SnT2017 T3.5-P46

Disclaimer

The views expressed on this poster are those of the author and do not necessarily reflect the view of the CTBTO

The International Data Centre Re-engineering Project, Phase 2: An Overview

H. Breitenfellner[1], D. Brown[2], J. Burns[2], D. Carr[2], J. Coram[2], B. Hamlet[2], J. Harris[2], R. Le Bras[1], P. Mialle[1], E. Peiszer[1], R. Prescott[2], E. Tomuta[1], and C. Young[2]

[1]CTBTO, [2]Sandia National Laboratories

Project Business Case and Goals

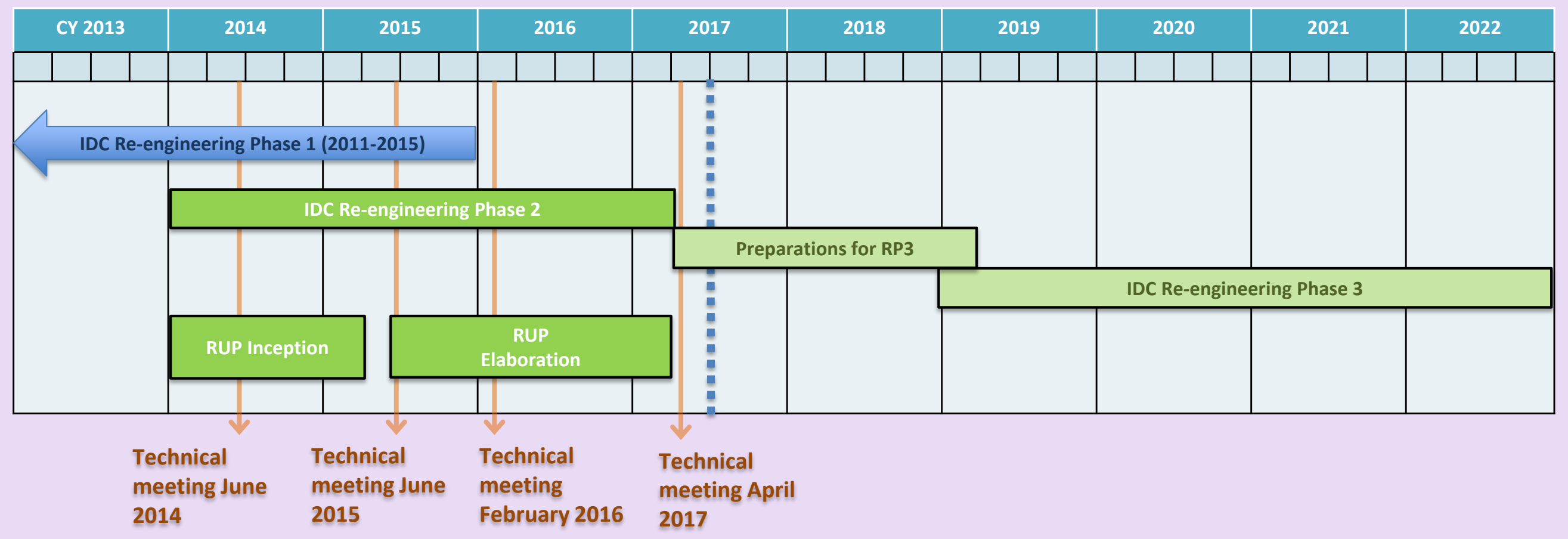
In the last 16 years, since the delivery of the first version of the SHI acquisition, processing and product dissemination software by the Prototype International Data Centre (PIDC), major components of the system have been replaced in response to advances in monitoring technologies, leading to new functional requirements and infrastructure changes. In the absence of an up-to-date overarching architecture, the result of these development activities is an increasingly fragmented software landscape with little software reuse, code duplication and outdated technologies. Such a system is increasingly difficult to maintain and enhance as new technologies become available.

IDC Re-engineering Phase 2 (abbreviated ‘RP2’ in this poster) developed an overarching, model-based component architecture for SHI software, as the foundation for a cost-effective, maintainable and extensible system that will allow the CTBTO to meet its treaty monitoring requirements for the next 20+ years.

