The background is a gradient of dark blue and purple. On the left side, there are several concentric circles and a large circular scale with degree markings from 140 to 260. Some of the circles have arrows indicating a clockwise direction. The overall aesthetic is technical and futuristic.

# NEXT GENERATION NATIONAL WATER MODEL DATA STREAM

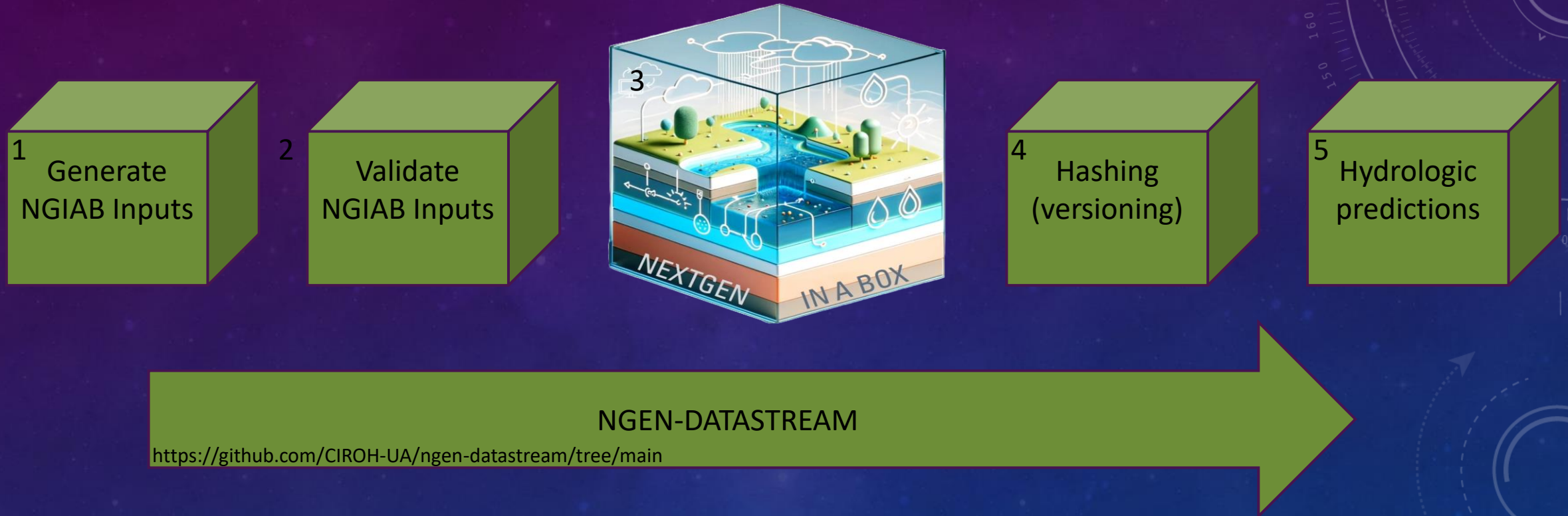
JORDAN LASER

ZACH WILLS

NELS FRAZIER

FALL 2023

# NEXTGEN Data Stream Conceptual Model



ngen-datastream refers to the software chain that builds valid ngen-run input packages, executes NEXTGEN with NGIAB, and versions the entire run. NEXTGEN is a framework that coordinates hydrologic model configuration, parameterization, and execution across spatial domains. NGIAB consists of the NEXTGEN docker container pre-built for the user's architecture as well as software that aids users in properly running the container.

<sup>1</sup> Generate  
NGIAB Inputs

GUI

<https://github.com/NOAA-OWP/ngen-cal>  
Configuration

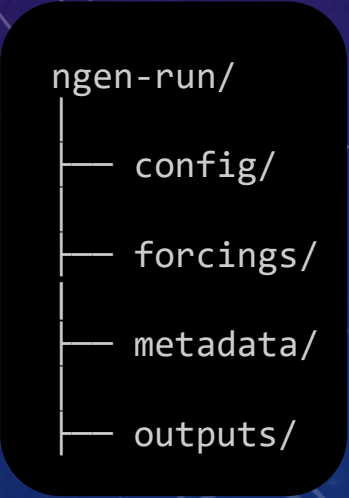
<https://github.com/LynkerIntel/hfsubset>  
Subsetting

