SCEC
an NSF+USGS center

Design, Implementation, and Status of the CSEP Testing Center at SCEC

Southern California Earthquake Center

AGU, December 12, 2007

GNS SCIENCE TE PŪ AO

W. M. Keck Foundation

Schweizerischer Erdbebendienst
Swiss Seismological Service

Collaboratory for the Study of Earthquake Predictability (CSEP)

- Goals:

1. Reduce the controversies through a collaboratory infrastructure that can support a wide range of scientific prediction experiments
2. Promote rigorous research on earthquake predictability through the SCEC program and its global partnerships
3. Help government agencies assess the feasibility of earthquake prediction and the performance of proposed prediction algorithms

CSEP System Requirements

- Earthquake Testing Center goals (as outlined by Schorlemmer and Gerstenberger (2007)) :
  - Controlled Environment
  - Transparency
  - Comparability
  - Reproducibility

- A Testing Center architecture should strive to meet these non-functional requirements to ensure that users trust the results.

System Design Principles and Approach

CSEP Testing Center is designed to meet several requirements:

- Must run for years to get statistically significant results
- Must be easy to modify without significant re-implementation or down time.
- Must be inexpensive to maintain and operate
- Must be able to “show our work”, that is demonstrate how we produced a particular result.

System Design Principles and Approach

Basic CSEP system design and development principles:

- Use simple design approaches and only add features when they are needed.
- Use existing software wherever possible.
- Utilize open-source software whenever possible to keep transparency and low-cost.
- Put the system under test to ensure reliability and to make the system easier to change.
- Open the CSEP testing center codes as open source, and provide communication tools (email-lists, wiki, version control, test suites) to groups interested in the details of the system.

CSEP Testing Center Use Cases

Existing CSEP Use Cases:

(3) Evaluate seismicity-based daily forecasts by filtering catalog, running forecast model, evaluating forecast, and plotting evaluation results.

(2) Evaluate seismicity-based 5-year forecasts on a monthly basis, by filtering catalog, scaling forecasts, evaluating forecast, and plotting evaluation results.

(3) Reproduce a given forecast and forecast evaluation from system archives

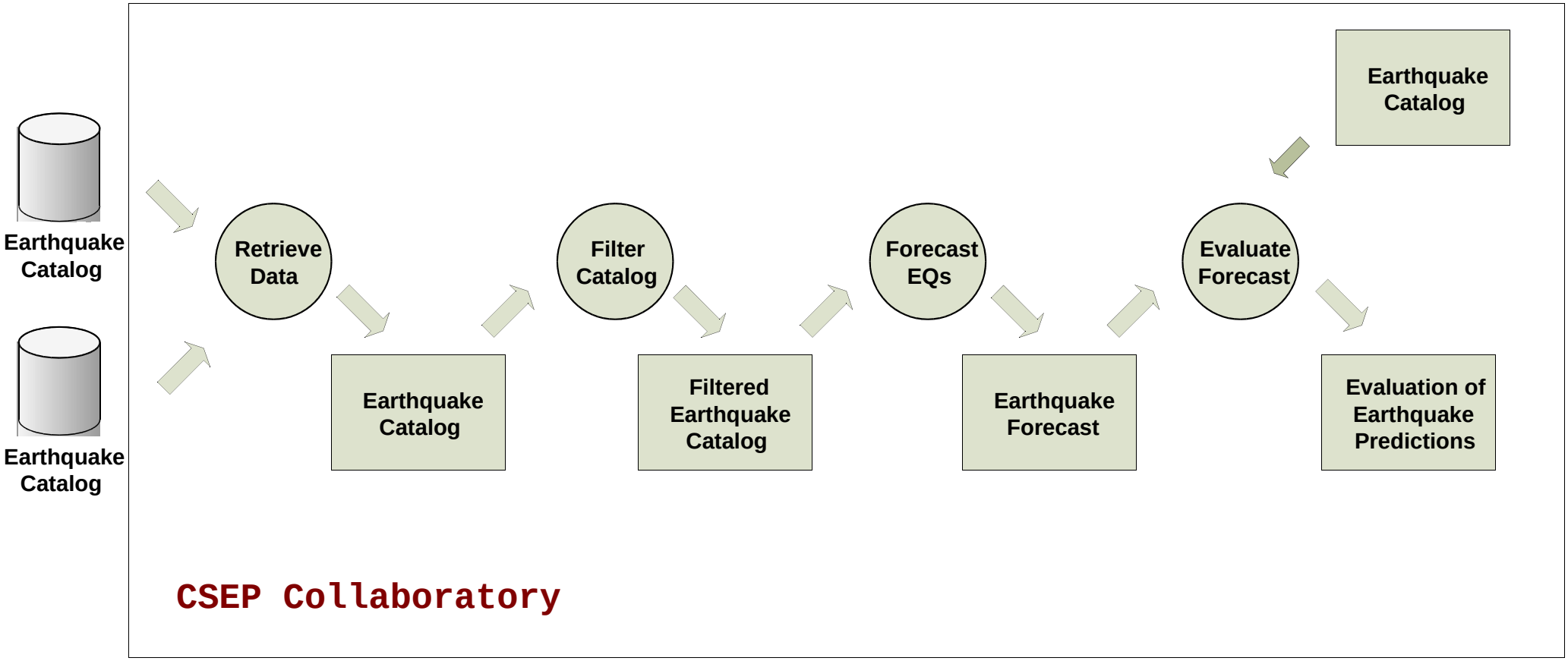
(4) Allow a model developer to run her forecast and determine how it would perform in the CSEP system, without entering it into the CSEP system.

