

NITE (NWP Information Technology Environment): an infrastructure to facilitate development of NCEP numerical models

Laurie Carson¹ and Ligia Bernardet²

¹National Center for Atmospheric Research/Research Applications Laboratory and Developmental Testbed Center

²Cooperative Institute for Research in the Atmosphere/NOAA Earth System Research Laboratory and Developmental Testbed Center

Motivation: The NITE project was established by the Developmental Testbed Center (DTC) to investigate and design a software infrastructure to encourage R2O (Research to Operations) in NWP, which would facilitate development of NCEP numerical models by scientists both within and outside of EMC. Recent reports from UCACN (UCAR Community Advisory Committee for NCEP, 2014) and the DTC Science Advisory Board (2013) recommend the development of software and IT infrastructure for conducting model experiments. Infrastructure is also critical to the success of NOAA's NGGPS (Next Generation Global Prediction System).

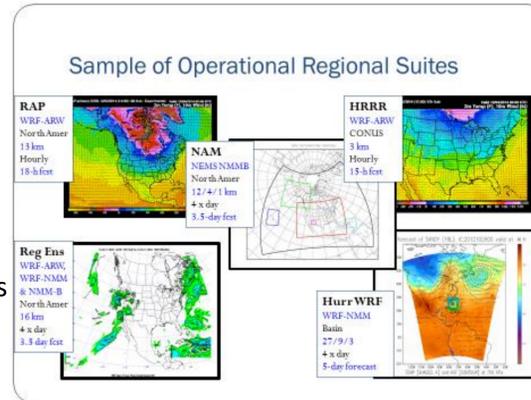
Background & Process

A survey of the potential NITE user community was conducted to understand the main problems in running experiments with NCEP operation systems.

- Surveyed potential users (EMC, NOAA Labs, NCAR, etc.)
- Identified key challenge areas

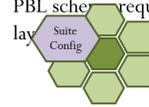
An assessment of selected existing NWP infrastructure systems was conducted, including documentation review, focus groups and site visits

- Held group discussions at NOAA ESRL, NCAR CESM, ESMF
- Visited NCEP, ECMWF, UKMO
- Reviewed some of the existing infrastructure and its documentation



Suite Configuration

- Configuration files for the overall suite
- Configuration files for each workflow component
- User-defined suites
- Predefined suites (CompSets), e.g.
 - NAM operational
 - NAM parallel
 - NAM for study X
- Compatibility checks, e.g. certain PBL scheme requires certain surface layer



Data Management

- Model data for initialization & boundaries
- Observations
- Experiment Outputs
- Standardized formats to allow plug-n-play with new components
- NOAA HPSS, MADIS, NOMADS
- Extensive metadata with browsing capability
- Smart data retrievals – subset, aggregate, plot



Documentation and Training

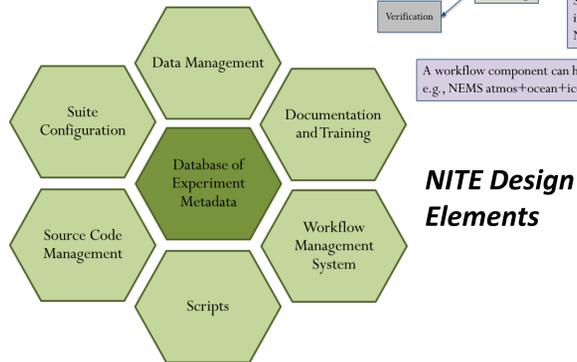
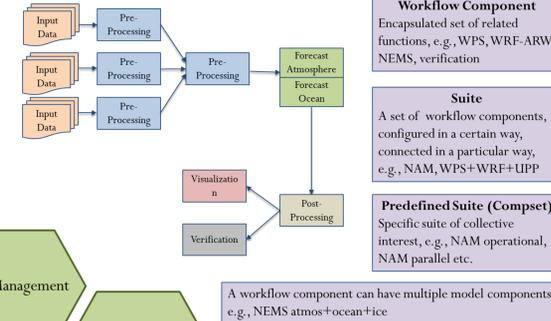
- Extremely important – NITE must make life easier
- Component documentation
 - Functionality, inputs/outputs, configuration, running
- Suite documentation
- Configuration of every component in a suite.
 - e.g. how is GSI employed in RAP
- Online documentation and training



