TA GUIDEBOOK

OSIsoft Academic Hub

OSIsoft Academic Hub is a cloud-based platform that enables experiential learning with industrial process datasets. Students use OSIsoft's Academic Hub to gain experience solving real-world problems with real-world data, while researchers use the Academic Hub for easy data sharing among institutions. This document is designed to provide TA's step by step instructions to prepare classroom and lab students to use the Academic Hub.

OSIsoft and University of Southern California

Advancing university educational programs for scientists and engineers has been a professional priority of mine throughout my 45+ year career in the oil and gas industry. A particular focus has been on how best to incorporate the ongoing growth in digital technologies into university curricula and research programs to prepare graduates for future careers in the industry. As the former Corporate Vice-President and Chief Technology Officer of the Chevron Corporation, I pursued a series of joint venture and cooperative arrangements with both universities and the technology industry to enable a growing intersection of digital technology with the traditional fields of petroleum engineering and geoscience in both industry and academia. In my second career at the University of Southern California, my colleagues and I continue to put into practice this concept and evolve graduate-level engineering education to incorporate the rapidly advancing areas of data science and machine learning.

OSISoft is a core technology supplier to the oil and gas industry through the wide-scale deployment of PI systems and OSIsoft Cloud Services (OCS) for real-time data acquisition, analysis, and management in field and plant operations. OSISoft has also developed an extensive university partnership program to enable academic programs to directly expose STEM students to the PI system and OCS, and help students develop a foundation in data-driven technology. When I was introduced to OSIsoft's new Academic Hub program, which incorporated the PI System into the modern cloud-based digital architecture of OCS, I saw the opportunity to create a much more flexible and dynamic learning and research environment. With the OSIsoft Academic Hub, universities can develop and expand educational and research tools, utilize the scale and ease of a cloud environment, and naturally collaborate with other universities. Universities contributing datasets and curriculum to the OSIsoft Academic Hub is key to growing the value of this collaboration space and preparing engineering students for the impact of data analytics across all industries over the course of their careers.

As long-time academic partners with OSISoft, the University of Southern California is excited to support the advancement of the OSIsoft Academic Hub, and to participate in the development of the learning guide detailed in this document. We anticipate that this guide will enable faculty, staff, and teaching assistants to efficiently utilize the OSIsoft Academic Hub's data and applications to expand their educational and research opportunities along with partner universities.

Dr. Donald L. Paul, PhD

Research Professor of Engineering and William M. Keck Professor of Energy Resources and Executive Director of the Energy Institute at University of Southern California

Table of Contents

Table of Contents

Industry and Academic Challenges with Data

Digital Transformation is underway across virtually every industry. Even with current economic headwinds, companies are developing and implementing plans to expand the utilization of digital data and advance digital technologies to lower costs, improve operating performance, and re-engineer workflows and decision processes.

However, the implementation of Digital Transformation requires companies to face two major challenges: transition their digital infrastructures to modern architectural paradigms (such as edge-cloud and IIoT), and build the organizational capabilities required to capture value from the exponential increase in both real-time data acquisition and advanced analytics. Meeting this second challenge will require an expanded form of interaction between industry and academia around data science, to both evolve the data analysis skills of their current workforce and to build additional data analysis fluency into the education of new engineering graduates.

University engineering schools are rapidly expanding both research and educational programs to incorporate elements of data science, informatics, and computer science to compliment core engineering content across disciplines and increase data and digital technology intersections. Artificial intelligence and machine learning are creating new opportunities to improve industrial operations and decision processes through automation and control system advances. However, universities will need the tools, systems, and data to provide the hands-on experiences necessary to both develop current engineering students and support current industry workforce improvement. In particular, access to industrial data for both research and education is a critical element.

Universities have historically faced twin challenges when it comes to utilizing industrial data for research and education: access to real world datasets and software. Availability of industrial-grade datasets of sufficient scale and realism is limited, and this is particularly true when it comes to real-time operational datasets. In addition, there has been a lack of convenient access to industry-standard software and training to utilize these systems. An additional component of the second challenge is the difficulty some academic environments face in implementing and supporting the necessary digital and data management infrastructures (including effective cyber-security protocols), at the level normally seen in industrial environments. This latter challenge will likely grow as data volumes rapidly expand and closer industry-academia ties lead to demands for high-performance connectivity.

Bridging the Academic and Industry Gap

In partnership with industry and academia, OSIsoft created the OSIsoft Academic Hub to address the principle challenges outlined above. The OSIsoft Academic Hub will develop and deploy a data sharing platform and user environment for both industry and academia to contribute datasets relevant to multiple industrial sectors and enable access and sharing for both educational and academic research programs.

The key elements of the OSIsoft Academic Hub include

- Development on OSIsoft Cloud Services (OCS), which means implementation and maintenance of a stand-alone system by faculty or TA's is no longer required
- Data transfer mechanisms, management structures, and protocols enabling industry and academia to contribute datasets for shared use in research and education are already created and ready for student, TA, and faculty use
- Applications and user toolkits specifically developed for university students to have access and practical experience with industrial datasets and the PI system
- Robust security protocols to ensure data integrity, system security, and user confidentiality to ensure cybersecurity and privacy for users and the institution

With university partners, OSIsoft will also develop and deploy an Academic Hub Teaching Assistant Guidebook containing step by step instructions for setting up student account registration, accessing existing datasets and curriculum, and preparing new datasets for sharing in the Academic Hub. Students will gain access to data and applications which support academic research and education programs.