<https://github.com/CIROH-UA/ngen-datastream/tree/main/forcingprocessor>  
ForcingProcessor

```
ngen-run/  
├── config/  
├── forcings/  
├── metadata/  
└── outputs/
```

# Input explanations for NGIAB/NEXTGEN

	Configuration	Hydrofabric	Forcings
Physical Explanation	Deciding how to model hydrologic data is not easy. There's a ton of water models with different versions, configurations, and parameterizations that makes running and interoperation difficult. To add to the complexity, the performance of these models has been found to depend strongly on the environment, implying some models work better than others depending on the spatial domain. NextGen gives users control over modeling this complexity.	The hydrofabric is the spatial data that describes the physical location of things like the borders and topology of a catchment, nexus points, and flowpaths.	Forcings are the physical variables needed to "force" NextGen from one time step to the next. Forcings variables include temperature, precipitation, and solar longwave and shortwave intensities. These files are per catchment time series of these variables.
Computer Science Explanation	Here is where the user supplies NextGen with the information it needs to execute a "run". This includes information like the time range, model selection, model configuration and parameterization. It is possible to specify the model selections on a per catchment basis. This allows the user to run different models over different areas and can dramatically increase the complexity of this file. The goal of NGIAB is to dynamically produce this file via a GUI.  e.x. ngen-run/config/realization.json	This data is most readily available in .geopkg per VPU. Users often subset these .geopkg's into smaller domains with tools like hfsubset. NextGen only accepts .geojson format, so users will then need to conver their .geopkg to .geojson with tools like og2og2  e.x. ngen-run/config/catchments.geojson	More on how to create these files later.  e.x. data_dir/forcings/cat-12.csv



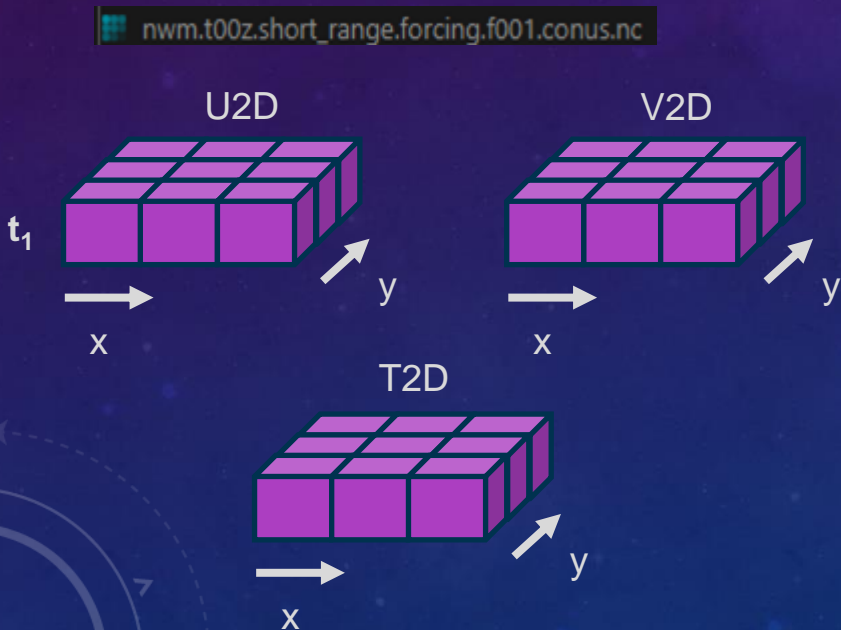


# Why do we need a forcingprocessor?

## NWM output forcing data

- Gridded
- CONUS wide
- One file for each forecast time

```
nwm.t00z.short_range.forcing.f001.conus.nc
nwm.t00z.short_range.forcing.f002.conus.nc
nwm.t00z.short_range.forcing.f003.conus.nc
nwm.t00z.short_range.forcing.f004.conus.nc
nwm.t00z.short_range.forcing.f005.conus.nc
```



