Develop a Data System for Your Pilot Plant

Mark Fisher
Thermo Systems, LLC

The data requirements of a pilot facility are just as important as the plant's physical components. Follow this guidance to design and deploy a robust process data system for your pilot plant.

The pilot plant stage of process development provides valuable data that must be transformed into actionable information to guide the design and operation of a commercial facility. This stage involves a diverse group of stakeholders, including engineers, scientists, and executives, each with their own unique information needs. A process data system that collects and analyzes data from your pilot plant must meet the requirements of each stakeholder.

This article explains how to design and deploy a process data system for a pilot plant. It identifies the various stakeholders involved in a pilot plant project, the data and information needs of different stakeholders, and the diverse types of data that the system needs to handle. The article lays out a step-by-step process to develop the pilot plant data system and discusses some of the challenges that may be encountered along the way.

A pilot plant data system

The design and implementation of a pilot plant data system is a project unto itself. It may be part of the overall pilot plant project, but it should be treated as a separate project. The disciplines and focus of the personnel who design and deploy these systems are, for the most part, different from those who design and deploy the physical components of a pilot plant.

A dedicated project manager should be assigned to oversee the project. If the company does not have in-house expertise in this field, a qualified and experienced systems integration (SI) firm should be contracted to manage, design, and deploy the system.

* Editor's note: This article is adapted from a paper presented at the 2015 AIChE Spring Meeting and 11th Global Congress on Process Safety.

Data vs. information vs. knowledge vs. wisdom

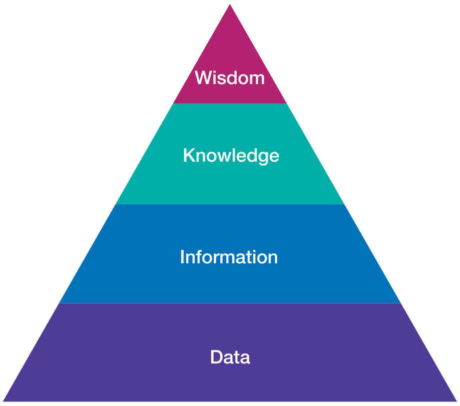
A distinction should be made between data and information (Figure 1) *(1–2)*.

Data are raw facts. They have no meaning in and of themselves. They are neither right nor wrong. They just are. Examples of data include:

- the object is red, round, and weighs 4 oz
- the pressure in the vessel is 125 psi.

Information uses data. When data are processed, organized, structured, and/or presented in a given context, they become information. Information can be right or wrong, depending on the context. Examples of information include:

- the object is a fruit. This information eliminates some objects, but does not give enough context to determine exactly what it is. It could be an apple or a tomato; both are fruits. If, however, more context is given, such as the



▲ **Figure 1.** The DIKW pyramid illustrates the relationship between data, information, knowledge, and wisdom. Data forms the foundation of the pyramid.

object grows on vines not trees, then it becomes clear that the object is a tomato, not an apple.

• the pressure is increasing but is below the pressure rating of the vessel (150 psi).

Knowledge uses information to come to some conclusion. When multiple pieces of similar information are combined and used as the basis to draw a conclusion, they become knowledge. Knowledge is the understanding of patterns of information. Examples of knowledge include:

• even though a tomato (the object) is a fruit, it is rarely used in fruit salad

• if the pressure continues to increase above the vessel's rating (150 psi), the vessel will burst.

Wisdom uses knowledge to correct, alter, or prevent a situation. Wisdom enables you to take action based on knowledge. For example:

