EnergySHR

EnergySHR Portal

Frequently Asked Questions

Manual

Introduction

Welcome to the Frequently Asked Questions for the EnergySHR platform. This is the place where we collect questions that are more frequent when using the EnergySHR platform

Who is the initiator / owner of EnergySHR?

TU Delft and Erasmus University Rotterdam jointly have taken the initiative to develop EnergySHR.nl as part of the Convergence. See https://www.energyshr.nl/about

Why do you only focus on the energy transition?

We believe that bringing together data scientists, data consumers and data providers, like grid operators, local government, energy service providers and energy companies, facilitates and accelerates cooperation within the energy transition domain.

Who can use EnergySHR?

Everyone is able to download datasets and algorithms that are published as 'downloadable' without login. Also executing compute jobs in case the dataset and algorithm is private is possible without login.

In this starting phase, only a selected user group from Delft University of Technology and Erasmus University is allowed to publish datasets and algorithms.

The ambition is to gradually improve the platform and expand access to publishing, so that the platform is accessible to researchers across Europe.

How can I join with a usecase?

If you have ideas for a use case for the next phase in which you would like to participate, please contact: c.g.duterloo (at) tudelft.nl or doron.gollnast (at)eur.nl.

How do I login?

To use the EnergySHR platform you must be a member of the EnergySHR collaboration (see previous question). If you are a member of the EnergySHR collaboration you can use your institutional account or an EduID_NL account (free) and its credentials to login

Why can't I login?

In this starting phase, only a selected user group from Delft University of Technology and Erasmus University is allowed to login and publish datasets and algorithms.
If you are part of this user group, but can't login with your *institutes credentials*, please request to join the collaboration here.
If you still have troubles to login, please send a mail to energyshr(at)tudelft.nl or energyshr(at)eur.nl

Where can I store my data best?

For datasets and algorithms that aren't sensitive (privacy or secrecy issues), you can host your datasets and algorithms on public repositories like 4TUResearchData , Zenodo or Github

Why would I use Compute2Data?

The compute2Data option is a usecase that is especially interesting in case the dataset can't be shared (due to for example privacy or secrecy issues) but you like algorithms to be developed based on this dataset.

Why is there a need for Compute to Data?

When publishing datasets or algorithms you might not always want this data to be available to anyone and overyone. Especially datasets, as they might contain Privacy sensitive or commercial information for instance. With Compute to Data you can make sure that the original algorithm and dataset cannot be downloaded. What happens is that an algorightm that is vetted by the dataset owner, can be run on the dataset. Only the output of the algorithm can then be downloaded by the end user.

How can I publish a dataset or algorithm?

In order to publish algorithms or datasets, you will need to join the EnergySHR collaboration. EnergySHR is using SURF. If you are a member of an Education or Research institute you can use your institute's credentials. Otherwise you will have to create en EduID_NL account (free). For this last option you will be asked to install an App on your mobile device. This app is needed to complete the creation of the EduID_NL account.

Are algorightms and datasets hosted by the EnergySHR plarform?

No any algorithm and dataset is externally hosted. The EnergySHR platform acts as a middleman. It protects the original URL of the dataset/algorithm. This means that users will not be able to deduce this URL. The so called provider within the platform plays the central role of acting as a proxy whenever you request to download an algorithm or dataset. The provider ensures that you are allowed to access the dataset and/or algorithm. It then requests the original URL and passes on the content of the data transparently to you. The provider is also responsible for the encryption and decryption of the metadata and original URL of the algorithm and datasets.

Why are there transactions involved?

The EnergySHR platform is build using the Ocean Procotocl Marketplace software. Ocean Protocol is a blockchain project. This decentralized approach means there are transactions involved in making sure actions are recorded in the distributed blockchain. Normally you would need a so called web 3 wallet to sign these transactions. However in the EnergySHR platform this is handled transparently for you in the background.

What languages are supported for algorithms?

If you are providing you algorithm only fow download, then any algorithm language can be used. Be aware you will have to provide a Docker image. You can simply leave the default Node based docker image selected during the publish phase. Unfortunately you will have to provide a docker image, even if the algorithm will never be used directly in the Compute to Data environment. In reallity for a downloadable algorithm the Compute to Data environment will never be used. So even if your algorithm is not using Docker or written in another language than Javascript, you can still use the NodeJS docker container in this case.