By utilizing the OCS model, the OSIsoft Academic Hub structure removes the access and support barriers for universities previously encountered to hosting their own PI systems. This should materially expand the pool of potential university partners. For current OSIsoft Academic Partners, the OSIsoft Academic Hub will also support configurations which integrate both traditional PI Systems and the cloud based OSIsoft Academic Hub.

The OSIsoft Academic Hub also anticipates future capabilities to allow universities to develop and contribute new real time, time series applications coming out of ongoing research and educational programs. Future capabilities will also accommodate industry-academia structures which include both shared and proprietary datasets.

Operational and Time Series Data and Its Uses

Operational data is data that reflects the physical state of an industrial asset (or "thing"). There are many sources and types of data in operations; real-time process control data, sensor, and device data; machine-generated data, asset data, web data, weather data, heat map data, and many more. One of the unique types of operational data is time-series data, where time is the index.

Relational databases are comprised of entities with attributes, such as a list of homes on a street (entities) and their respective attributes (number of bedrooms, square footage, year built). Time series data is not comprised of entities, but instead is a series of values, each paired with a timestamp. These values can be defined by names and a set of attributes (units of measure, scan rate, digital state). Time is the major axis, and the gaps between data received can be routine or erratic. Time series data is particularly useful for evaluating change, whether it be in the past, present, or predicting how something might be different in the future.

OSIsoft helps deliver value using real-time operational time series data across multiple use cases from asset health optimization, to saving energy, process optimization, tracking quality, ensuring regulatory compliance, and improving safety. OSIsoft provides the industry standard in enterprise infrastructure, for management of real-time data and events. For over forty years OSIsoft has been at the forefront of time series data collection and aggregation, along with other types of operational data, allowing people to contextualize and federate large volumes of industrial, operational data.

Asset Health	Energy Efficiency	Process Optimization	Quality Tracking	Regulatory Complianc	Safety
DCP Midstream	**Air Liquide**	**ArcelorMittal**	**Deschutes**	**TasWater**	**Qatar Power**
$20-25 million cost savings in first year ↗	10x ROI from operational savings in first 8 months ↗	Shipped additional 26M tons for $120 million in added revenue ↗	Postponed $8 million capital upgrade ↗	Reduced response time by 13 hours ↗	Over 3,452 days without lost-time accidents ↗

OSIsoft: Company Overview

OSIsoft was founded in 1980. The company's headquarters are in San Leandro, California. Our company is a global operation with over 1,400 employees. The company is privately held and has always been profitable. With installations in 140 countries spanning the globe, OSIsoft software is used in manufacturing, energy, utilities, life sciences, data centers, facilities, and process industries.

Oil & Gas	Power & Utilities	Metals & Mining	Pharma & Life Science	Chemicals	Pulp & Paper
80% of the top companies globally	**1000+** utilities worldwide	**8** of the global Fortune top 10 companies	**24** of the top 25 pharmaceutical companies	**9** of the top 10 chemical companies rely on the PI System	**138.5M** meters per year of production monitored

OSIsoft is led by engineers and industry experts. We have a large network of partners, including leading technology companies, such as Microsoft and ESRI. With 65% of industrial Fortune 500 customers using OSIsoft products to monitor over two billion data streams globally, it is likely engineers will work with an OSIsoft product during their career. Data analysis and specifically time series data analysis skills are quickly becoming a standard requirement for engineers and managers alike. OSIsoft data is ready to integrate into sophisticated visualization tools, industry applications and AI/ML platforms in use today globally.

As a system of record, OSIsoft provides the foundational data that enables better decisions, improves business processes, and fuels digital transformation. OSIsoft is the backbone of industrial operational data flow. Hundreds of different types of instrumentation and systems use OSIsoft to enable real-time analytics, create predictive statistical modeling and machine learning, and visualize it all using dashboards, assessments, and diagnostics.

"Data is the only resource that gets more valuable the more people use it."

Dr. Pat Kennedy, OSIsoft Founder and CEO

OSIsoft Academic Program and Academic Hub

OSIsoft Academic Program

The mission of the Academic Program at OSIsoft is empowering the workforce of tomorrow with data-focused skills that industry needs. OSIsoft wants to create a community of educators, students, and industry contributors that enables sharing of industrial datasets, best practices, and educational material. A key component of strategy is the OSIsoft Academic Hub, a cloud-based platform that enables experiential learning with industrial process datasets. Students use OSIsoft's Academic Hub to gain experience solving real-world problems with real-world data, while researchers use the Academic Hub for easy data sharing among institutions. Smart campus initiatives utilize OSIsoft technologies to transform campuses into living labs, providing students with real-world operations data and opportunities to apply their skills to sustainability efforts that better their campuses and communities. Leading universities currently partnering with OSIsoft include University of Southern California, Massachusetts Institute of Technology, McMaster University, North Carolina State University, University of Oxford, Rose-Hulman institute of Technology, University of California Davis, Lehigh University, and Carnegie Melon University.