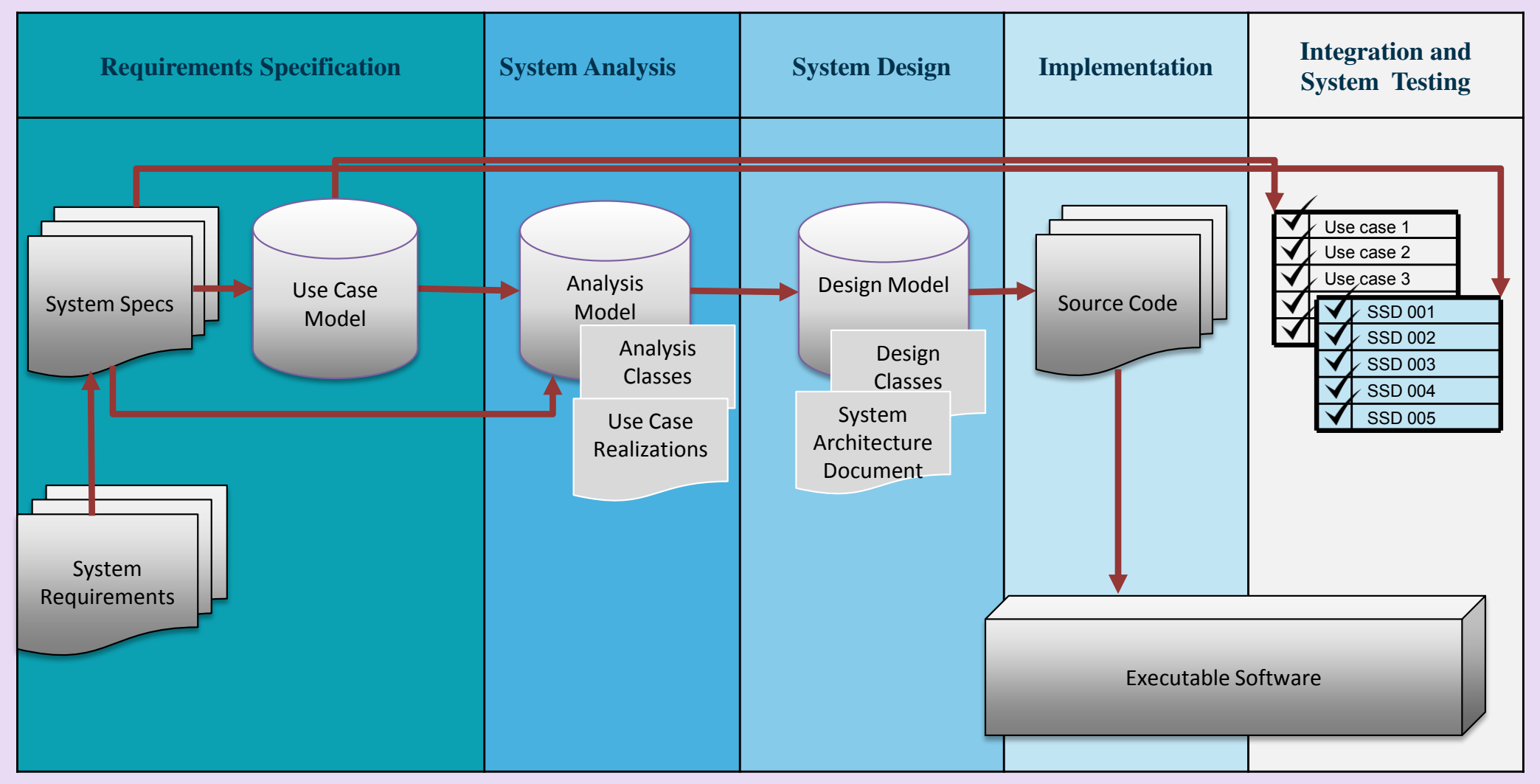
The goals of the RP2 system architecture were to:
1. Unify seismic/hydroacoustic/infrasonic (SHI) processing software used in all processing stages and provide a modern approach to integrating new components.
2. Provide a basis for enhancing and extending SHI interactive software with a modern UI design, improved security and support for remote analysts.
3. Ensure reproducibility of results of SHI analysis.
4. Add new data and product acquisition and distribution mechanisms following latest industry standards.
5. Provide a basis for an IT Disaster Recovery System in accordance to CTBTO Business Continuity requirements. (A ‘disaster’ is any situation that interrupts the routine operation of the processing system to the extent that the system hardware and software can no longer adequately function for a sustained period of time on the order of several days or longer).
6. Provide a training platform to integrate new staff.
7. Integrate new data sources, including data from Contributing National Facilities, National Technical Means, and other openly-available SHI data and meteorological data required to analyze infrasound signals.
8. Enable the IDC to efficiently provide special event analysis at the request of the Member States.