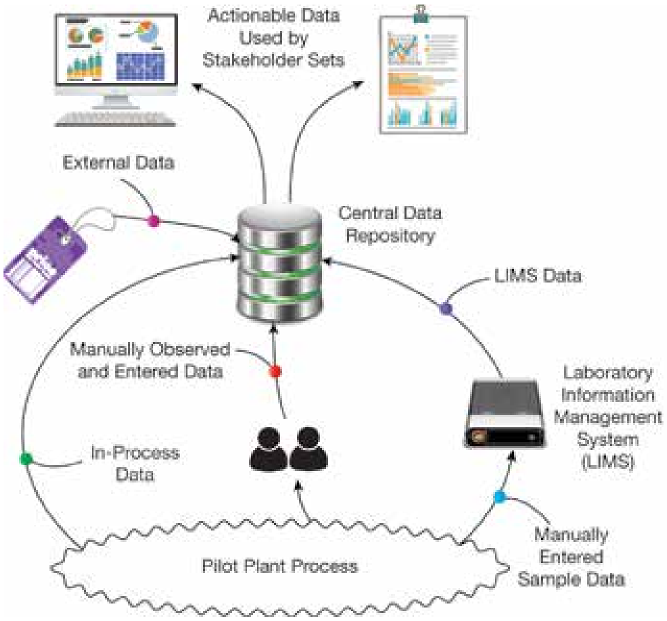
• I won't put tomatoes in my fruit salad

• I will open the pressure relief valve on the vessel to relieve the pressure before it bursts.

Actionable information. The term "actionable information" is used throughout this article to refer to data that have been transformed into information from which knowledge can be gleaned and then judged to take a wise action.

Diverse stakeholders and data sources

When developing a pilot plant data system, it is important to keep in mind both the end (data out) and the beginning (data in). At the end are multiple stakeholders with diverse information needs; at the beginning are data of many different forms.



▲ **Figure 2.** The data system for a typical pilot plant involves four types of data, each of which must be transferred to a central data repository before it is transformed into actionable information.

Multiple stakeholder sets. In many situations, sometimes the best place to start is at the end. This is true of data transformation in a pilot plant. It must be understood that several different types of people will need access to the data from your system. These groups of people are referred to as stakeholder sets. A typical pilot plant data system may have many stakeholders:

• scientists
• engineers
• maintenance personnel
• operators
• supervisors
• quality assurance personnel
• plant executives
• corporate executives.

Your particular pilot plant may include some or all of these stakeholder sets and maybe others. The important thing to remember is that each set will have its own requirements concerning how the data stored in the system are transformed into actionable information.

Divergent data sources. With the end in sight, now let's focus on the beginning. Most pilot plants have more than one source of data. A typical pilot plant involves four divergent sources of data (Figure 2).

In-process data are data that are measured in real-time at the pilot plant process. They are generated from field instrumentation installed directly in or near the process stream to safely and efficiently monitor and control the pilot plant process. Although the data from some of this instrumentation may be used to meet the information requirements of some of the key stakeholder sets, its main purpose is process control. Note that some in-process instrumentation may be installed for the sole purpose of meeting a key stakeholder set's information requirements and have nothing to do with actual process control. This instrumentation is almost always wired in to the process control system that monitors and controls the pilot plant process.

Laboratory equipment data. It is impossible to measure some variables in or near the process. For these types of data, a sample must be taken manually (or automatically in some instances) and analyzed in a laboratory. The data from the laboratory must then be entered into the pilot plant data system's central data repository, either manually or, preferably, via some automatic electronic transfer mechanism.

Manually observed and entered data. Some variables and conditions cannot be measured at all, even with available laboratory equipment. To obtain this type of data, an operator may need to tour the plant at regular intervals and enter their observations manually either with paper and pen via log sheets or, preferably, using an electronic device.

External sources of data. The types of data that may enter the pilot plant's data system from external sources include costs of raw materials, raw material inventories, and costs of energy. Such data may be entered into the system manually, but ideally they would come in automatically in semi-real-time through RSS feeds or a related technology.

Data transformation

Transforming raw data into actionable information is more complex than Figure 3 implies. It is an intricate process that involves multiple steps *(3)*:

• collection — the communication and aggregation of the raw data from the multiple data sources at a central location

• storage — placing the aggregated data onto a hard drive or other storage device so that they can be retrieved later

• processing and manipulation — qualifying the raw data against validity and accuracy rules and, in some cases, flagging data as out-of-range or otherwise changing them so they can be presented properly in subsequent steps of the transformation process

• retrieval — obtaining the qualified data from storage so they can be presented

• presentation — displaying the qualified data in a format that gives them context, thus turning the data into actionable information.