OSIsoft Academic Hub

OSIsoft Academic Hub is an online service offering for engineering education designed to empower educators to equip their students with process data literacy. OSIsoft hosts and provides students with access to real, industrial datasets and related learning modules. Educators also work with industry to develop process data analytics educational materials based on challenges and representative datasets from industrial assets and processes. Students gain experience in several areas that help to prepare them for careers in today's workplace such as navigating broader, less defined problem statements and analyzing large datasets to improve business decisions. Students also get to practice integrating timeseries data with other types of data using industry current AI/ML platforms, visualization tools, and programming languages including MATLAB, Jupyter, R, and Python. Leading universities are using the OSIsoft Academic Hub to improve learning outcomes in laboratory-based engineering courses such as Chemical Engineering Unit Operations and Controls and interdisciplinary courses focused on sensor data acquisition, processing, and analysis.

Another key use of OSIsoft's Academic Hub for the data education community is as an ecosystem for collaboration. Researchers use OSIsoft technologies to enable strong partnerships with industry and accelerate industry innovation. OSIsoft allows for easy data sharing among institutions and consistent methods of managing and enhancing data; this enables the adoption of industry standards in research and accelerates the idea-to-market process. OSIsoft and universities partner across a wide variety of industries including Microgrid & PMUs, Pharmaceutical Operations, Distributed Energy Resources, Food and Beverage, and Smart Campus and Cities.

Software

Web Browser

The OSIsoft Cloud Services portal and the OSIsoft Academic Hub recommend using one of the following desktop browsers. If you are using an unsupported browser or version, please download or upgrade to a supported browser or version from any of the sites listed below. If you choose not to upgrade your browser, you may not be able to use certain features within the OSIsoft Cloud Services portal and the OSIsoft Academic Hub. You also need to enable JavaScript.

Browser	Version	Website
Microsoft Edge	Latest Version	https://www.microsoft.com/en-us/windows/microsoft-edge
Mozilla Firefox	Latest Version	https://www.mozilla.org/en-US/firefox/
Google Chrome	Latest Version	https://www.google.com/chrome/browser/desktop/index.html

Anaconda, Jupyter Notebooks, Python, and Pandas

The OSIsoft Academic Hub curriculum material is in the form of Jupyter Notebooks to house instructions, texts, exercises, and solutions. All learning modules can be executed with a standard Jupyter server using Python distribution such as Anaconda. Anaconda is highly regarded and used internally (Version 3.6 and later) to develop and test the notebooks against OSIsoft Academic Hub datasets in OCS. Pandas and NumPy are among standard numerical Python packages the notebooks rely on for data manipulations.

Pandas is the de-facto Python data science library for tabular data manipulation and analysis. Before Pandas, Python's role was confined to data munging and preparation. With Pandas, it is possible to perform data analysis and modeling fully and simply in Python. Since process data obtained from OCS with Data Views is in tabular format, the Pandas library was the natural choice for the learning modules to introduce students to this important tool.

Anaconda: https://www.anaconda.com/products/individual

Jupyter: https://jupyterlab.readthedocs.io/en/stable/getting_started/installation.html

Pandas: https://pandas.pydata.org/

Python: https://www.python.org/downloads/release/python-385/

University Requirements

Licensing and Agreements

In order to legally access OSIsoft's software, both individuals and the university must obtain and adhere to licensing contracts and usage agreements. OSIsoft is committed to protecting the privacy of student information, and has a formal OSIsoft Student Information Privacy policy available for public review.

University:

- An Educational Software Licensing Agreement needs to be in place, ideally at a university-wide level, to enable data and curriculum sharing for all departments

- OSIsoft Academic Hub Agreement needs to be in place, which covers the university departments for the courses and labs that will use the Academic Hub

Individual:

- Each individual user including students, TA's, research staff, and faculty, needs to create an OSIsoft Academic Hub Account Registration

- Each individual must agree to the Academic Hub Subscription Agreement terms of service, and acceptance of the Academic Hub Terms of Service is signified by checking a box during the Academic Hub Account Registration process.

Security and Access

- Access to the academic hub should be limited to related Faculty Members, TAs, RAs, Course Producers (CPs) and enrolled students.

- The security of the account is dependent on the University's Identity Provider
    - Additional security can be provided by Two-Factor Authentication, Duo

Additional Considerations

There are different requirements for simply accessing existing datasets on the Academic Hub, versus contributing dataset to the Academic Hub. Datasets that can be publicly disclosed can be uploaded onto the hub freely. The university must have permission to publish datasets it wishes to contribute, and the legal burden of verifying permission falls to the university. Protected datasets, such as proprietary work with commercial partners, should only be accessible by users with a signed NDA. The university should be diligent to limit access to each dataset based on users.

There are also machine and bandwidth requirements for university equipment used to send data to the Academic Hub, depending on the volume and frequency of any streaming data to be shared.

OCS Account Creation for Academic Hub

Identity Provider

Access to OSIsoft Cloud Services (OCS) resources requires user authentication. OCS uses industry standard OAuth 2.0 and OpenID Connect for this part. In the case of the Academic Hub, Google is the only Identity Provider (IdP) supported. Since many schools rely on Google G Suite (https://edu.google.com/), students may have already own a Google account they can reuse with OCS once they register and are invited.

