

Chat during March 15, 2021, session

from jon vandegriff to Everyone: 4:18 PM

<http://hapi-server.org/servers>

from Rebecca Ringuette to Everyone: 4:50 PM

Since Kamodo can manipulate data from HAPI, can't we already compare model and observational data?

from Rebecca Ringuette to Everyone: 4:58 PM

If they can just make the data compatible with a HAPI server, then the file format doesn't matter.

from jon vandegriff to Everyone: 5:00 PM

Darren is the farthest along on using HAPI for data-model comparisons. I know he's getting ephemeric data from HAPI, and also I think measurement data too. Is that right Darren?

from ROBERT CANDEY to Everyone: 5:04 PM

The ISTP Guidelines provide more structure and metadata than is transferred in HAPI. Unfortunately I have to attend another meeting.

from Rebecca Ringuette to Everyone: 5:06 PM

Can HAPI work towards communicating more of the information provided by ISTP?

Asti Bhatt comment

Just wanted to bring this up as a reference. We have created a user-friendly wrapper around Docker for geospace scientists called 'Reproducible Software Environment (Resen)', which works on a similar concept, but it comes with images that has several space science Python libraries installed. So users can use geospace software without having to install it themselves. <https://ingeo.datatransport.org/home/resen> While this can be used on your own machine, we have also created an online version of Resen where users can use the whole system in the cloud. <https://ingeo.datatransport.org/home/resen/resen-online>

Chat during **March 16, 2021, session**

from Asher Pembroke to Everyone: 5:02 PM

Docker/Singularity container infrastructure

- * may be used to house cononical data preparation pipelines
- * raw data may be mounted into cleaning container
- * can contain all code/algorithms/environment needed
- * container registries allow users to pull the latest version of any container
- * allows training/cleaning to be performed in the cloud

Resources:

- * Why use containers for Machine Learning <https://aws.amazon.com/blogs/opensource/why-use-docker-containers-for-machine-learning-development/>
- * Singularity <https://sylabs.io/guides/3.7/user-guide/>
- * Docker <https://www.docker.com/resources/what-container>

from Asher Pembroke to Everyone: 5:04 PM

ML training on AWS <https://aws.amazon.com/blogs/opensource/why-use-docker-containers-for-machine-learning-development/>

from Viacheslav Sadykov to Everyone: 5:23 PM

Sorry, I have to disconnect now. Thanks everyone for the discussion!

from Rebecca Ringuette to Everyone: 5:25 PM

Multiple versions of the same data should be available simultaneously, not just the best available versions. For example, the community tests new versions and calibrations against earlier ones. This is how some errors in calibrations have been corrected.

from Rebecca Ringuette to Everyone: 5:27 PM

For models, the doi per run could point to a container hosting all input data and parameter values

from Karin Muglach to Everyone: 5:30 PM

I agreew with rebecca!

from Asher Pembroke to Everyone: 5:30 PM

The "doi" would just be the container tag + version

from Rebecca Ringuette to Everyone: 5:30 PM

The container, possibly a RESEN container, should also have the model specific doi and either the output data or a doi that points to the output data for that run

from Asher Pembroke to Everyone: 5:31 PM

well container id (which is a hash of the image) + tag/version

from Rebecca Ringuette to Everyone: 5:31 PM

what do you mean by version, asher?

from Rebecca Ringuette to Everyone: 5:33 PM
But the container shouldn't change....

from Asti Bhatt to Everyone: 5:33 PM
RESEN for example has multiple images (resen-cores) that you can use in a container, so I suppose the container id could contain that as well

from Rebecca Ringuette to Everyone: 5:42 PM
Could python go now?

from Rebecca Ringuette to Everyone: 5:55 PM
For model input/output, is could have the ccmc acronym, an acronym for the model used, and followed by the run number

from Asher Pembroke to Everyone: 5:57 PM
This is Zooko's Triangle

from Rebecca Ringuette to Everyone: 5:57 PM
?

from Arnaud Masson to Everyone: 5:57 PM

<https://doi.org/10.1016/j.asr.2021.01.035>

from Rebecca Ringuette to Everyone: 5:58 PM
so some combination like that used by papers: ccmc-(randomstuff)

from Rebecca Ringuette to Everyone: 6:00 PM

There are some instruments with the same name. For instance, there is a CubeSat mission coming up with the same name as a popular astronomy mission. (Gaia I think.)