In case you want your algorithm to run in the Compute to Data environment you can chooise between the standard provided NodeJS (javascript) and Python images. However in a lot of cases you might need additional libraries, modules and/or languages for your algorith, In order to facilitate that you can choose to create or use a custom Docker image. This docker image needs to be hosted using a Docker registry that is available to the EnrgySHR provider. Currently private repositories using authentication/authorization not supported. If you want to make sure that a Docker Image needs to be protected, be aware that no one will know the actual URL of the docker image. However most publicly available registries would allow others to find out the URL, as they might be indexed by search engines and/or provide a search functionality on their registry platform. One option would be to use/host a registry service that would allow you to only allow certain IP addresses. End users will never interact with the docker repository. Only

the Provider will have access, meaning you can quite easily use a firewall or configure the repository software to only allow access from this IP.

What is the relationship between the algorithms docker image and algorithm itself?
When publishin an algorithm you will have to provide both a Docker Image and a URL for the algorithm. The Docker image is reponsilbe for providing the supporting libraries, modules and algorithm language support. The actual URL

Does EnergySHR store files/data?

No EnergySHR does not store data. When publishing you are providing the URL of the dataset or algorithm, which will never be known to end users. When a user downloads the dataset or algorithm this URL will be access by the provider, which acts as a proxy. The end user only interacts with the provider.

The EnergySHR platform only makes algorithms and datasets available to end users and allows algorithms to execute against datasets providing results that can be downloaded. The actual datasets and algorithms are hosted elsewhere .This could be using IPFS or a regular URL, like for instance on AWS, Google Cloud a regular webserver. That URL will never be communicated to end-users as the EnerySHR platform will always act as a middleman.

Can I use private Docker repositories with authentication?

Most publicly available docker registries would allow others to find out the URL, as they might be indexed by search engines and/or provide a search functionality on their registry platform. Unfortunately current platform does not allow for private repositories protected by authentication and authorization. One option would be to use/host a registry service that would allow you to only allow certain IP addresses. End users will never interact with the docker repository. Only the Provider will have access, meaning you can quite easily use a firewall or configure the repository software to only allow access from this IP

Can I make my algorithm or dataset available both as publicly downloadable and compute to data?

Unfortunately you will have to choose whether a dataset or algorithm can be downloaded or that it can be run in the Compute to Data environment. You cannot select both at the same time.
If you really want your asset to be downloadable and executiable you can decide to publish 2 versions of your dataset and/or algorithm. One will be available as download, the other as compute to data asset.

What if I want to make all the algorithm code available in the docker image itself without providing a URL to an algorithm/script?

That is an option that is possible. However be aware that in current EnergySHR platform it is not possible to easily update a docker image once it is published as part of an algorithm. You will have to retract the algorithm and publish a new one. Algorithm URLs themselves can however be updted. So we suggest that the Docker image contains any supporting modules and/or libraries as well as the language support for your algorithm, but that the algorithm code itself is published using the URL you need to provide in the publish step.

If you do want to include all the code including the algorithm in the Docker container, do make sure you create a small bash script as algorithm for instance. The reason is that the algorithm will be used as a Docker entrypoint when running the docker container. So the script in this case would need to call the algorithm in the location within the Docker container

Can anyone create algorithms for my dataset and publish them for the Compute to Data environment?

When using the Compute to Data option of the EnergySHR platform the owner of the dataset always determines which algorithms can be run on the dataset. Anyone with an account would be able to publish an algorithm, but in the end the owning organization of the dataset needs to allow the algorithm to be run on the dataset. After that permission is given, end-user will be able to select the algorithm to run against the dataset.

I do not want my dataset to be publicly available. How do I protect it?

When publishing a dataset or algorithm you have to make a choice whether the dataset or algorithm is publicly downloadable or can only be used in Compute to Data jobs. In all cases the algorithms and dataset cannot be access directly by end users. They will never see the original URLs, as the central EnergySHR provider acts as a proxy.

How does the Compute to Data solution work?

Whenever you run a compute job (algorithm). The EnergySHR Compute to Data environment that is running as a Kubernetes Cluster, will download the Docker image associated with the algorithm, from the URL provided by the algorithm owner. Then the dataset URL will be

accessed by the provider and will be mounted as a file within the docker image at the path /inputs. Next the algorithm itself will be downloaded and made available to the docker image. Then the algorithm will run, yielding outputs, which will be placed in the /outputs folder. After the algorithm is finished the Kubernetes cluster will cleanup all resources and only the resulting output will remain available for a maximum of 30 days.

How do Docker image work, or how can I create a Docker image?

Docker is a well-known way of creating binary images that allow software to run, no matter the target environment, operating system, hardware or the architecture of a hosting provider. How docker works, or how to create docker images is out of scope of this FAQ. You can find many resources on the internet about docker and can always visit https://www.docker.com/ for more information and documentation