The TA needs to verify whom the Identity Provider is for their campus.

There are two possible scenarios: either Google is the university's Identity Provider, or an entity other than Google is the university's Identity Provider. The next steps for each of these situations are below.

Google is University's Identity Service Provider

University email accounts can be used for OCS account creation.

Something Other Than Googles is University's Identity Service Provider

Students will need to create Gmail accounts to use for OCS account creation, and OSIsoft and the TA will need to keep correlated list of Gmail and university email accounts.

Workflow for OCS Account Creation and Vetting

There are two main subprocesses of the OCS account creation workflow: collecting the list of students enrolled in each course, and OSIsoft actually creating the accounts.

For the first piece, the TA or faculty member can provide OSIsoft a list of names and email addresses for students enrolled. Alternatively, the following process will be used.

- TA shares OCS Account registration link with students enrolled in course
  - o TA posts the registration link on the course Blackboard site and gives students a deadline to register. The TA should check the box to send email notification for the post to make sure students receive the message.
  - o If the university does not have Blackboard, another Learning Management System could be used to contact the students or even simply email.
- OSIsoft's OCS Account Registration form asks for user's full name, university and Gmail email addresses, class name, class code, and click box Accepting OSIsoft Academic Hub Subscription Agreement.
- OSIsoft will provide list of students who have registered to TA and faculty member for verification of course enrollment before OSIsoft creates the accounts.
- TA and faculty member will verify students on list are enrolled in the course and notify OSIsoft via email.

Once OSIsoft has the verified student list from TA and faculty member, OSIsoft creates accounts for class.

- OSIsoft will send an email to each student acknowledging OCS Account creation
  - The email will come from cloudservices@osisoft.com
  - Each student then needs click on the acceptance link, boxed in red below
  - Please note – every invitation expires in 21 days!

[EXTERNAL] Accept your invitation for OSIsoft Cloud Services account for OSIsoft Academia

OC   OSIsoft Cloud Services <cloudservices@osisoft.com>
       To   Jenny Danielsen

Caution: This email came from outside the company.

Dear Jenny,

Welcome to OSIsoft Cloud Services!   **Click to Accept**

To accept your invitation:
1. Click: https://dat-b.osisoft.com/identity/invitation/process?id=gXcV7L-92TptweLequrz3eic2_kX9HIvVn84kJftjzc
2. Sign in using Azure Active Directory

After accepting your invitation:
1. Access OSIsoft Cloud Services at: https://cloud.osisoft.com/
2. Click Sign in and enter the company alias: **osisoft-academia**.

Please refer to the introduction page of the OSIsoft Cloud Services documentation for instructions to get started.

If you have questions, please don't hesitate to contact us at cloudservices@osisoft.com.

This invitation will **expire** in 21 days on Friday, May 8, 2020 at 8:08 PM (*Coordinated Universal Time). Please click here to convert this to your local time zone.

- OSIsoft then provides list showing OCS Account creation status and student acceptance status, allowing TA to follow up with any students without OCS Accounts.

- After initial set of OCS Accounts, if additional students need to be added to the course, TA will need to work with the Academic Account Manager to have additional OCS Accounts created.

- Student OCS Accounts will be deactivated at conclusion of course enrollment.

Timeline Expectations

Currently, there are manual steps in the OCS account creation process. At least three business days need to be allotted for the steps of OSIsoft collecting and then sharing the student registration lists, and also OSIsoft creating the OCS accounts.

OSIsoft Academic Hub Background

OSIsoft Cloud Services

The OSIsoft Academic Hub is powered by OSIsoft Cloud Services (OCS). OCS is a highly flexible cloud-based platform that provides scalable, elastic, centralized environment to aggregate data for reporting, advanced analytics, and third-party applications. In OCS, data is organized first into Tenants which represent an organization that owns a collection of OSIsoft Cloud Services. Tenants are divided into one or more units called Namespaces. Namespaces serve as the destination for incoming stream data. Each Namespace is distinct and separate from other Namespaces to prevent accidental or malicious data access.

Authentication

Each individual user will register for an OCS Account, providing their name, email, and course title and class code (provided to students by TA). These criteria will be used for provisioning of accounts and associating accounts with all course specific data. Students will have an OCS account that allows them to access the OSIsoft Academic Hub Tenant and Namespace, along with course specific data. The authentication process will vary based on the method used to access data, such as Python or the data portal. OSIsoft provides an OCS RESTful API client library for Python that helps with authentication, discovery of available data views, and data access. More detail is provided below in the Python section on the next page.

Data Views

OCS uses data views to store and map data from a particular asset. Data views have a tabular form, where rows are observations (times) and columns are data streams (parameters such as temperature). The user selects the desired data view, start and end time for the data view, and the interpolation interval.

Here is an example of a data view from a fermenter showing the volume, temperatures, and apparent degree of fermentation at five different times.

