

## Abstract

The Collaboratory for the Study of Earthquake Predictability (CSEP) is developing the infrastructure for facilities to conduct earthquake forecast experiments. It provides controlled integration environment with standardized software stack for developing and installing of forecast experiments. The processing infrastructure has to allow for rapid computations using distributed computing facilities but also needs to run on desktop computers for research activities. Module design of the CSEP software focuses on reproducibility of any forecast experiment. Furthermore, program codes need to be validated and distributed to other than Southern California Earthquake Center (SCEC) testing facilities. We will discuss the design challenges and present the software concept, development strategies, ways for participating in development, and the flexibility for customizing of our open-source software.

## CSEP Objectives

- Establish rigorous procedures for registering and evaluating prediction experiments.
- Construct community standards and protocols for comparative testing of predictions.
- Develop an infrastructure that allows groups of researchers to participate in prediction experiments.
- Provide access to authorized data sets and monitoring products for calibrating and testing prediction algorithms.
- Accommodate experiments involving fault systems in different geographic and tectonic environments.

## CSEP Goals

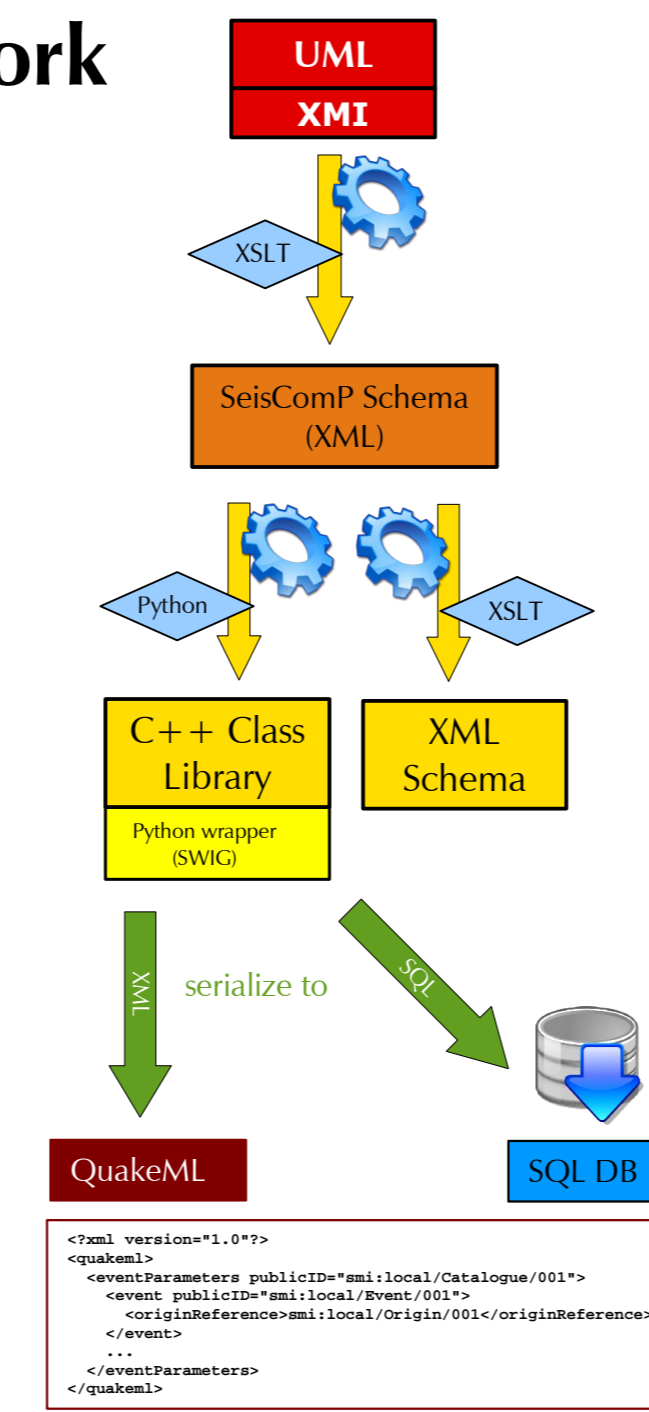
- Reduce the controversy surrounding earthquake prediction through a collaboratory infrastructure to support a wide range of scientific prediction experiments.
- Promote rigorous research on earthquake predictability through the SCEC program and its global partnerships.
- Help the responsible government agencies assess the feasibility of earthquake prediction and the performance of proposed prediction algorithms.