Describing the CSEP Architecture

- Basic CSEP architecture is a file-based workflow system architecture
  - A workflow is a set of tasks that have data dependencies between them.
  - Currently CSEP tasks run serially.

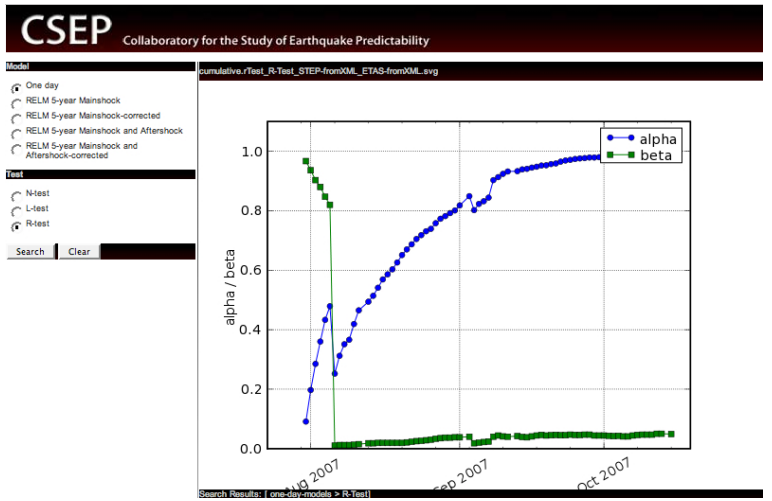
- Functionally, CSEP has three main elements:
  - Components (executable programs in the workflows)
    - Functional Tasks
  - Data Sources and Stores (Input and outputs)
    - Interfaces between programs
  - Dispatcher (Executes the components in the correct order)
    - Program Control

Conceptual CSEP Processing Model For Seismicity Based Forecasts (Use cases 1 & 2)