	Timestamp	Volume	Bottom TIC OUT	Bottom TIC PV	Middle TIC OUT	Middle TIC PV	Top TIC OUT	Top TIC PV	ADF	Fermentation ID
0	2017-03-17 07:00:00	716.56600	10.935327	29.884571	0.0	29.356380	0.0	29.613152	0.659725	Fermentor 31201731179653
1	2017-03-17 07:15:00	716.56600	41.584870	30.140123	0.0	29.487812	0.0	29.451567	0.659725	Fermentor 31201731179653
2	2017-03-17 07:30:00	716.56600	19.585953	30.120740	0.0	29.532250	0.0	29.469800	0.659725	Fermentor 31201731179653
3	2017-03-17 07:45:00	716.56600	85.105540	30.309180	0.0	29.579490	0.0	29.475550	0.659725	Fermentor 31201731179653
4	2017-03-17 08:00:00	716.56600	72.413750	30.334532	0.0	29.509760	0.0	29.769472	0.659725	Fermentor 31201731179653

OSIsoft Academic Hub Data Access Methods

Python

OSIsoft provides an OCS RESTful API client library for Python that helps with authentication, discovery of available data views, and data access. OSIsoft recommends a standard Jupyter server using Python distribution such as Anaconda. Anaconda is highly regarded and used internally to develop and test Jupyter notebooks against OSIsoft Academic Hub datasets in OCS. Pandas and NumPy are among the standard numerical Python packages notebooks rely on for data manipulation. Pandas is the de-facto Python data science library for tabular data manipulation and analysis. Since process data obtained from OCS with data views is in tabular format, the Pandas library is practical.

Here is a summary of the functionality offered by the Academic Hub Python Library:

- OCS:
    - Initialize client object
    - Authenticate using Hub OCS account
- Datasets:
    - Get the list of all published datasets
    - Select/get current dataset and associated OCS namespace
    - Refresh datasets information from Hub
- Assets (of current dataset):
    - Get list of assets with their Asset ID and description
    - Get list of data views IDs associated to an Asset ID
    - Get metadata (static information) about an asset
- Data Views:
    - See the structure of a data view: Asset_ID, column names, unit of measure, stream names
    - Get a Pandas dataframe (one row == one observation) with interpolated data from a given data view

For more details please consult the latest Hub Library QuickStart notebook on Github

OSIsoft Academic Hub Data Portal

OSIsoft provides another method for data access, but it is much less robust than Python. The Academic Hub Data Portal allows users to download CSV files, but each download is limited to a single asset for a single time range.

OSIsoft Academic Hub Data Portal

Community Datasets

Collaboration and Existing Datasets

A key use of OSIsoft's Academic Hub for the process data education community is as an ecosystem for collaboration, and the Hub currently houses two community datasets.

Deschutes Brewery provided a dataset containing process parameters such as temperatures, pressures, and volumes as well as brands of the brews used in 29 fermenters for the years 2017-2020. UC Davis Facilities provided a dataset containing meter data parameters about electricity, heating, cooling, gas and building square footage of 44 buildings spanning 2017 - 2019. OSIsoft wants universities and industry to share authentic industrial process datasets for use in the classroom and for research purposes, in order to increase the number of available industrial datasets. Access to authentic industrial datasets of sufficient scale and realism is limited, and this is particularly true when it comes to real-time operational datasets. The more universities that share process datasets along with process data analytics educational materials, the more students will have experiential learning opportunities.

Existing Data Views

Links to full documentation for the currently available community datasets:

[Deschutes Brewery Dataset Documentation](#)

[UC Davis Facilities Dataset Documentation](#)

[All Dataset Documentation](#)

Contributing Data to the OSIsoft Academic Hub

When a university has a new dataset they would like to contribute to the OSIsoft Academic Hub, the following steps need to occur:

1. Assure Dataset Availability in Local PI System
2. Develop Asset Model for Dataset
3. Document the Dataset – Description and Metadata
4. Transfer Dataset to Academic Hub
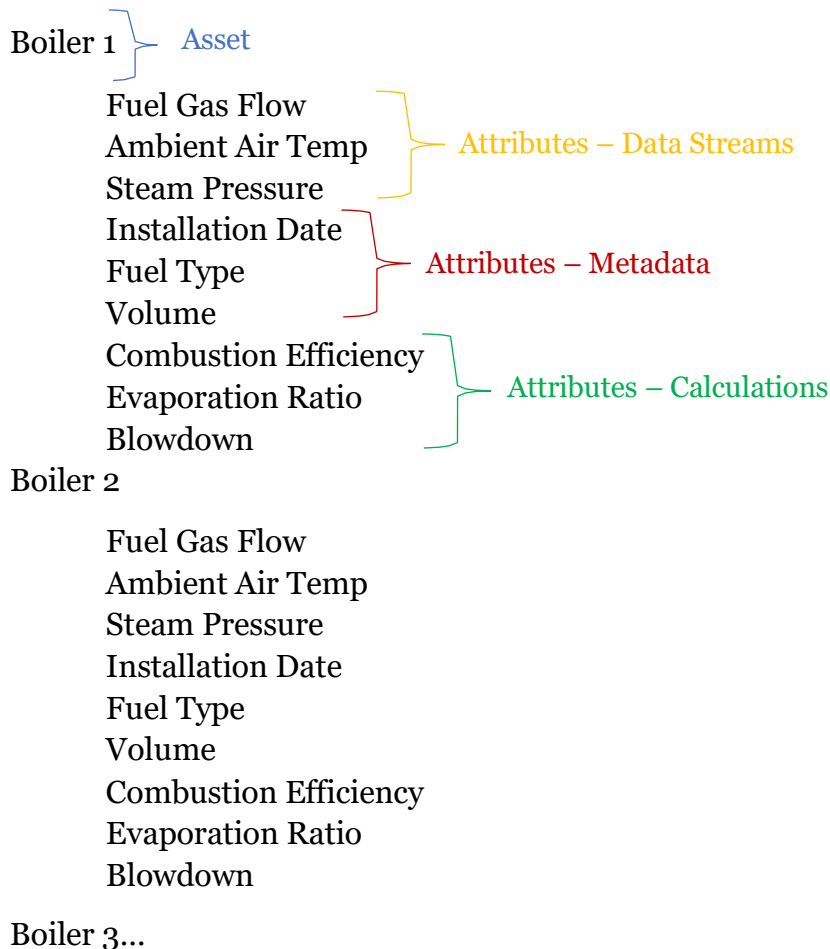5. Create New Data Views

