Streamlining the Model On-boarding Process
Agenda

- Opening remarks - Jon Linker

- 01 CCMC challenges/needs - Leila

- 02 NASA LWS – Simon Plunkett
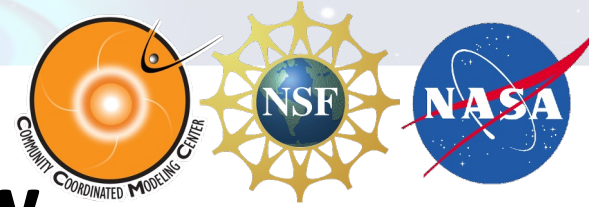
Modeler Developer Perspectives:

- 03 John Dorrelli

- 04 Tamas Gombosi

- 05 Shaela Jones

- 06 Jon Linker

- 07 Slava Merkin

- 08 Lulu Zhao

# CCMC Thoughts on Streamlining the Model Onboarding Process

Leila Mays and Tina Tsui

Our goal: make it easier and faster to onboard a new model/model upgrade to get it into the hands of the community faster

Ongoing Modernization:
Easier Onboarding and Implementation of New Models

- NASA LWS Strategic Capabilities program
  - all deliverables must be **installed and tested** at CCMC by the end of performance
  - development efforts include **ongoing inputs** from CCMC

- CCMC is maintaining **shared collaborative environments** on the AWS cloud (with NASA security compliance) and at NASA HECC systems.

- Establishing and following **best practices**. Shared **GitLab** repositories are possible.

- CCMC supports **containerizing** models and applications

- Working toward a **streamlined onboarding process**
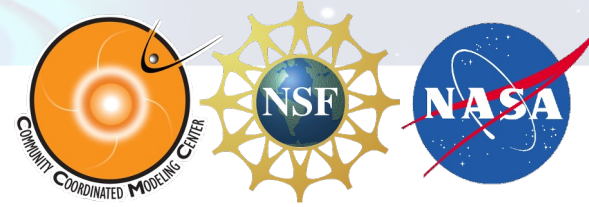
Making a model available at CCMC ROR can require significant effort from both sides

- ROR infrastructure needs to accommodate model requirements: data inputs, libraries and tools, web interface, workflow

- Model must accommodate requirements of ROR:
  - operational-level stability,
  - organized file inputs/outputs,
  - security and environment constrains,
  - metadata, and documentation:
    - Model and simulation run metadata are critical for improving CCMC services to users, but it's not easy to get these from developers
    - Model onboarding questionnaire – simplification in progress. *How else can we make this easier?*
   - Model documentation – *how can we better encourage this?*

In progress:
Work with developers to communicate and test these requirements from project start - what else can re: requirements
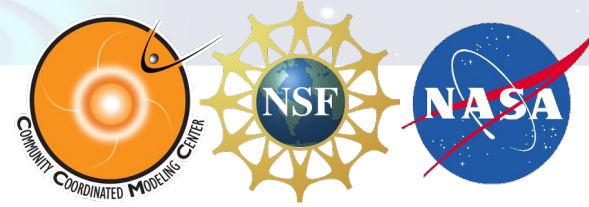
Model Onboarding

- Ongoing improvements: formalized the model onboarding process
https://ccmc.gsfc.nasa.gov/model-onboarding/
How can we improve the process communication? Is it the process transparent enough?

① **Pre-onboarding**

If your project is at the proposal stage, we ask you to fill out a short pre-onboarding questionnaire ☐.

A note on computational resources: CCMC will generally aim to meet all computational and technology resource requirements of the proposed model for Runs-on-Request. If we can' t meet the requirements with our in-house infrastructure, we can use resources at NASA HECC ☐ to execute the Runs-on-Request. We are open to all proposed innovative ways that CCMC can implement the proposed project.
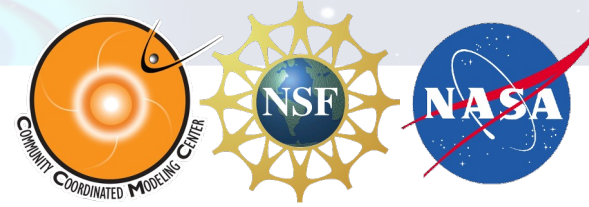
② Preparation

Modeler contacts CCMC to receive and complete pre-installation questionnaire (Word ⧉ or PDF ⧉). A CCMC point-of-contact is assigned. The questionnaire information is used to add the model to the CCMC Metadata Registry (CMR) ⧉ which automatically populate CCMC Model Catalog.

