

InGeO: The Integrated Geoscience Observatory

Todd Valentic
SRI International
Menlo Park, CA
todd.valentic@sri.com

Asti Bhatt
SRI International
Menlo Park, CA
asti.bhatt@sri.com

Leslie Lamarche
SRI International
Menlo Park, CA
leslie.lamarche@sri.com

Pablo Reyes
SRI International
Menlo Park, CA
pablo.reyes@sri.com

Ashton Reimer
SRI International
Menlo Park, CA
ashton.reimer@sri.com

Abstract

Geoscientists arrive at scientific results by analyzing observations from a diverse set of instrumentation and often assimilate them into a model. Effective collaboration between scientists can be hampered when they are using different resources, resulting in a lengthy and laborious process. Individual researchers need to assemble many of the community resources on their own before they can conduct successful research or get credit for their work.

The Integrated Geoscience Observatory (InGeO) project tackles the problem of seamless collaboration between geoscientists by creating a platform where the data from disparate instruments can be brought together with software tools for data interpretation provided by the instrument operators.

The primary goals of the InGeO project are to:

1. Provide tools that make it easy for geospace researchers to collaborate, share work, reproduce results, and build on tools that have already been developed.
2. Educate the geospace community on best practices for software development and data archiving to ensure data and the tools needed to work with it are available to a broad range of researchers in the community with minimal barriers to entry.

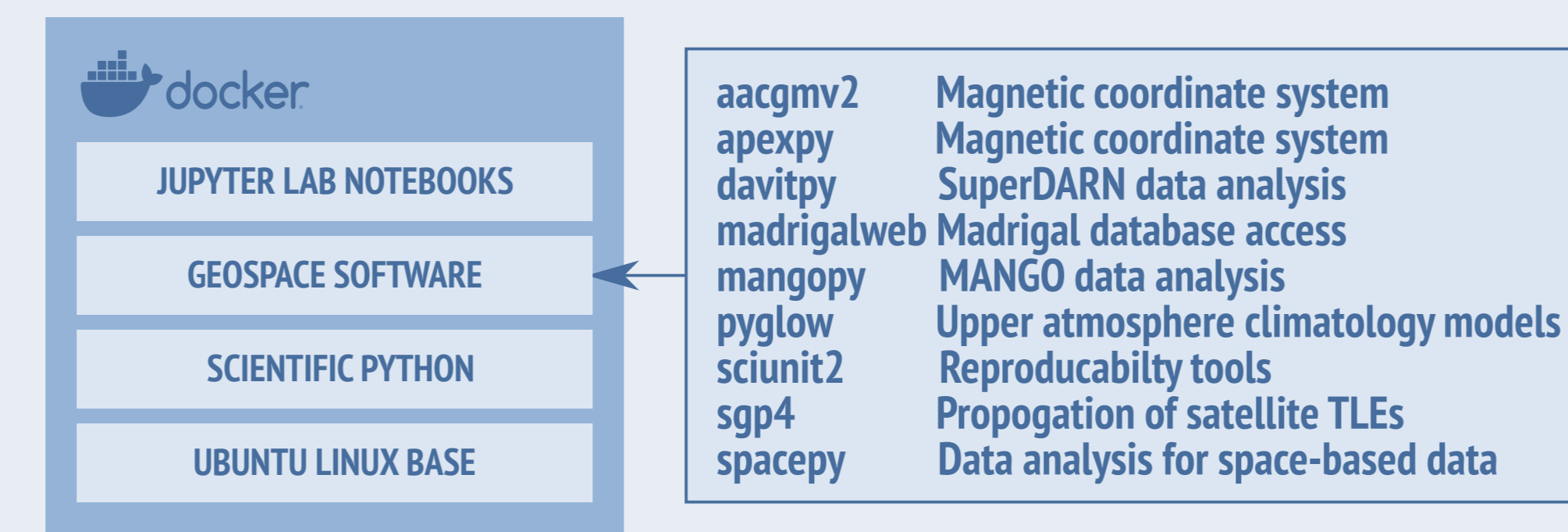
Our solution to the first goal has been Resen - a tool to enable reproducibility and collaboration. Resen allows community developed toolkits to be easily accessed and creates a convenient mechanism for researchers to save and share their results along with the analysis that produced them. Resen is written in Python and uses Docker containers. The current software packages in Resen allow you to access common geospace data sources including MANGO, Madrigal and SuperDARN. We plan to add functionality for other data sources in the future releases.