Dataset Availability on University's PI System

The new dataset needs to exist on the university's PI System, and if it is not already there the TA needs handle this process. The two likely categories for data transfer are from lab systems (historical or live), and the second is a public or preexisting dataset. In the case where the university is collecting its own data from its own devices to put into the university's PI System, the TA will need to determine and install the correct interfaces and connectors. This process is covered in the Configuring a Simple PI System training course, linked below at the end of this section under Online Curriculum. In the case where the university wants to contribute a public or preexisting dataset, the TA will need to import the external dataset and manually organize it in the local PI System before transferring the dataset to the Academic Hub. The OSIsoft Live Library has information on the multitude of interfaces and methods available to import data into the PI System such as the Universal File Loader and over three hundred other interfaces. The PI Software Development Kit, (PI SDK) provides COM access to OSIsoft historians. PI SDK provides an object-oriented approach to interacting with PI Systems in contrast to the procedural methods used in the PI Application Programming Interface (PI API).

Develop Asset Model for Dataset

Most operational data can be thought of and organized in terms of assets. Assets can be large stationary equipment such as turbines, or containers like tanks, or even mobile resources like trucks and train cars. Each piece of information is stored in an attribute, and attributes are defined for each asset. Attributes include data streams, calculations, and metadata. Data streams include temperatures, pressures, and flows. Calculations can be anything from simple time weighted averages to soft sensors such as complex chemical calculations. Metadata is the collective name for all the various information about an asset, such as manufacturer, model, and serial number. A process status data stream is extremely helpful to understanding and accurately modeling the process. Process status examples include heating, cooling, separating, drilling, blending, compressing, and fermenting. Process status can be defined by using logical and mathematical equations as calculations in an attribute.

```
PI AF Elements
        Boiler 1 —⌐ Asset
                Fuel Gas Flow
                Ambient Air Temp   —⌐ Attributes – Data Streams
                Steam Pressure
                Installation Date
                Fuel Type          —⌐ Attributes – Metadata
                Volume
                Combustion Efficiency
                Evaporation Ratio   —⌐ Attributes – Calculations
                Blowdown
        Boiler 2

                Fuel Gas Flow
                Ambient Air Temp
                Steam Pressure
                Installation Date
                Fuel Type
                Volume
                Combustion Efficiency
                Evaporation Ratio
                Blowdown
        Boiler 3...
```

The TA will use PI AF to create a digital representation ("digital twin") of the asset's data structure. This digital twin will be a virtual model of the asset, providing a detailed, intricate view of a physical asset that is likely not physically in the same place.

1) Create AF elements that represent each physical asset
    a. Add attributes for PI data streams like temperatures, pressures, and flows
    b. Add attributes for metadata like manufacturer, model, type, and capacity
    c. Add attributes for real-time calculations and soft sensors
        i. Create calculations that write data to PI tags, which are also referenced as attributes of the respective AF element
        ii. Soft sensors are also frequently used in PI calculations, for example engine performance, emissions, and weighted averages
2) Document the digital representation
    a. Describe the elements, assets, and attributes
    b. Provide a graphical explanation of the AF structure

Dataset Documentation Template

Part 1: Description of Dataset

Explain the process occurring over time throughout the dataset. Ideally, there will be a status tag that correlates with the stages of the process. This does not need to be extensive, a sentence or two for each stage is adequate. Students and others without relevant industry experience may not know what others assume to be industry common knowledge. Include the following type of information:

Facility Type	Industry	Equipment Types
Process Definition	Product Type	Time Range of Data
Assets	Batch Numbers	Process Status Stages
General Data Types	Categorical Data	Acronyms or Abbreviations

Part 2: Industry Challenges

Discuss the industry challenges of the entity from which the dataset comes from.

- What are the biggest pain points, costs, and risks being faced?
- Which processes and data included in the dataset impact each of these challenges?
- What correlations and causations are being studied?
- What are common goals in this industry – minimizing or maximizing certain parameters?

Part 3: Stream Data

Include a table of the stream data. This should include columns for index, Stream Configuration, Stream Name, Unit of Measurement, and Description.