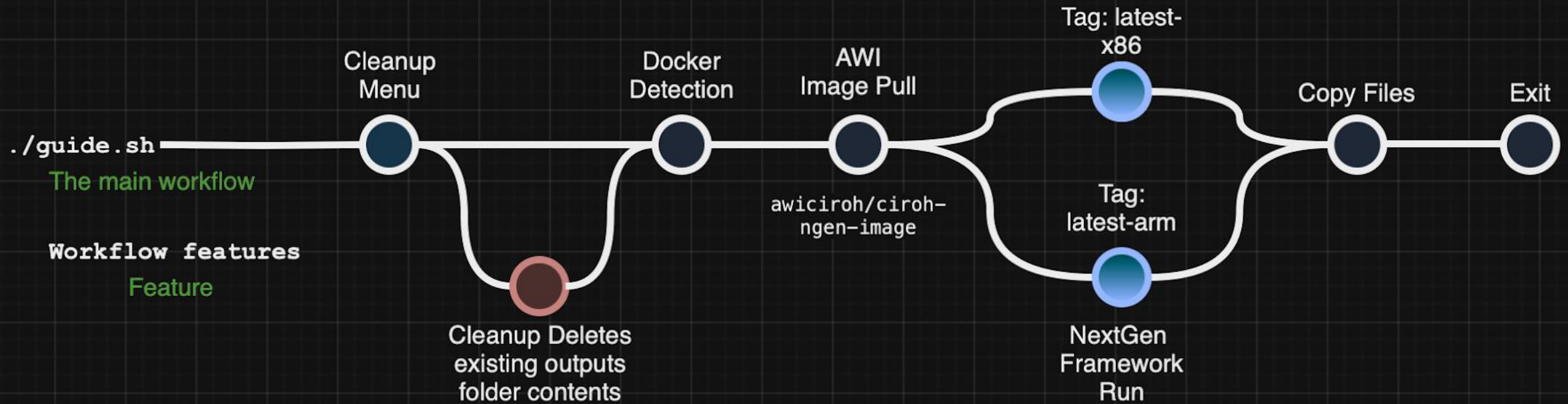
## NextGen input forcing data

- Non-Gridded
- Time Series within each file
- One file for each catchment

```
cat03W_106885.csv
cat03W_106886.csv
cat03W_106887.csv
cat03W_106888.csv
cat03W_106889.csv
```



## CIROH NextGen-in-a-Box Guide Script Workflow



# TERMS

- NGIAB – Next Generation National Water Model in a Box
- NGEN - Next Generation National Water Model
- Catchment – geographic area characterized by a single location, a nexus, where all precipitation in the area runs off through. A drainage basin.
- Nexus – the singular point where water flows into or out of a catchment. Often a point along a river.
- Subsetting – To reduce a large geopackage (many catchments) down to a smaller geopackage (fewer catchments) . In effect, this is choosing the domain over which ngen will run.
- Hashing – SHA256 algorithm applied to files to generate a unique id for a file. Useful for preserving and distinguishing unique inputs.
- Validation – Ensuring the ngen input directory data\_dir has been constructed properly. Properly meaning that NextGen will not crash and will generate output data.

# LINKS

- DATASTREAM <https://github.com/CIROH-UA/ngen-datastream/tree/main>
- FORCINGPROCESSOR <https://github.com/CIROH-UA/ngen-datastream/tree/main/forcingprocessor>
- REALIZATION GENERATION AND NGEN-RUN FOLDER VALIDATION <https://github.com/NOAA-OWP/ngen-cal>
- HYDROFABRIC SUBSETTING <https://github.com/LynkerIntel/hfsubset>
- HASHING/VERSIONING <https://github.com/aaraney/ht>
- NGIAB <https://github.com/CIROH-UA/NGIAB-CloudInfra>
- <https://docs.ciroh.org/>
- <https://docs.ciroh.org/docs/products/tools/nextgeninabox/ngiab-intro>
- <https://github.com/NOAA-OWP/ngen/wiki>
- <https://mikejohnson51.github.io/hyAggregate/>
- <https://ciroh.ua.edu/>