Acknowledgements

Funding for this research was provided by the National Science Foundation under the grants EarthCube ICER-1541057 and Cyberinfrastructure for Sustained Scientific Innovation CSSI-1835573.

Fall AGU 2019 - IN33B-0827

1 The Foundation: resen-core Docker Image



DockerHub: earthcubeingeo/resen-core

The resen-core Docker image builds upon a standard Ubuntu Linux image and comes pre-installed with a stack of commonly used geospace libraries and software packages. This approach allows users to avoid the arduous and often time-consuming task of manually installing software on their local systems and provides a consistent platform for collaboration and reproducibility. New releases are made twice a year to incorporate newer versions of the packages. The releases are available on DockerHub.

2 The Bucket: A software, data and platform bundle



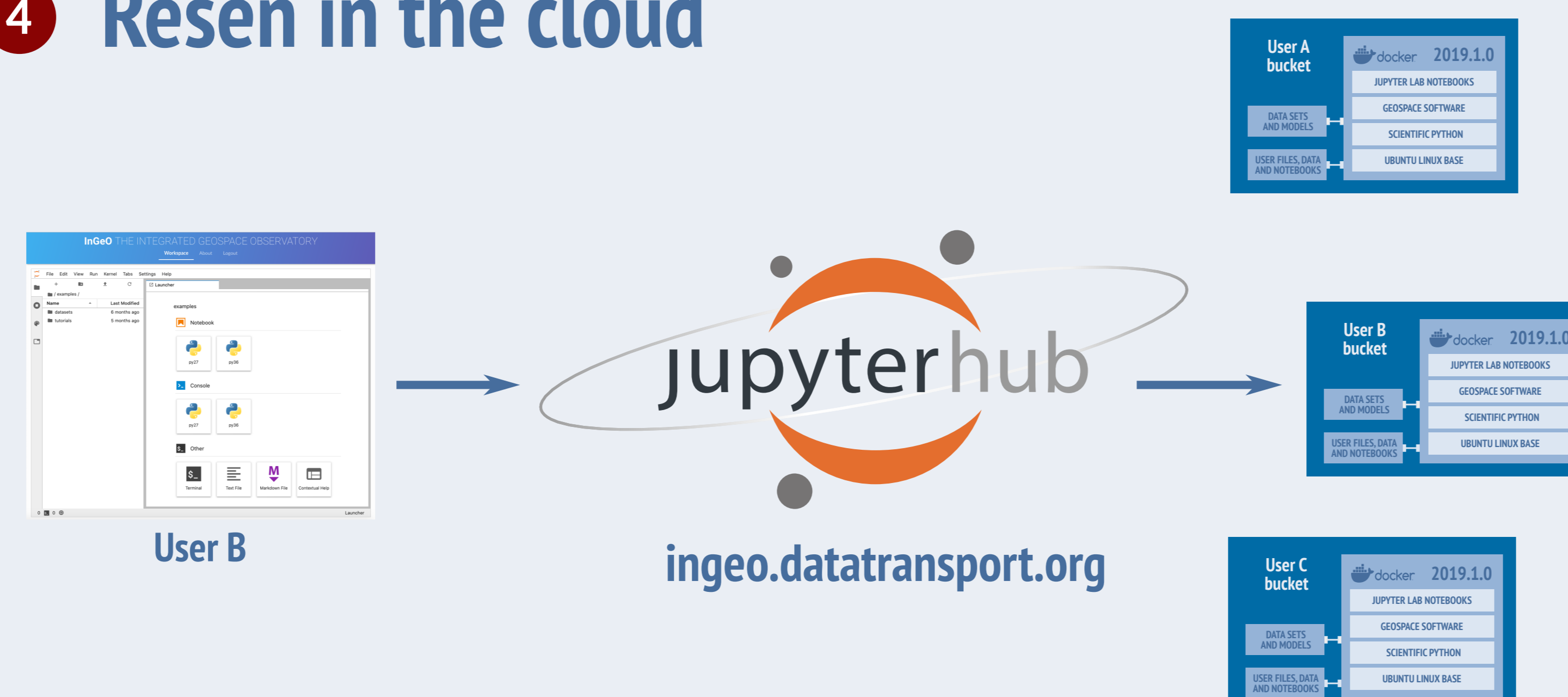
Within Resen, data analysis environments are called buckets, consisting of a Docker image and mounted file volumes where a user's files will be stored. In addition, there is a mount for common data sets and models. The resen create command is used to make a new bucket, which is based on one of the available resen-core images.

```
[resen] >>> create
Enter bucket name: case_study
Select a version: 2019.1.0
Mount storage? yes
```

Resen: The Reproducible Software Environment

A tool to enable reproducibility and collaboration in the geospace sciences.