Create your pilot plant data system

Developing a data system to collect, store, process and manipulate, retrieve, and present raw data from a pilot plant involves several steps.

Step 1. Identify every stakeholder that might require access to the pilot plant data system. Assign a single representative and an alternate (if possible) from each stakeholder set to champion the requirements of their set throughout the process. Meet and interview each representative to determine the key performance indicators (KPIs) and presentation style preferences of their group.

Step 2. Identify the instrumentation, equipment, and data sources that are necessary to calculate/derive the KPIs identified in Step 1.

Step 3. Generate a data requirements specification (DRS) that lists each stakeholder group's KPIs, as well as the instrumentation, equipment, and/or data sources that are required to calculate/derive them. This is not a detailed specification; it is a requirements specification. Detailed equipment and software needs are specified later in the process. The DRS should be reviewed and approved by all stakeholder representatives before the project moves to the next step. If requirements change after the DRS is approved, then the DRS should be revised and reapproved accordingly. It should be a living document that exists throughout the data system's life.

Step 4. Generate a detailed data system design specification (DDSDS) that specifies the instrumentation, equipment, communication infrastructure, hardware, and software necessary to meet the requirements outlined in the DRS. All stakeholder representatives should review and approve the DDSDS before moving to the next step. If requirements and/or specifications change after the DDSDS is approved, the DDSDS should be revised and reapproved accordingly. It, too, should be a living document that exists throughout the data system's life.

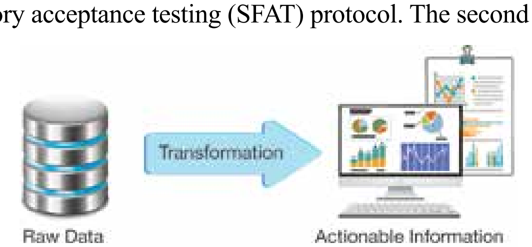
Often, the in-process or near-process instruments needed to calculate or derive KPIs are already being used for the pilot plant's process control system. Whenever possible, the same instrument should be used for both systems, unless some other requirement, such as safety, redundancy, or maintenance, warrants duplication. In situations where an instrument that is required for KPI purposes already exists for process control purposes, the DDSDS should flag it as such. These types of instruments will already be specified in documents generated for the process control system. The DDSDS should not respecify these instruments, but should instead provide a reference that points the reader to the process control system document where those specifications can be found. Only the instruments that are required solely for data system purposes should be specified in the DDSDS.

Step 5. Procure, implement, and deploy the pilot plant data system, adhering to all of the specifications contained in the DDSDS. Prototypes of every report generated by the data system should be provided to the appropriate stakeholder representatives for approval prior to final implementation.

Step 6. Generate testing protocols to ensure that the data system meets all of the requirements laid out in the DRS. If possible, the data system implementation should be staged, starting in a laboratory before it is installed at the pilot plant.

If you contract an SI consultant to deploy the data system, the SI firm will typically test the system at its factory, although you may decide to also test the system at your facility or another location as well.

When the system is tested at two locations, two separate testing protocols are needed. The first, referred to as the factory acceptance testing (FAT) protocol, is executed prior to installation at the pilot plant site, the SI's factory, or another agreed-upon location. The FAT protocol is sometimes divided into two separate protocols: the hardware factory acceptance testing (HFAT) protocol and the software factory acceptance testing (SFAT) protocol. The second test-



▲ **Figure 3.** Raw data must be transformed into actionable information.

ing protocol, referred to as the site acceptance testing (SAT) protocol, is executed once the system is deployed at the pilot plant. All testing protocols should be reviewed and approved by stakeholder representatives before they are executed.

Step 7. Conduct training for all of the stakeholder representatives on the operation, maintenance, and troubleshooting of the data system. Other training for the engineers, service, and IT personnel may also be required if they are going to maintain and make changes to the system in the future.