RP2 Status, Design and Development Methodology

The RP2 project ran from January 2014 to April 2017, covering the Inception and Elaboration phases of the Rational Unified Process (RUP) software development methodology. The remaining RUP phases of Construction (software development) and Transition (software implementation/deployment) are planned to be executed in a third phase of IDC re-engineering (RP3) to be initiated in 2019, and with a duration of about 4 years.
RP2 leveraged requirements and design artefacts created in a project executed by the U.S. National Data Centre (US NDC) and aiming to modernize the US NDC software. RP2 deliverables were reviewed by experts from Member States in a series of technical meetings held once a year from 2014 to 2017.



One of the central principles of RUP is to enable traceability of software components, through design artefacts, back to requirements. The types of requirements and design artefacts created during RP2 with this principle in mind are:
System Requirements Document (SRD): Declarative statements describing system requirements.
System Specification Document (SSD): similar in form to the SRD, but more complete and formally correct.
Use Case Model Survey (UCMS): A high-level description of all use cases, with architecturally significant use cases identified.
Use Case Model: Full use-case descriptions, starting with architecturally-significant ones.
User Interface Storyboards (UIS): User interface mock-ups for use cases that require a graphical user interface.
Data Model: Abstract (in RUP terminology “analysis”) classes that identify major objects in the system, some of their attributes and their relationships.
Use Case Realizations: Analysis classes and how they interact to enable the system to support the functionality specified in the use cases.
Component Interface Specifications: Specifications of the interfaces of major components in the system.
Architecture Document: Captures the general principles used throughout the design of the system, serves as a guide for further development.
More details on the data model, component interface specifications and the architecture document are in the companion poster “IDC Re-engineering Software Architecture and Data Model”.



The figure above shows the RUP disciplines each deliverable is associated to (requirements specification, system analysis, etc.), as well as, in red, the main dependencies between artefacts. These enable backward traceability from executable software, all the way to requirements.