Part 4: Data Quality

When the entire dataset is viewed as one entity, discuss any known data quality issues.

- Are there any large gaps that can be pointed out and explained ahead of time?
- Were some of the assets added or deleted at some point?
- Do some assets have substantially more instrumentation than others?
- Was the facility shut down for several weeks for a turn around?

Dataset Documentation – Description and Metadata

Example of Dataset Documentation for the Academic Hub:
Deschutes Brewing dataset documentation

Transfer Dataset to Academic Hub

In order to transfer data from a university's local PI System up into the OSIsoft Academic Hub, a publication must be set up using PI Cloud Connect. PI Cloud Connect (PICC) is a cloud-based Software as a Service (SaaS) offering that enables data sharing between PI Systems. PICC leverages components running in Windows Azure, the public cloud offering from Microsoft and does not require virtual private networks (VPN's).

PI Cloud Connect uses a publish and subscribe mechanism to manage the data flow between PI Systems. TA's will create an account on PI Cloud Services, and then sign into the PI Cloud Connect customer portal to install the components required to connect their university PI System securely and reliably to OSIsoft's Academic Hub and share data. TA's will use the PI Cloud Connect customer portal to manage publications, subscriptions, users, and nodes. OSIsoft's Live Library – PI Cloud Connect.

> **Publication**: an automated transfer of data over the cloud from a PI System to a node. (University will publish their dataset)
>
> **Subscription**: contract that enables a specific user to receive a read-only version of the data associated with the publication. (University's subscription consists of their dataset)
>
> **User**: an account setup to provide a specific user to have access to create publications and add subscribers to publications. (University will be a user)
>
> **Nodes**: computers that have had the PI Cloud Connect setup kit installed on them and can receive data over the cloud from the connected PI System. (OSIsoft Academic Hub will be the node)

Since PI Cloud Connect is transferring data from a PI AF structure to another, it is expected that the PI AF configuration is identical at both the source and the destination. The PI Cloud Connect Customer Portal does not provide functionality to create or organize structures in PI AF. This is why developing the dataset's asset model is a previous step in the process of contributing a dataset to the Academic Hub.

Data Sharing Workflow

As a destination for data sharing, users can create a subscription associated with publications that they have been invited to access. The association between a publication and a subscription is a contract between the publisher and subscriber that specifies what data is being shared. When the configuration of the publication and the associated subscriptions is complete on both sides and the publication and the subscription are started, the exchange of data commences and continues until either the publication or subscription is stopped.

Download Setup Kit → **Create Publication** → **Add Subscribers**

1	Download the PI Cloud Connect setup kit: The first step in sharing data is to download the PI Cloud Connect Setup Kit. The setup kit can be downloaded from the System page. See the System topic for more information.
2	Create the publication: The second step is to select a data source to include in the publication. A publication is configured by selecting a PI AF Element from any PI AF server that is accessible from a registered PI Connect node. See the Create publications topic for more information.
3	Add subscribers: Once a publication is configured, the publisher can invite other PI Cloud Connect users to subscribe to that publication. In the case of OSIsoft Academic Hub, the subscriber will be AcademicHub@osisoft.com. See the Add subscribers topic for more information.

PI Cloud Connect Onboarding Checklist	Yes/No
Define the PI Cloud Connect account and User Account for the PI Connect node Windows Service	
Define/create the Microsoft account used for administrating the account	
Read and agree with Terms of Use and Privacy Statement	
Validate that an activation email has been received from the PI Cloud Connect customer portal, and activate within 48 hours	
Signing in the Customer Portal is successful	
Navigation within the Customer Portal is successful	
Data to be shared is organized in PI AF on university's PI System	
Security around accessing the data is defined/known	
PI AF data targeted for publication does not contain unsupported objects	

Creating New Data Views

OSIsoft will automatically generate data views based on the elements in the new dataset, so the data views will be asset centric. Custom data views can be created as well if students want to integrate their Academic Hub data with other analytical tools. OSIsoft encourages students to create additional materials such as notebooks, code, scripts, curriculum, applications, and papers based on the community datasets, and in the future plans to provide a dedicated space to upload these items to share with the academic and research community. Application development provides hands on experience in requirements gathering and problem definition and developing an idea from an identified need. Application development also offers experience in designing, developing, and testing. These skills and the typically iterative process and nature of application development teach students valuable lessons that are directly applicable to the current work environment and roles the students will encounter once they graduate.

Online Curriculum

To complete the process outlined above, TA's will need to complete some of OSIsoft's online training. Brief descriptions and links of relevant training courses are provided below, along with the type of user that will likely benefit from each training course.

All courses are found on [OSIsoft's Learning Site](#). TA's may contact their OSISoft Academic Account Manager to request the discount code to receive free access to the training materials.

The following courses will help TA's learn to administer their university's local PI System and configure a digital twin, which will have an associated asset structure and asset-based analytics. Shown are the Course Title, Course Length, and which roles would benefit from completing each course.