Identifying stakeholders

Before you can identify the requirements of the data system, you must first identify the stakeholder sets that will be involved with the pilot plant data system. This is typically done by asking the overall project manager which departments will need access to the data from the pilot plant and who within those departments can best determine their requirements. It is imperative for the right people to get on board as early on in the process as possible.

The first person you speak to may not be the right person. However, eventually you will find the person who has the information you need and the authority to speak for the entire stakeholder set. If at all possible, also try to find an alternate representative from each stakeholder set in case the primary representative is unavailable for extended periods of time.

Establish in writing the responsibilities of the stakeholder set representatives. At this point, some people may decline to serve as the representative for their group and will instead suggest another person.

Higher-level stakeholders (*e.g.*, CEO, executive vice president) will likely not be a representative, even though they will use the data generated from the pilot plant to make decisions. These higher-level managers, directors, and executives typically assign someone to the team who has the authority to speak for them and will report back to them directly.

Once the key stakeholder set representatives are identified, interview each of them to determine their group's requirements. Two key questions to ask in these interviews are:

- What KPIs does your group want to monitor?
- How does your group want those KPIs presented?

Bring examples of KPIs and different ways they can be presented to the interviews.

The output from the interviews with the stakeholder set representatives will be a list of KPIs, which will be documented in the DRS. This information will be used to determine the instrumentation and/or other equipment necessary to calculate or derive each KPI.

Dealing with divergent data sources

Each type of data must be properly interfaced with the data system's central data repository.

In-process data interface. In- or near-process instruments are typically wired in to the process control system (*e.g.*, programmable logic controller [PLC], distributed control system [DCS]) that monitors and controls the pilot plant process. Data obtained from these instruments will be communicated to the central data repository via the control system. Therefore, the central data repository interface should support the same communication protocol as that supported by the process control system. Most commercially available central data repository systems support many different types of process controller interfaces, so this will most likely not be a problem.

Since most of the data collected from a pilot plant will likely be in- or near-process data, it may be valuable to build fault tolerance into the interface between in-process data and the central repository. If the interface fails for some reason, the data would be sent locally to the process control system side of the interface until the interface function is restored. Another option is to install fully redundant interfaces, so that a backup interface would kick in if the primary interface were to fail. Often, fault-tolerant and/or redundant interfaces at the pilot plant level may be overkill and cost-prohibitive, but it is an option to consider. Most commercially available central data repository systems support fault tolerant and/or redundant interfaces.

Laboratory equipment interface. Unlike in-process data, which are continually transferred to the central data repository, data from lab equipment will most likely consist of batches of data obtained manually by analyzing samples with lab equipment. And, although some lab equipment may support one or more of the same communication protocols that the central data repository supports, that is typically not the case.

There are two common ways to transfer lab equipment data to the central data repository. One method involves setting up the lab equipment to copy data, in the form of comma-separated-variable (CSV) files or another type of delineated-data file, to a specific location on a network drive when the analysis is complete. When the central data repository sees a file at this location, it opens the file, parses the data, and inserts the parsed data into the repository. This function is not available in all central data repository systems or laboratory equipment.

Another option involves purchasing a laboratory information management system (LIMS), which gathers data from the lab equipment and sends it to the central data repository. Not all commercially available central data repository systems support interfaces to a LIMS.

Manually observed and entered data interface. Several factors must be considered when designing a system that supports manually entered data.

The central data repository must support manually entered data and provide tools to create online forms that the operators can use to enter data. Manually entered data are prone to errors, so the forms used to collect the data should make extensive use of drop-down lists and other graphical user interface (GUI) objects that minimize the potential for data entry errors. Limits and other error-checking methods should be employed to further minimize the chance of data entry errors. However, regardless of these precautions, data entry errors will occur. Thus, a supervisor should be required to review all manually entered data before they are transferred to the central data repository.

In addition, if operators will need to regularly tour the plant and enter observations as they make their rounds, the central data repository must be compatible with portable electronic devices.