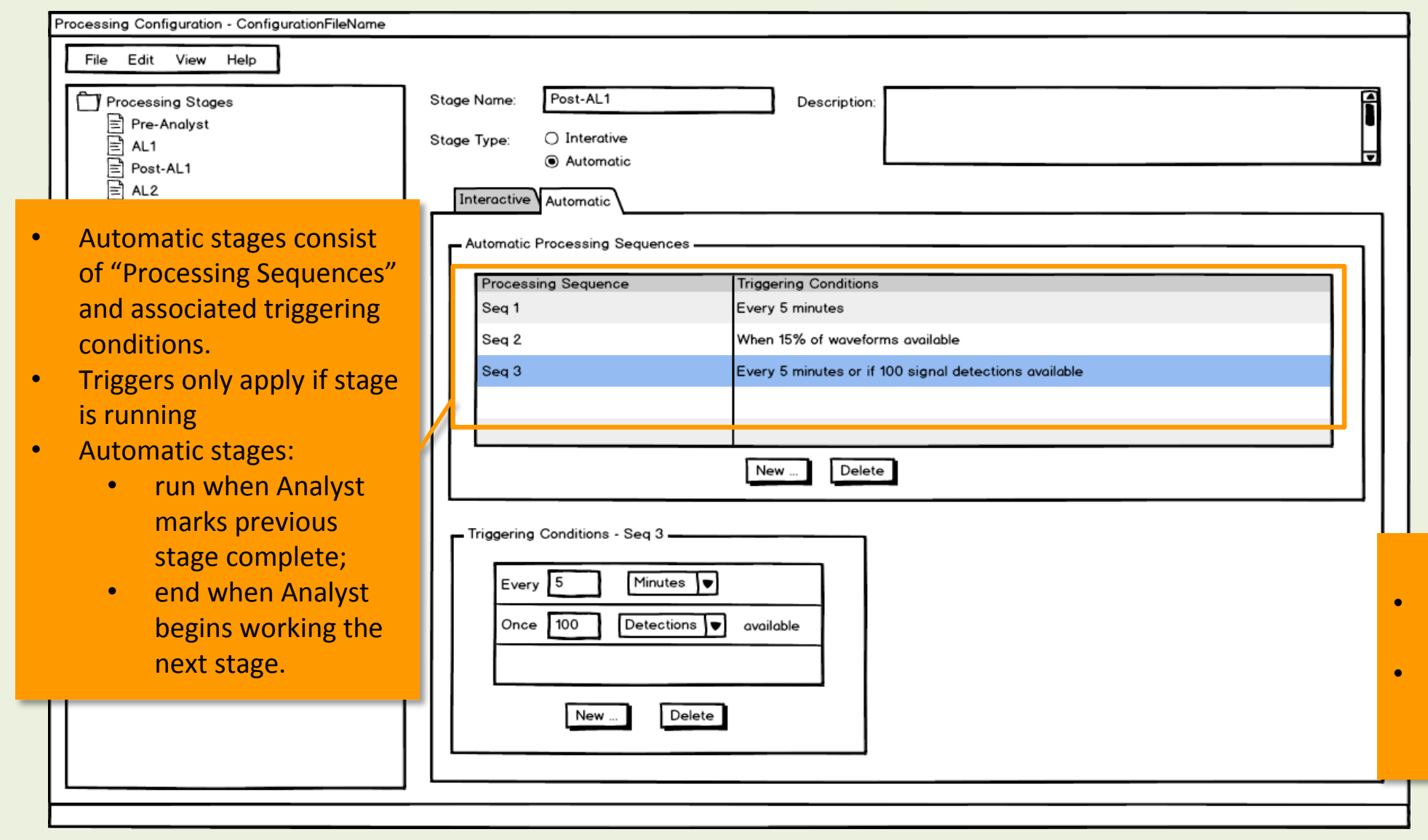
Key Improvements in the RP2 Architecture

The RP2 architecture brings major improvements in several areas of IDC processing of SHI data.

1. The design of analyst tools preserves the user interface concept analysts are familiar with, while enhancing specific aspects such as: UI flexibility, analyst review workflow, event management, event cross-correlation & comparison, map tool functionality and integration, waveform quality control masks, frequency-wavenumber displays, support for analyst training. For more details, see the companion poster “IDC Re-engineering Phase 2 (RP2): Analysis Interface Improvements”.
2. A componentized design based on documented component interface specifications facilitates a new model for collaborative software development following best practices in open source software development.
3. The architecture supports comprehensive capturing of data provenance allowing to understand how processing results were arrived at and to investigate the evolution of a result as the available information changes. Functions like the ability to view the history of an event, presented in the companion poster on analysis interface improvements and available to both internal and external system users build upon this functionality.
4. Support for Disaster Recovery by switching processing to a backup site in near-real time (hot backup) is built-in.
5. Capabilities for flexible pipeline configuration are supported by graphical tools.
6. Enhanced monitoring and testing capabilities are provided through test data set replays.
7. Extensibility and modularity are major features built in all components.