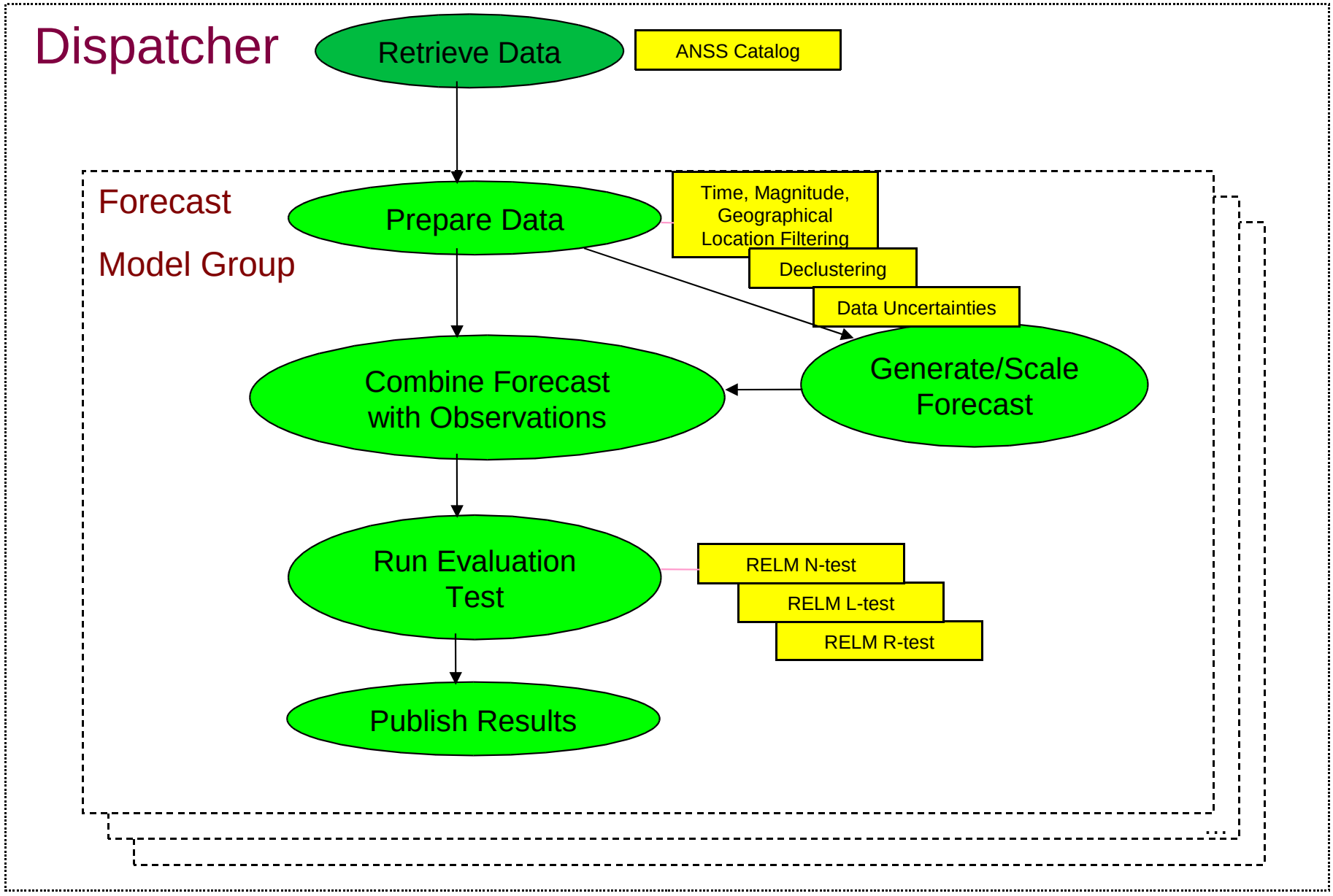


CSEP Software

- Must perform the following tasks:
  - Retrieve and prepare catalog data
  - Invoke forecast models
  - Evaluate forecasts
  - Publish evaluation results
- "Dispatcher"
  - Top-level class to automate these tasks
  - Focuses on
    - End-to-end processing
    - Reproducibility of any forecast experiment

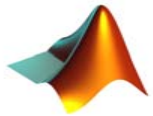


CSEP End-to-End Processing



Standardized Software Stack

- Software Version Control System
  - Open-source Subversion (SVN)
    - Tagging of CSEP releases
- System Configuration
  - Standardized software stack
    - For the testing center
    - For installed models




Physical View of CSEP System

Operational Computer:

- Dispatcher runs “official” forecasts and evaluations
- Saves data archives for reproducing results

Integration Computer:

- Configured to match “official” computer as closely as possible with latest software and models. Acceptance tests are run on nightly basis.

Development Computer:

- Modelers run their codes in their own accounts using CSEP Software Stack
- Modeler tools including version control runs on this computer