## Code Generation Framework

XML data models available as UML class diagram (and XMI)

Automated code generation for:

- Native C++ Library
- Native Python Library
- SQL database scripts
- Serialization



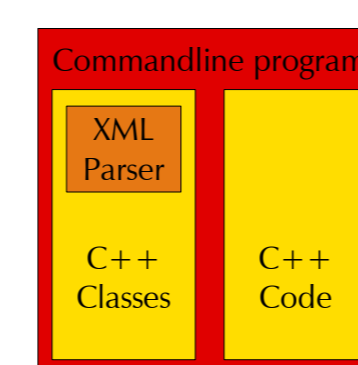
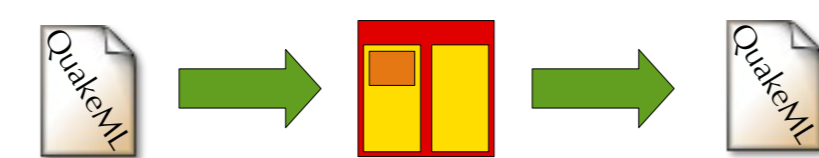
## Software Modular Design

Each major task will be wrapped into a command-line tool reading and writing (pipe) QuakeML or other XML-definitions (forecast, test results). All tools will constitute the CSEP toolkit to be used by researchers or testing centers.

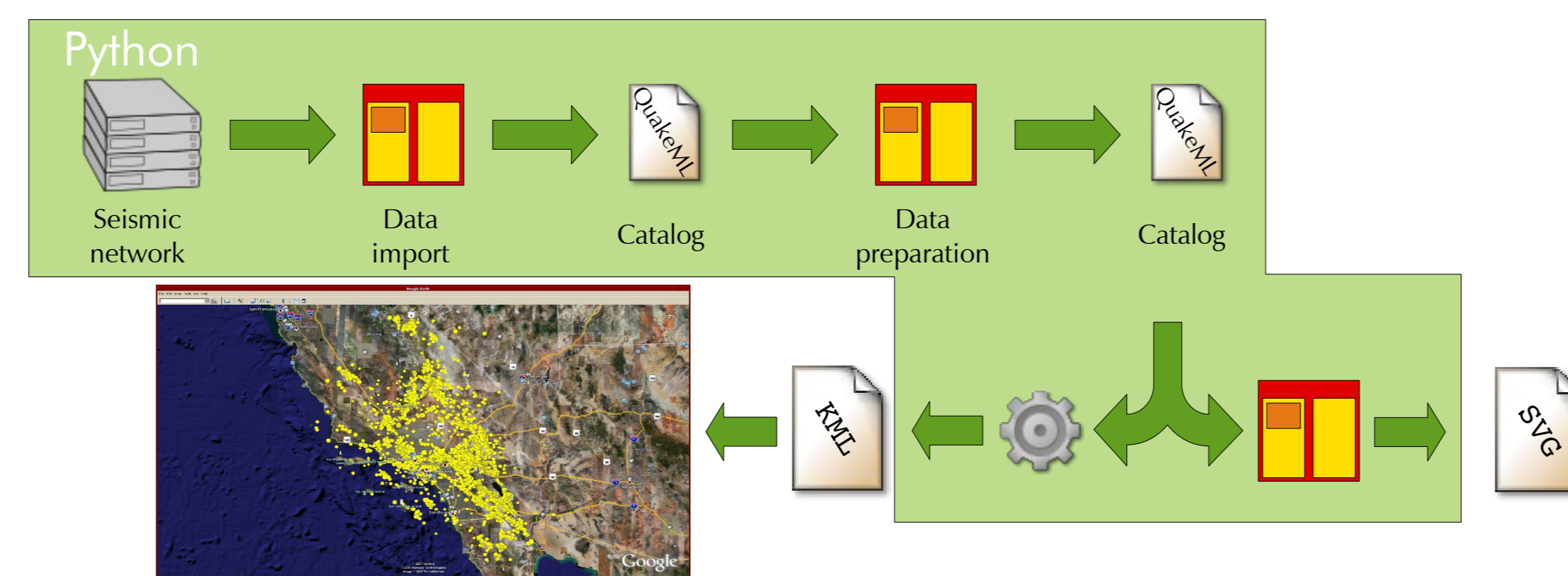
- Interface formats (XML-based)



- CSEP Core (Toolkit C++)



- Natural laboratory testing class business logic (Python)



