as possible.
Whatever data is agreed to be shared by the party, the lease should specify the method to be used in providing the data. This gives both parties assurance of how they can expect to receive and provide information, and gives them clear expectations about what they need to do to fulfill their data obligations under the lease.
Agriculture has seen its share of revolutions in the past century, shifting from animal power to petroleum power, the advent of chemical fertilizers and pesticides, the green revolution in plant breeding, and the potential of genetic engineering to bring new crop traits to market that unlock tremendous potential to produce larger crops with fewer inputs than ever. The use of agricultural data stands poised to be the next great revolution in agriculture, with the potential to help producers of all sizes and enterprises produce more with less. With the prospect of 9 billion people to feed with decreasing land, water, and other resources, farmers will need every tool at their disposal to be profitable and sustainable. Certainly, farmers need to evaluate the costs and benefits, opportunities and challenges of agricultural data tools as they determine how best to guide their operations into the future. We hope this handbook will help you with those choices.

Chapter 8
Conclusions