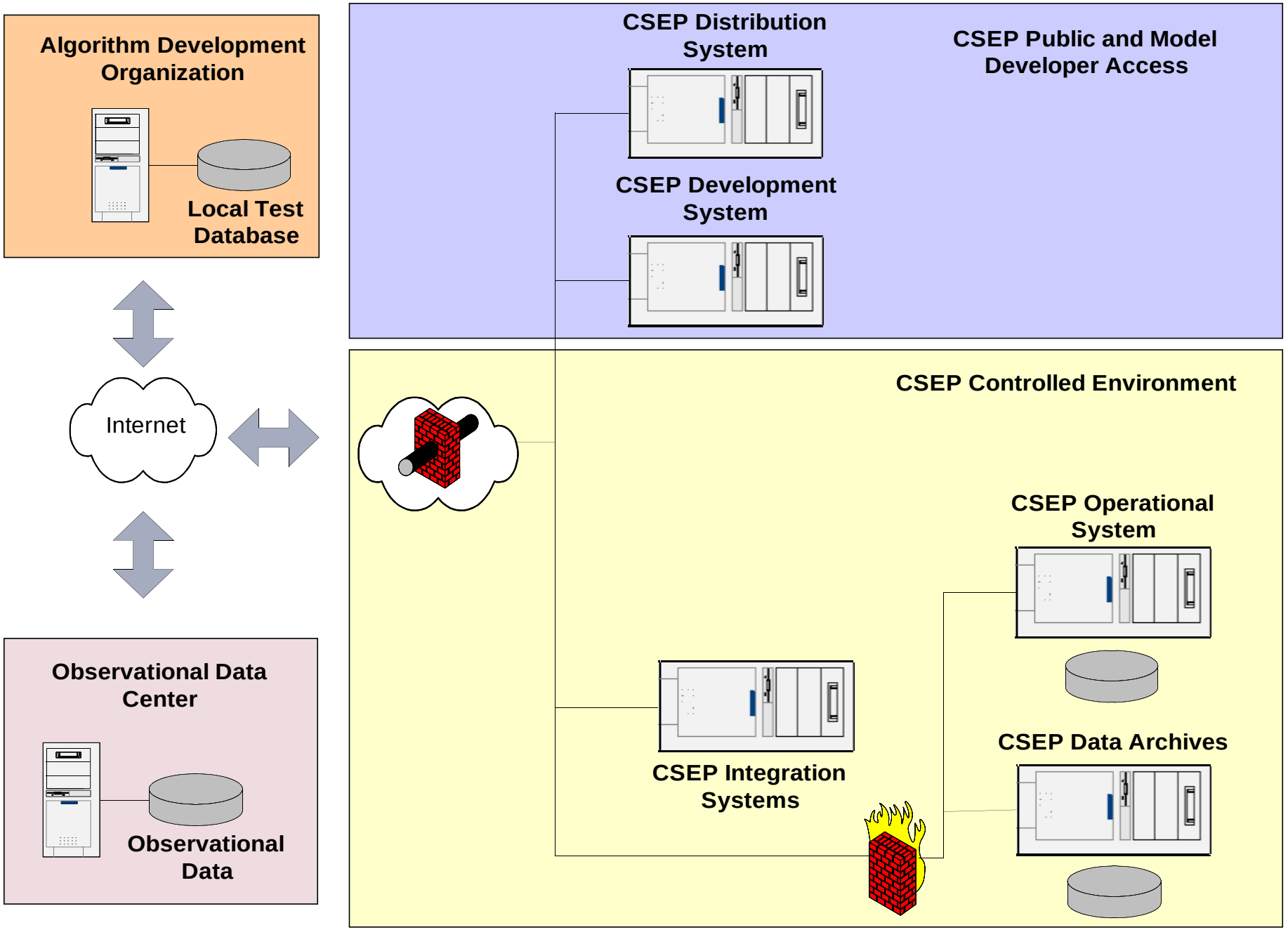
Distribution Computer:

- CSEP Forecast evaluation results posted here on web site. Testing Center Web site hosted here.

Data Archives:

- Second copy of reproducible data sets needed

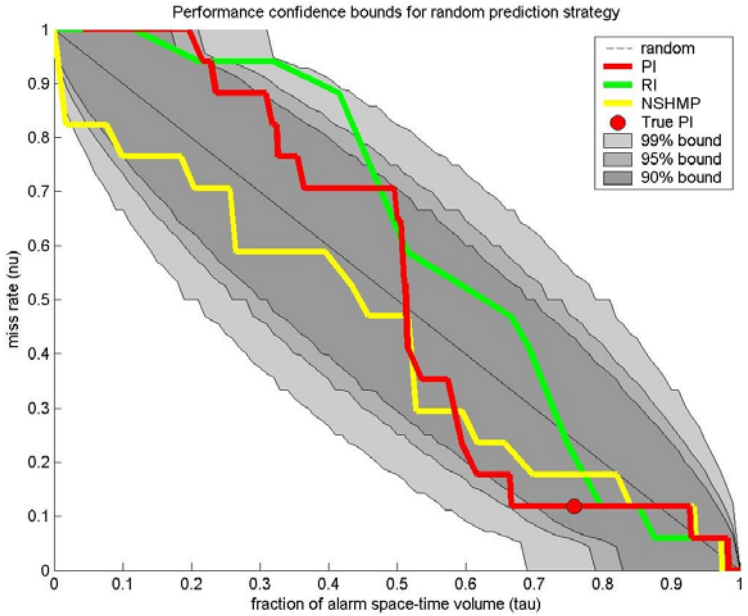
CSEP Information Technology (IT) Infrastructure



CSEP Development View

Primary Development Categories relate to “who’s doing the development” and “how distribution is done”

- SCEC CSEP Development Staff
  - Framework (Dispatcher) development
  - Web Site development
- Collaborative Developments with ETH
  - Forecast Template Development
  - QuakeML Catalog representation
  - QuakePy Catalog Processing Tools form
- Re-use of RELM Software Components
  - Catalog Processing and Evaluation Tests
- Forecast Model Development
  - Done by variety of Modeling Groups



CSEP Testing Center General Development

General Development Goals:

- Each Quarter add one or two new forecasts under-evaluation
- Reduction of data storage required to operate
- Replacement of proprietary software in CSEP developed components (e.g. evaluation tests, filtering tests)
- Wider adoption of Standard data formats (e.g. QuakeML catalog representation)
- Adoption of advancements in available processing tools (e.g. QuakePy)
- Improvements delivery of data products to scientists
  - More product that are useful to scientists
  - Existing products are easier to access
  - Products that are easier to interpret and understand

End

Please see www.cseptesting.org