The modeler also provides sample model output data, visualizations (possibly with data reader or visualization source code) and identifies how to prepare inputs and run the model.

CCMC confirms the model metadata in CMR, identifies computing resources, creates a model visualization sample page (similar to this page ⧉) and a simulation output page.
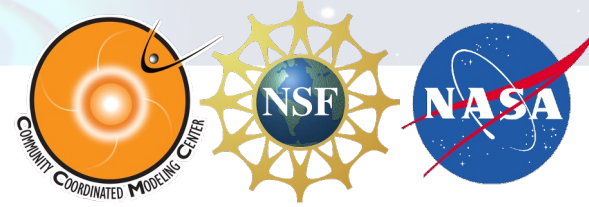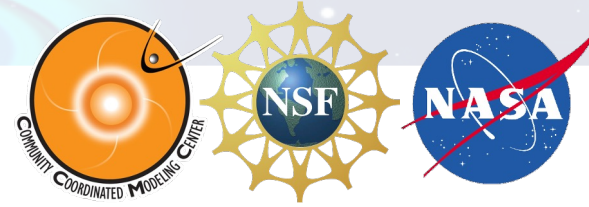
③ Model Implementation

All model onboarding and initial testing at CCMC is now done collaboratively with the model developer, on the cloud. We call this the *CCMC/Model Developers collaborative cloud environment.* For this purpose CCMC will provide account access to a cloud-based server, including a model-specific account that only CCMC and the model developers have access to. *CCMC uses the Amazon Web Services (AWS) Commercial Cloud. These cloud services meet the security criteria of Federal Information Security Management Act (FISMA) Low standards, and have NASA Authority to Operate (ATO).*

For certain projects, we may also work collaboratively in a shared space on NASA HECC supercomputers.

Modeler supplies model source code and documentation (makefiles, preprocessing, run and postprocessing scripts, user manual), and begins the installation process on the provided cloud instance. The model developer and CCMC iterate and perform test simulations together.

CCMC creates internal documentation to track the model onboarding process, source controls the code and documentation, re-installs the model on on-site servers at GSFC and performs tests. CCMC also creates a prototype model output and web visualization page for a test simulation

④ Test/Evaluation

CCMC creates a prototype run submission web interface for <u>Instant Runs (IR)</u> or <u>Runs-on-Request (RoR)</u>, for review by the modeler. CCMC and the modeler test the pipeline from end-to-end (executes model run, post-processing, and publishing results to the web).

CCMC recommends that the model development updates are installed and tested periodically on the collaborative cloud environment and/or the shared space on NASA HECC, once the proposed project begins. The model developers and CCMC are expected to continually document the installation and testing process.
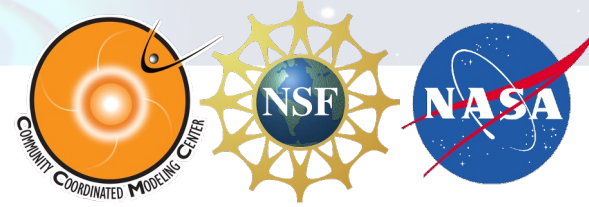
⑤ Production and Upgrades

Once the modeler is satisfied with the results, CCMC offers the model to the research community. CCMC will continue work with modeler to address any bug fixes and visualization issue(s) as issues occur. Model upgrades can be provided at any time and these are onboarded in the (4) Test/Evaluation step.