1. [PI System Basics](#) (1 h): All TA's and interested faculty should complete this
2. [Configuring a Simple PI System](#) (16 h): TA's configuring data collection and anyone responsible for administering the PI System should complete this
3. [Building Asset Hierarchies with PI AF](#) (12 h): TA's designing the asset models and building the digital twin model should complete this
4. [Configuring Analytics with PI AF](#) (12 h): TA's creating real-time data analytics (e.g. soft sensors) for the digital twin model should complete this
   Recommendation: Do NOT implement Event Frames for your setup, there is no easy migration path for these to the Academic Hub and OCS

Additional Engagement Opportunities

OSIsoft likes to engage with all different types of customers, especially university students, staff, and faculty. Each university has a designated Academic Account Manager, who is responsible for the relationship between OSIsoft and the university and is the liaison for user feedback, support issues, suggestions, and additional engagement opportunities both for the university and individuals. The Academic Team as a whole can be reached at Universities@osisoft.com.

There are several ways to participate in the OSIsoft community at large, encompassing customers and employees, end users and developers alike. OSIsoft utilizes [GitHub](#) to provide sample code for OCS. The [PI Dev Club](#) is the place to connect and collaborate with OSIsoft engineers, partners and fellow developers online through forums, webinars, blogs and online courses. There are [OSIsoft User Groups](#) including very active Pharmaceutical, Oil and Gas, and Power Transmission and Distribution. A dedicated Academic space is targeted for 2021. Once in person events return, OSIsoft will resume hosting their annual [Users Conference](#) in the San Francisco Bay area. The UC has typically featured an [Academic Symposium](#) where students and faculty presented their work and shared ideas and experiences with colleagues from around the world.

Students

In addition to the learning and community opportunities detailed in this section, one way to continue engaging with OSIsoft is to come work for OSIsoft as an intern or a full-time employee! The bulk of new hires come from universities and our intern pool, as OSIsoft likes to grow employees and promote from within. New hires begin with a "rotation orientation process," a comprehensive introduction to every facet of the worldwide organization and the myriad industries and top enterprise companies served. It's a unique opportunity to explore the wide-ranging options available at OSIsoft, and after that most new hires move into one of two roles.

Software Developer: Be part of a specialized, energetic development team designing, coding, testing, documenting, and supporting a variety of software products.

Product Support Engineer: Become a member of the Customer Support team that provides best-in-class product support to our global customers in order to achieve their business outcomes. In this role, you will have the opportunity to use your problem-solving skills every day on a range of challenges, across multiple industries.

OSIsoft hires people with bachelor's degrees, master's degrees and PhD's in Engineering, Math, Computer Science, Natural Sciences, or related fields. Additional language skills a plus, specifically Spanish, Portuguese, or French. Please apply on [OSIsoft's career site](#) for recent graduates and interns. Also, on the career site is information about OSIsoft's benefits such as paid time off for volunteering and educational expense reimbursement, along with strong compensation, 401K + match, and paid time off.

Faculty and Research Staff

Faculty members can engage with OSIsoft's Academic Group across semesters and years, creating additional content in the form of curriculum, notebooks, datasets, apps, papers, and presentations. OSIsoft regularly works with partners such as faculty and customers to publish white papers, and present at OSIsoft's User Conference as well as different industry organization conferences globally every year.

Facilities and Sustainability Departments

Facility managers are monitoring billions of building data points - from HVACs and lights to fans and elevators - in real time with a unique data infrastructure. OSIsoft technologies transform campuses into living labs. Students engage with real-world operations data and apply their skills to sustainability efforts that better their campuses and communities. In a living lab, academic research is tied to campus data, enabling a virtuous cycle of innovation and operational improvements that benefit the university. Visit OSIsoft's Operational Insight for Facilities page for more information.

OSIsoft Learning and Training Resources

Continuous learning is engrained at OSIsoft, and our employees complete over 80,000 hours of training in a typical 12-month period. OSIsoft has extensive and diverse training available to students and faculty. OSIsoft's Learning main page is the jumping off point for a wide variety of training options.

The OSIsoft Live Library houses the depth and breadth of PI System documentation, organized by topic and continually reviewed for accuracy. Topics include: PI Servers, interfaces and connectors, cloud services, developer technologies and more.

The Virtual Learning Environment on the OSIsoft Learning site gives customers access to a working PI System 24 hours a day, 7 days a week. Customers use the Virtual Learning Environment to learn new skills, test knowledge, or try things with the PI System that cannot be done safely on a production system. OSIsoft has dozens of learning labs on topics ranging from simple displays and reports to system management and programming. Each lab has a workbook and video. Users who complete and pass courses can earn certificates, which can be displayed on LinkedIn where the certificates can be seen and searched for by potential employers.

The OSIsoft Learning Channel on YouTube is a convenient way to learn about the PI System anytime, anywhere. Customers can view videos to learn a how to install, configure, and use OSIsoft products. The videos capture basic and advanced tasks step by step. Playlists are groups of similar videos that make it easier to learn more about specific topics and products. Anyone can subscribe to the channel to get notifications when OSIsoft releases new videos.

The OSIsoft Resource Library contains the most comprehensive repository of all things related to the PI System and OSIsoft. Choose from a wide spectrum of articles, case studies, presentations, webinars, and white papers, and sort by year published to narrow down your results. There are over 4,000 presentations, and more are continually added.