Key Processing Improvements: Flexible Pipeline Architecture

The re-engineered software will allow System Maintainers to define Processing Configurations, which include Processing Stages, Processing Sequences and Analyst Activities. At the highest level, a Processing Stage is associated with a Processing Sequence. The System Maintainer defines a Processing Sequence to be executed, which includes other Processing Sequences and/or Processing Steps. The System Maintainer defines Analyst Activities which are associated with Processing Stages. Once the System verifies that the configuration is valid, the System Maintainer saves all changes made to the Processing Sequence. The configuration changes take effect only after they are installed as a software update.
A graphical user interface, illustrated through mock-ups shown below, supports the System Maintainer in defining a Processing Configuration.

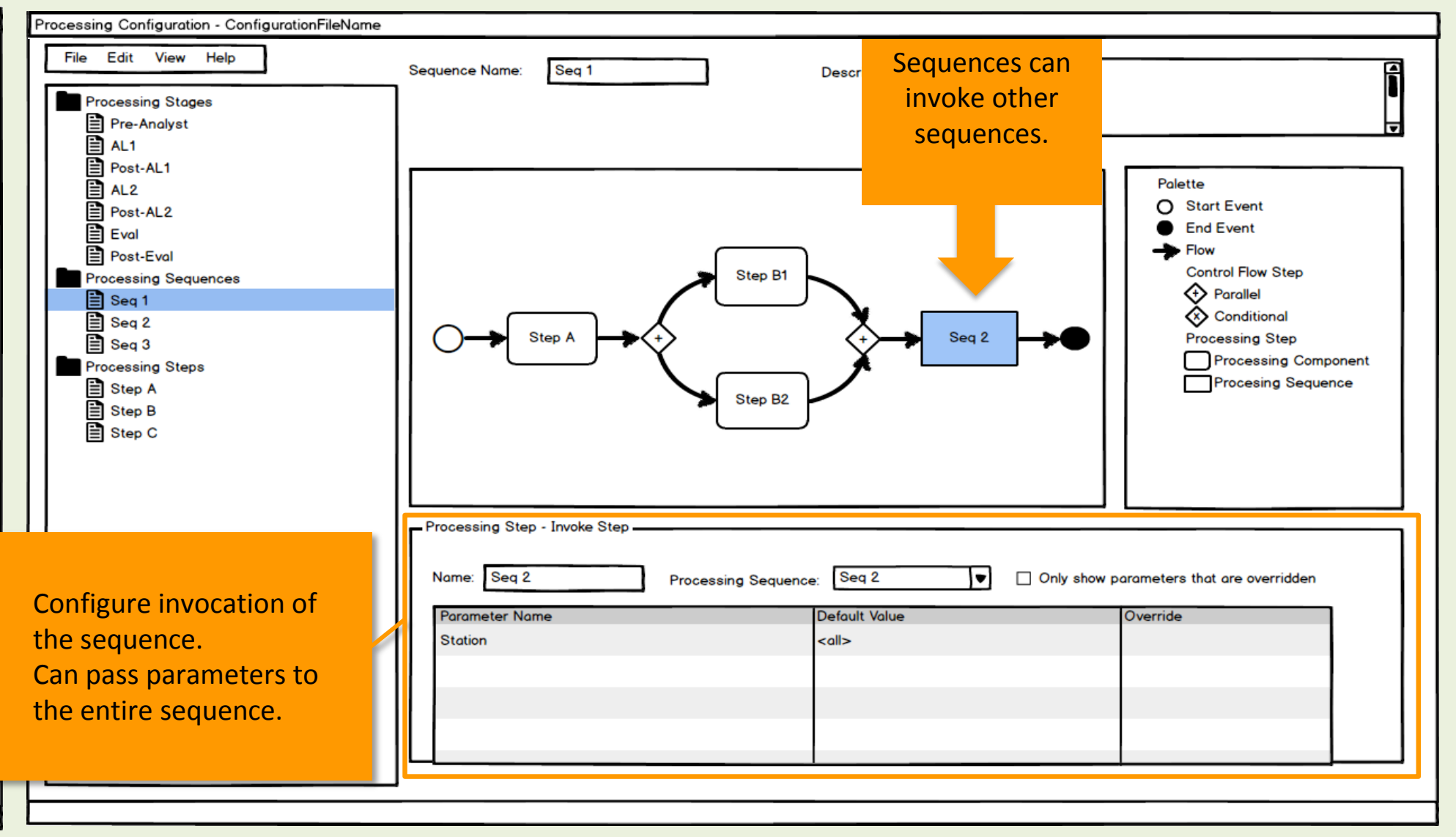


Stage Name: Post-AL1
Description:
Stage Type: ○ Interative ● Automatic

Processing Sequence	Triggering Conditions
Seq 1	Every 5 minutes
Seq 2	When 15% of waveforms available
Seq 3	Every 5 minutes or if 100 signal detections available