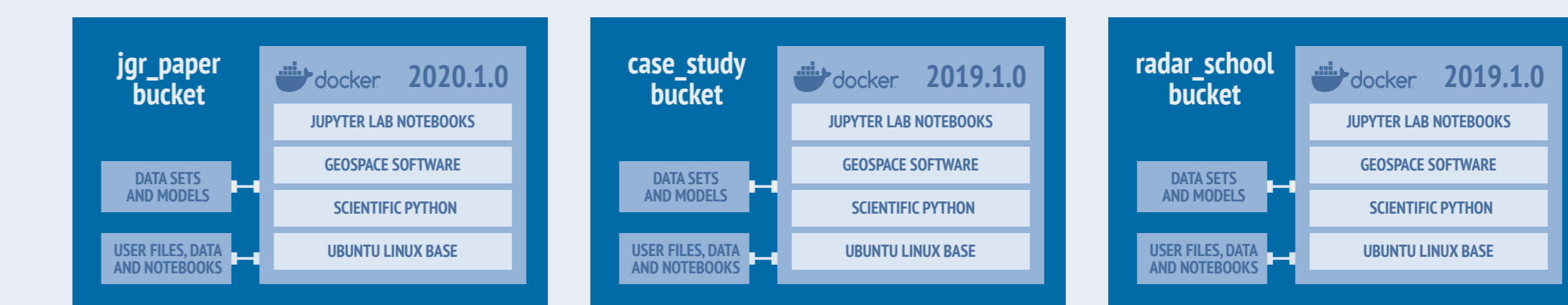
4 Resen in the cloud



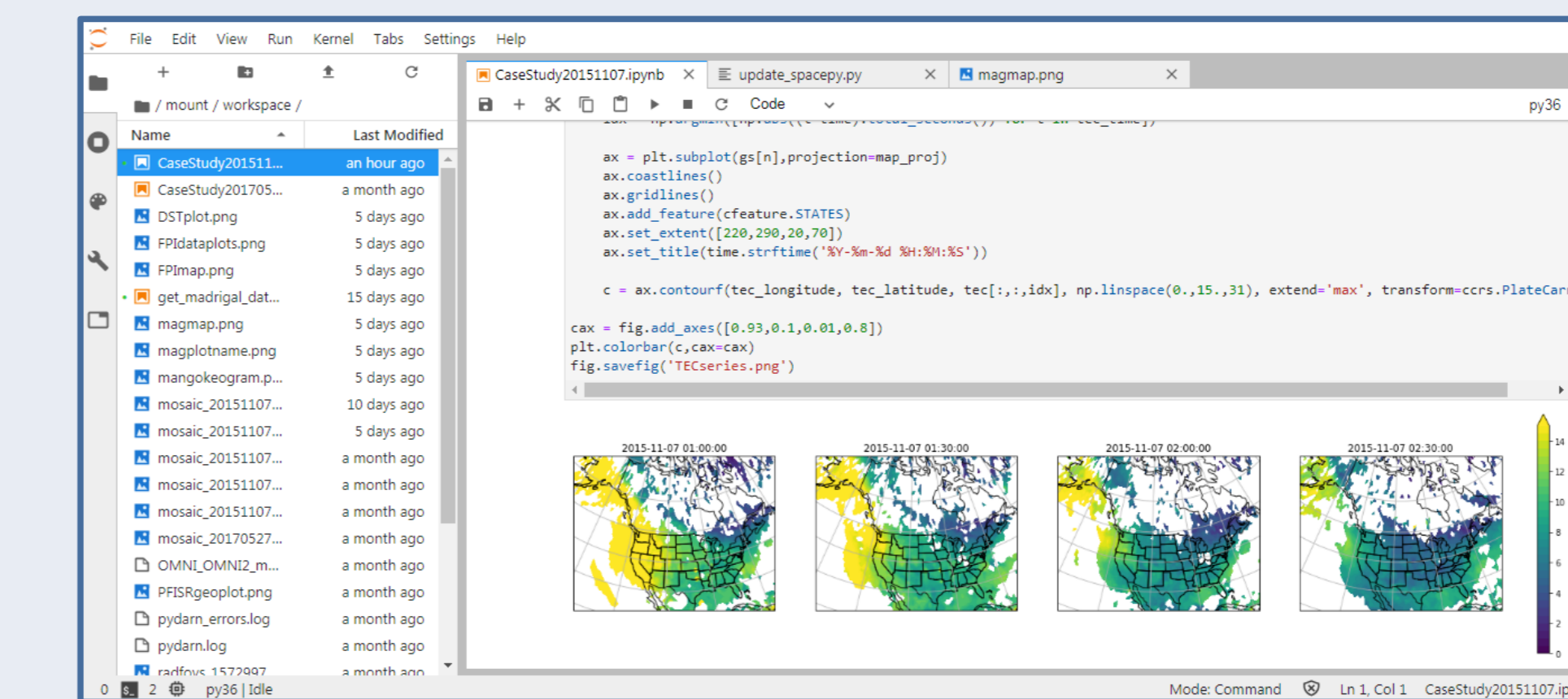
Resen buckets can be run in the cloud using JupyterHub to connect users with their own server-hosted buckets. Users can quickly try out the system simply by signing up. Nothing needs to be installed on their computer. Special spins of the Docker images could be made to support activities such as schools, which would come with tutorials and data sets preinstalled.