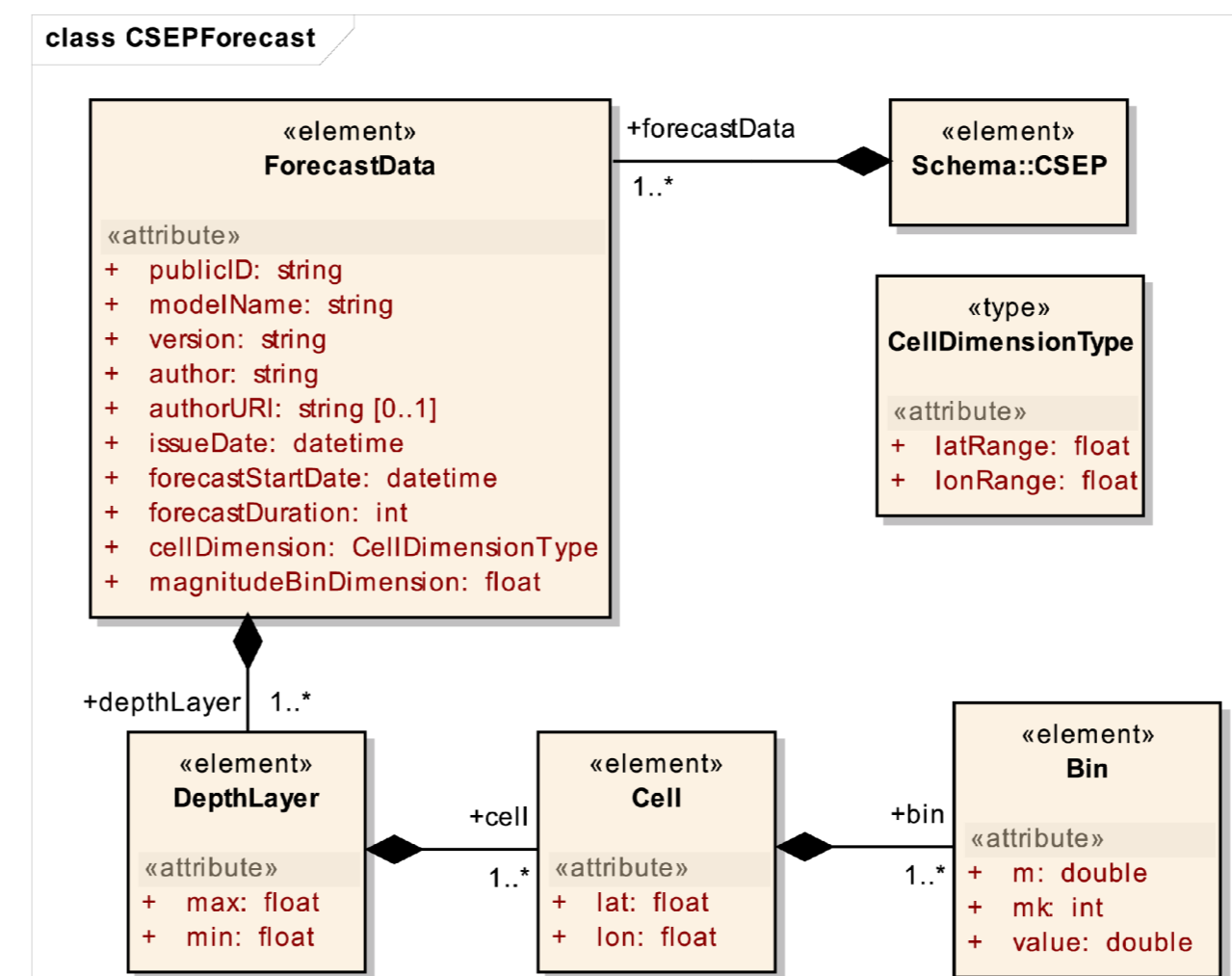
## Implementing the Business Logic

The business logic is being implemented in Python by combining the command-line tools. "Factory" software pattern is used to register custom Python modules that represent forecast models and catalog filtering specific to the Natural Laboratory (NL). This offers the necessary flexibility for customizing the CSEP software for different testing centers or use by researchers.

## XML-Definitions

- Earthquake Catalog Format
- QuakeML
- CSEP XML Formats
- Test Result
- Forecast (OpenSHA)