NITE Design Goals

- Single system for NOAA and partners
- Datasets (inputs and outputs) are readily retrievable with well-documented format standards
- Predefined suites are available
- Customized suites can be created
- Past experiments can be browsed and reproduced
- Basic visualization and verification tools available for any suite
- Infrastructure is general and applicable to NCEP's various suites, including future ones
- NITE experiments are relevant for R2O at NCEP
- Easy to use and well documented
- Modular and allows incremental implementation and growth

Sample Workflow



EXPERIMENT METADATA DATABASE

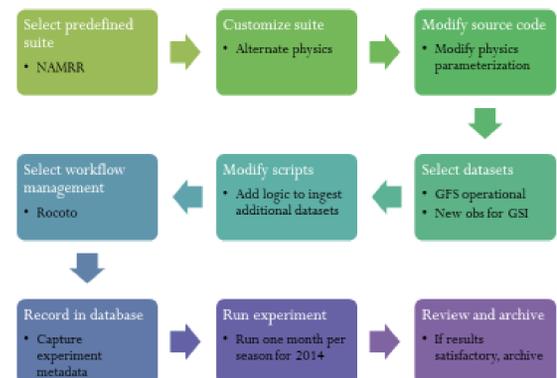
- Record provenance of codes, scripts, configuration files, inputs, computational platform, compilation options, job running options etc.
- Recommendation: the user edits configuration files and captures them in database before running
- Goal is **accountability, reproducibility, and relevance** to NOAA
- Database can be connected to a wiki where scientists record additional information, put links to papers, figures, reports etc.
- Database structure needs to be considered carefully so searches are fast



Recommended Phased Implementation

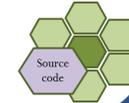
	Phase I	Phase II	Phase III
Analyses/Forecasts	GFS operational	GEFS operational	GFS reforecasts
Observations	GFS PrepBUFR	Raw observations and processing tool	Multi-year retrospective
Timeline	4 months retrospective plus case studies	Realtime data available in NOAA HPSS	Multi-year retrospective
Access	NOAA HPSS	NCAR HPSS	Data service
Formats	No change – various		Standardized
Level of processing	No change – postprocessed		Native
Metadata	Initial capability		Metadata database
Retrieval	No change – manual		Subsets/processed data
Archive Results	Postprocessed files to NOAA HPSS	NCAR HPSS	Data service
Computational Platform	Zeus and WCOSS	Yellowstone and other NOAA research	Generic
Workflow Management Systems	Rocoto and eFlow		
Predefined suites	NAM and NAMRR	NGGPS	Others (HWRP, RAP etc.)
Suite composition and scripts	All configuration abstracted to config files		
Experiment Metadata	Initial implementation	Completed relational database	Expanded search/query/summarize tools
Visualization and Verification	Initial capability	Fully implemented	
Documentation and training	Basic, for friendly users	Complete, including tutorials	Expanded

Example of NITE use



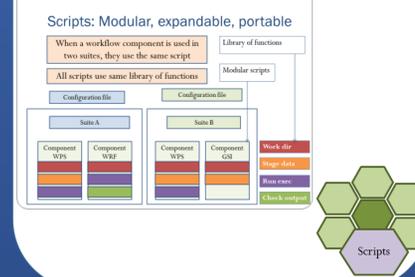
Source Code Management

- Access to source code repositories for components
 - Stable releases
 - Latest experimental code
- Develop and share code with colleagues
- Efficient, parallel build systems to compile all components of a suite



Scripts

- Execute workflow components
- Recommend script unification at NCEP
- Each workflow component has a script



Workflow Management Systems

- Automate running suites
- Inefficient to run scripts by hand
- Check dependencies and status
- Adds fault tolerance
- Examples:
 - Rocoto
 - ECFLOW