Onboarding Improvements for Continuous/Real-time Runs
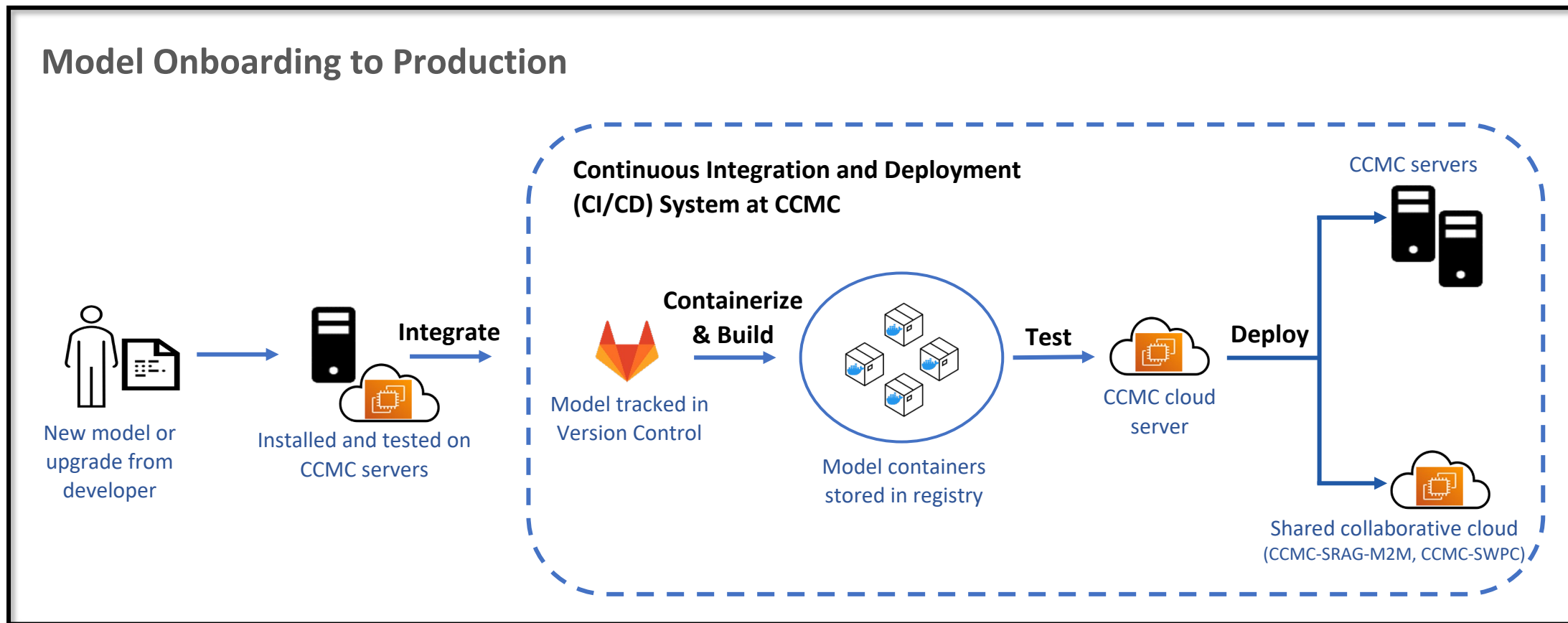
Moving the model (reinstallation of the model and its dependencies) between servers is not always straightforward. Ongoing improvements: containerize serial models, follow version control best practices, better documentation.
Container benchmakring/testing for parallel models in progress

CCMC Infrastructure and Computational Resources

➔ ***In-house Resources:*** *beowulf clusters, servers, web servers, storage servers*
- ◆ <u>Models in production (each model is installed on at least 2 servers):</u> Runs on Request, continuous/Real-time runs, instant runs
- ◆ Website, web tools/apps, iSWA and ROR archive

➔ ***Amazon Web Services (AWS) Infrastructure:***
EC2 (c5, m5, m5n, t2, HPC compute nodes: c6i.32xlg and m5n.24xlg), EFS, S3
<u>Model on-boarding and development</u> (CCMC/Model Dev shared environments)
- • Onboarding: 10 parallel models, 8 serial models, and 2 web applications with collaborators
- • CCMC development/deployment: Instant Run (IR) framework and dev website
- • Performance benchmarks
- ◆ <u>CCMC-SRAG-M2M and CCMC-SWPC shared environments</u>

➔ ***NASA High-End Computing Capability (HECC) HPC Supplement*** to in-house computational capabilities and model collaborative resource for computational intensive models: *Aitken, Pleiades, Electra*
- ◆ <u>Available now:</u> CORHEL, SWMF/AWSoM-R
- ◆ <u>In progress</u>: WACCM-X, HYPERS, GITM