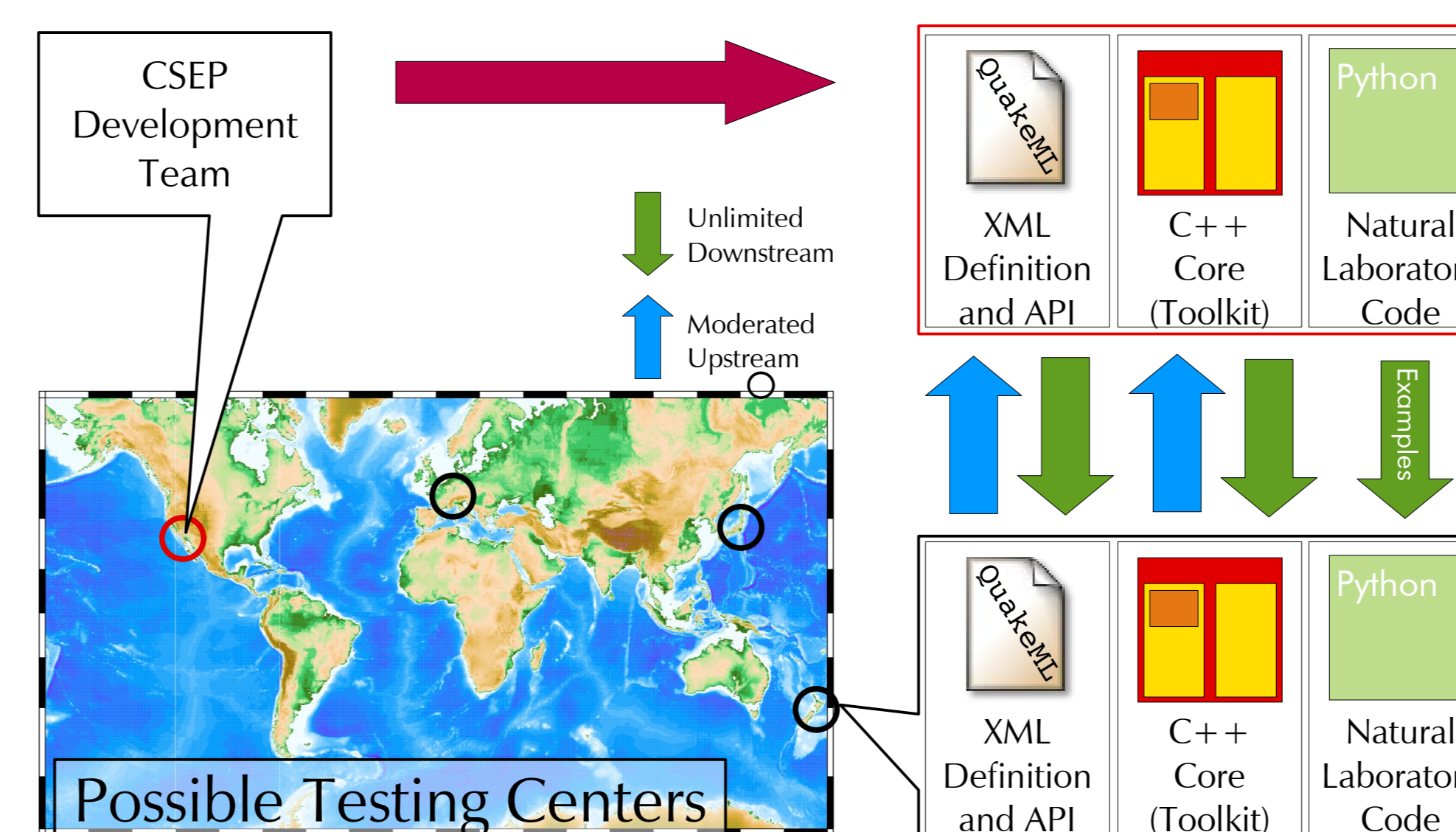
```
<?xml version="1.0"?>
<CSEPForecast xmlns="http://www.scec.org/xml/ns/csep/forecast/0.1">
  <ForecastData publicID="smi//org.scec/csep/forecast/1">
    <modelName>Kate Moss</modelName>
    <version>0.1</version>
    <author>Danijel Schorlemmer</author>
    <authorURI>smi://org.scec/persons/schorlemmer_danijel</authorURI>
    <issueDate>2007-03-15T16:00:00</issueDate>
    <forecastStartDate>2007-03-16T16:00:00</forecastStartDate>
    <forecastDuration>365</forecastDuration>
    <cellDimension lonRange="0.1" latRange="0.1" />
  </ForecastData>
  <depthLayer min="0" max="30">
    <cell lon="122" lat="34.1">
      <bin m="4.1" mk="1">0.0004</bin>
      <bin m="4.2" mk="1">0.0002</bin>
    </cell>
    <cell lon="122" lat="35.1">
      <bin m="4.1" mk="1">0.000432</bin>
      <bin m="4.2" mk="1">0.000433</bin>
    </cell>
  </depthLayer>
</CSEPForecast>
```



## Development organization

CSEP Software will be released under open-source licenses and distributed to regional testing centers. We will host the XML-definitions, the software core, and the scripts for Natural Laboratories (NL) hosted in our center. We provide unlimited downstream and moderated upstream of codes.