3 The Workflow

```
[resen] >>> list
```



```
[resen] >>> start case_study
```



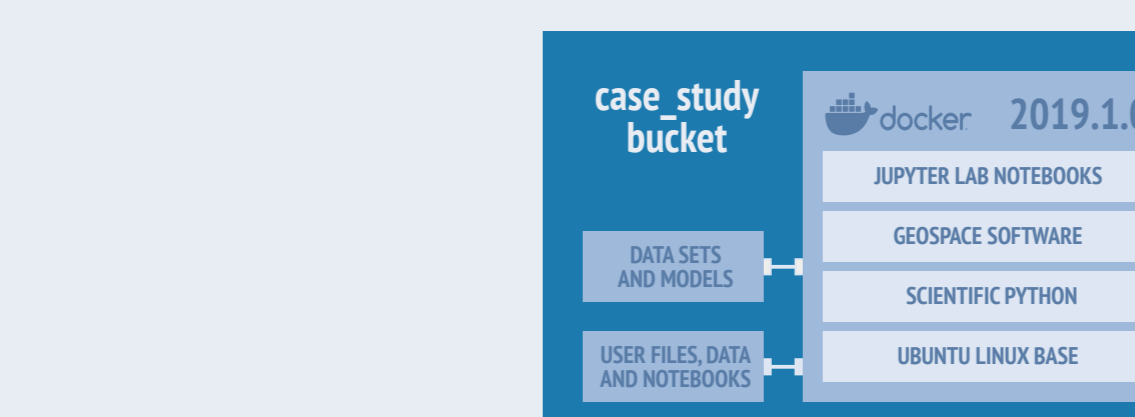
```
[resen] >>> export case_study
```



Download



```
[resen] >>> import
```



Use resen to start a bucket from your collection. Each of your data analysis projects will have its own bucket. Disk usage is minimized because the docker images are shared. At startup, resen launches the docker container, mounts the data volume, sets up the networking ports and starts running JupyterLab.

You access your JupyterLab compute environment through a web browser. You can upload data files, document your work in Jupyter notebooks and create plots. You can also open a terminal to access the command line to install additional software, develop programs and use the traditional Linux tools.

To share your results with others, you need to export the contents of your bucket. The resen export command builds an archive file containing your Docker image and a copy of the mounted user and data volumes.

An exported bucket can be shared with a collaborator or uploaded to a data repository such as Zenodo, where the bucket will be publicly available and citable with a DOI.

Anyone may download, execute and reproduce the research results or perform additional analysis without having to install dependencies or configure their own system.

Each exported archive contains all of the analysis code, results, data and docker image. To use it, resen import builds a new bucket from the archive and adds it to your collection. You can then start the bucket to access the contents.

For more information

InGeO website: <https://ingeo.datatransport.org>
InGeO github: <https://github.com/EarthCubeInGeo>
Resen documentation: <https://resen.readthedocs.io/en/latest/>
DockerHub images: <https://hub.docker.com/r/earthcubeingeo/resen-core>