Every 5 Minutes
Once 100 Detections available

- Automatic stages consist of “Processing Sequences” and associated triggering conditions.
- Triggers only apply if stage is running
- Automatic stages:
  - run when Analyst marks previous stage complete;
  - end when Analyst begins working the next stage.



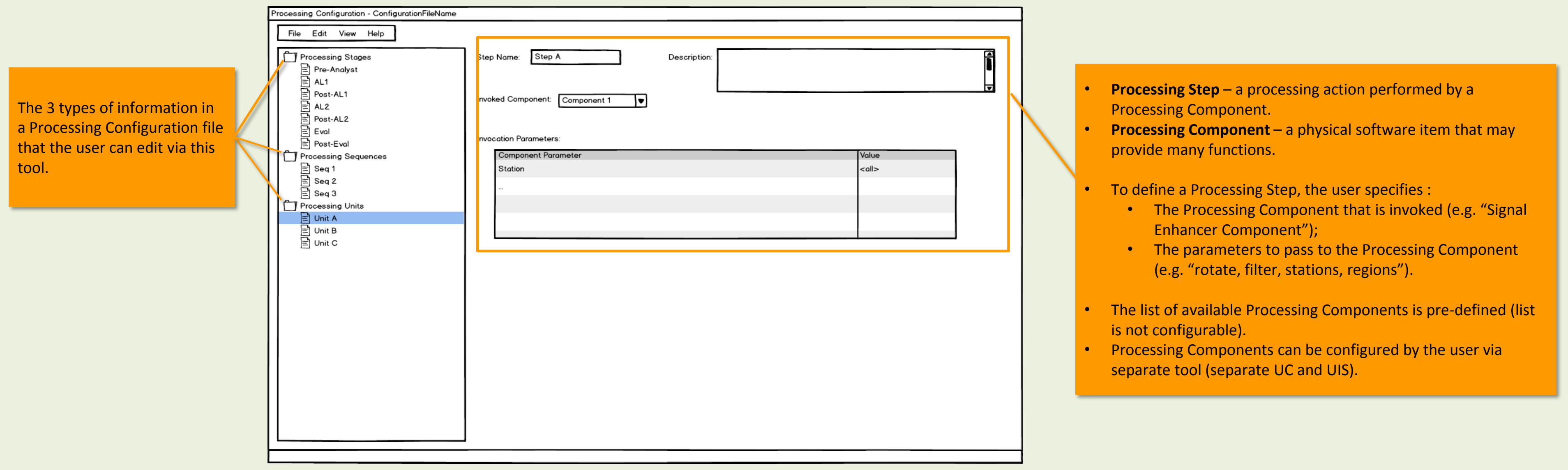
Sequence Name: Seq 1

Sequences can invoke other sequences.

Name: Seq 2
Processing Sequence: Seq 2
☐ Only show parameters that are overridden

Parameter Name	Default Value	Override
Station	<all>	

- Configure invocation of the sequence.
- Can pass parameters to the entire sequence.



Step Name: Step A
Description:
Invoked Component: Component 1
Invocation Parameters:

Component Parameter	Value
Station	<all>
...	

The 3 types of information in a Processing Configuration file that the user can edit via this tool.