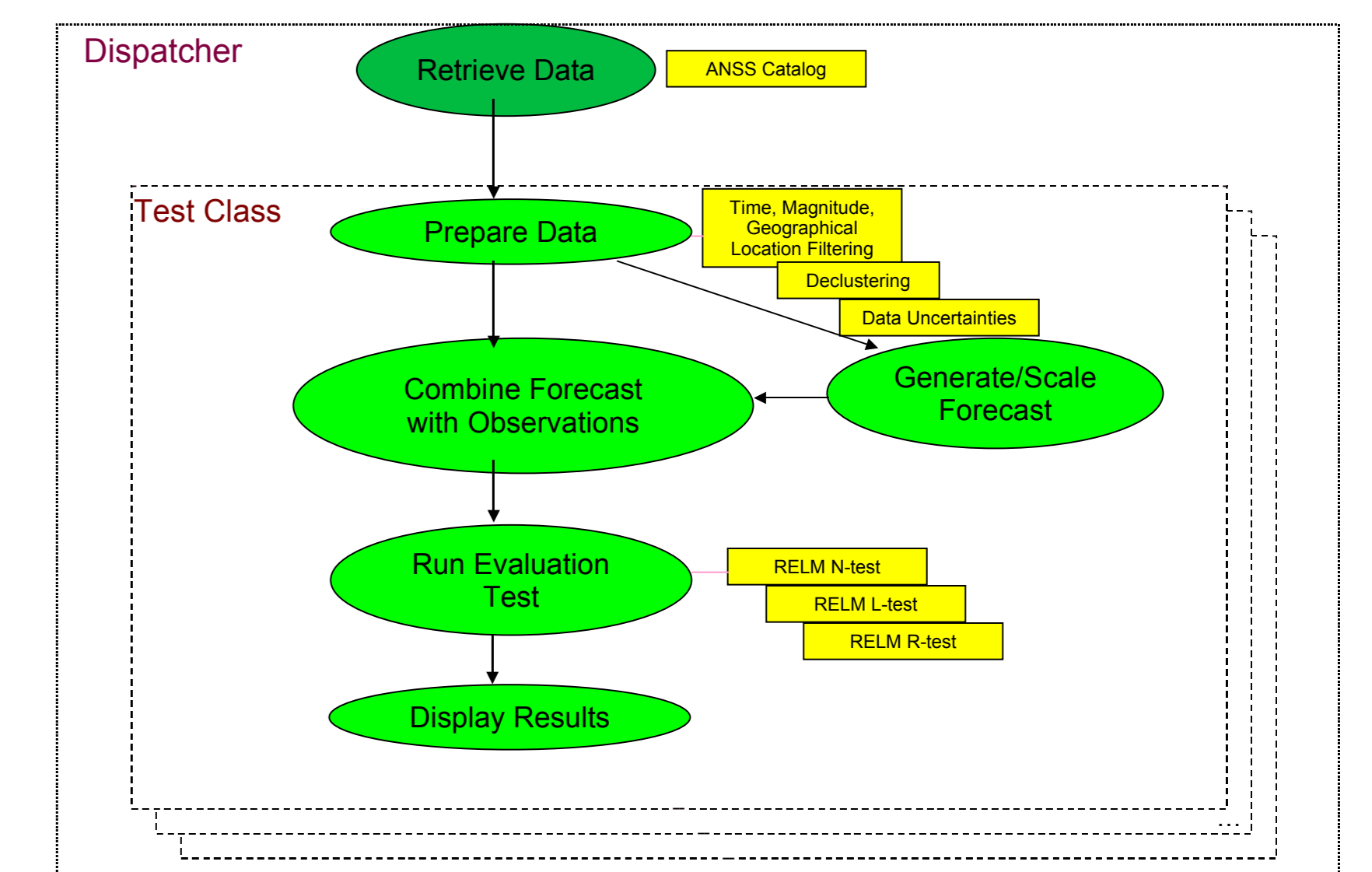
## Software distribution



## CSEP Version 1.0

- Full Reproducibility - the Testing Center will keep:
- All input data (earthquake catalogs)
- All simulations (random numbers)
- All results
- System and software configurations used for computations (metadata)
- End-to-End Processing

## CSEP V1.0 – End-to-end Data Flow



## Software Stack

We ask the modelers to use open source software only. Ask us for including other packages. Current list:

- Linux
- MPICH2
- GCC, G77/gfortran
- Python
- R (Matlab)
- Subversion
- Trac
- Eclipse

## System infrastructure