Some central data repository systems support software that can be used to create online forms that run on an operator's tablet and essentially walk them through their rounds. The forms prompt the operator to confirm that they have "visited" important locations during their rounds. It is highly recommended to place labels (*e.g.*, bar codes, radio-frequency identification [RFID]) at these locations and connect scanners that can read these labels to the operator's tablet to ensure that the operator actually visits each location.

When operators are required to take samples, the tour forms are a good way to capture the location from which the sample was taken.

External source data interface. If the data system involves data from external sources, the central data repository system must be able to communicate with those sources. RSS feeds are a typical mechanism to bring in data from external sources.

The most important factor to consider when opening your data system to outside sources is security. A well-defined and implemented cyber security plan must be in place to ensure that the system is not exposed to outside threats.

Central data repository selection criteria

The central data repository is where all divergent data sources send their data. The selection of the right central data repository is perhaps the most important decision during the design of a pilot plant data system.

Several companies offer central data repository systems. When choosing one, it is important to remember that the input side of the data warehouse is just as important as the output side. Do not be lured by fancy graphics and reports that the vendor pitches during their sales call. What is important is that the system can communicate with all of your data sources when you take it out of the box, without the need to write custom communication drivers.

Next, you must realize that, even in small pilot plants, large amounts of raw data can be generated and accumulate very quickly. The repository selected should support both exception-based and polling-based data source communication. It should also incorporate sophisticated data-compression algorithms.

To emphasize the importance of selecting the right central data repository, consider the pyramid in Figure 1. Data are at the bottom of the pyramid and thus form its foundation. If the data (*i.e.*, your data repository) are weak, the pyramid will crumble and come crashing down — just like your pilot plant's data system will if you fail to pick the right central data repository for your application.

Final thoughts

The pilot plant is the bridge between the laboratory and the commercial production plant, and its data system will operate long after the physical construction of the pilot plant is complete. It is the tool that senior management will use to determine whether it is worthwhile to stay the course and build a full-scale production facility or to abandon the process altogether and move on to other endeavors.

The best-designed and implemented pilot plant will ultimately fail if its data system does not deliver the information to the decision-makers who need it, when they need it, and in the format that they need it. CEP

Literature Cited

1. **Ackoff, R.,** "From Data to Wisdom," *Journal of Applied Systems Analysis*, **16** (1), pp. 3–9 (Jan. 1989).
2. **Bellinger, G., *et al.*,** "Systems Thinking: Data, Information, Knowledge, and Wisdom," www.systems-thinking.org/dikw/dikw.htm (2004).
3. **Rudo, P.,** "Six Important Stages in the Data Processing Cycle," *Enterprise Features*, www.enterprisefeatures.com/6-important-stages-in-the-data-processing-cycle/ (Apr. 24, 2013).

Additional Resources

Fisher, M., "Collecting, Organizing, and Presenting Data from Small Scale Systems," presented at the 2015 AIChE Spring Meeting and 11th Global Congress on Process Safety, Austin, TX (Apr. 15, 2015).

Meyer, C., "Four Steps to Turn Big Data into Action," *Forbes*, www.forbes.com/sites/mckinsey/2013/10/22/four-steps-to-turn-big-data-into-action/#6a10c2f554c0 (Oct. 22, 2013).

MARK FISHER is a senior director at Thermo Systems, LLC, a systems integrator headquartered in East Windsor, NJ (Website: www.thermosystems.com; Email: mark.fisher@thermosystems.com). With almost 30 years of industry experience, he has designed and deployed mission-critical automation and information systems in several industry segments, including life sciences, food and beverage, consumer products, water/wastewater treatment, mining, chemical processing, chilled-water generation and distribution, packaging/sortation, data center utilities, and district energy. He has authored several publications, including "Controls, Instrumentation, and Metering," Chapter 8 of the International District Energy Association (IDEA) *District Cooling Best Practice Guide*. He has an AAS in electronics engineering technology from Mercer County Community College and a BS in electrical engineering from the College of New Jersey. He is an ISA-Certified Automation Professional (ISA-CAP) and an ISA-Certified Cyber Security Specialist (ISA-CCSS — ISA99/IEC62443).