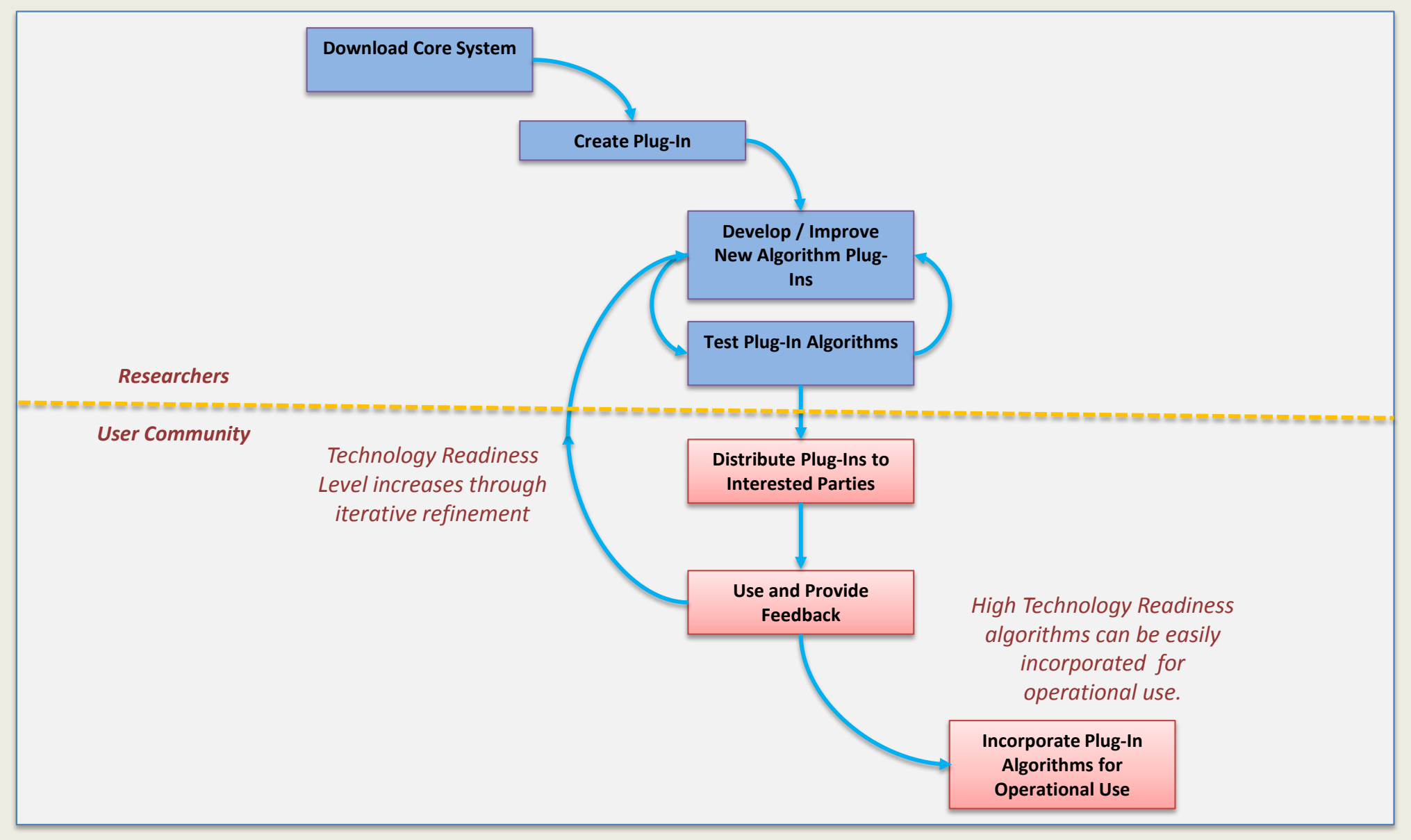
- **Processing Step** – a processing action performed by a Processing Component.
- **Processing Component** – a physical software item that may provide many functions.

- To define a Processing Step, the user specifies :
  - The Processing Component that is invoked (e.g. “Signal Enhancer Component”);
  - The parameters to pass to the Processing Component (e.g. “rotate, filter, stations, regions”).

- The list of available Processing Components is pre-defined (list is not configurable).
- Processing Components can be configured by the user via separate tool (separate UC and UIS).

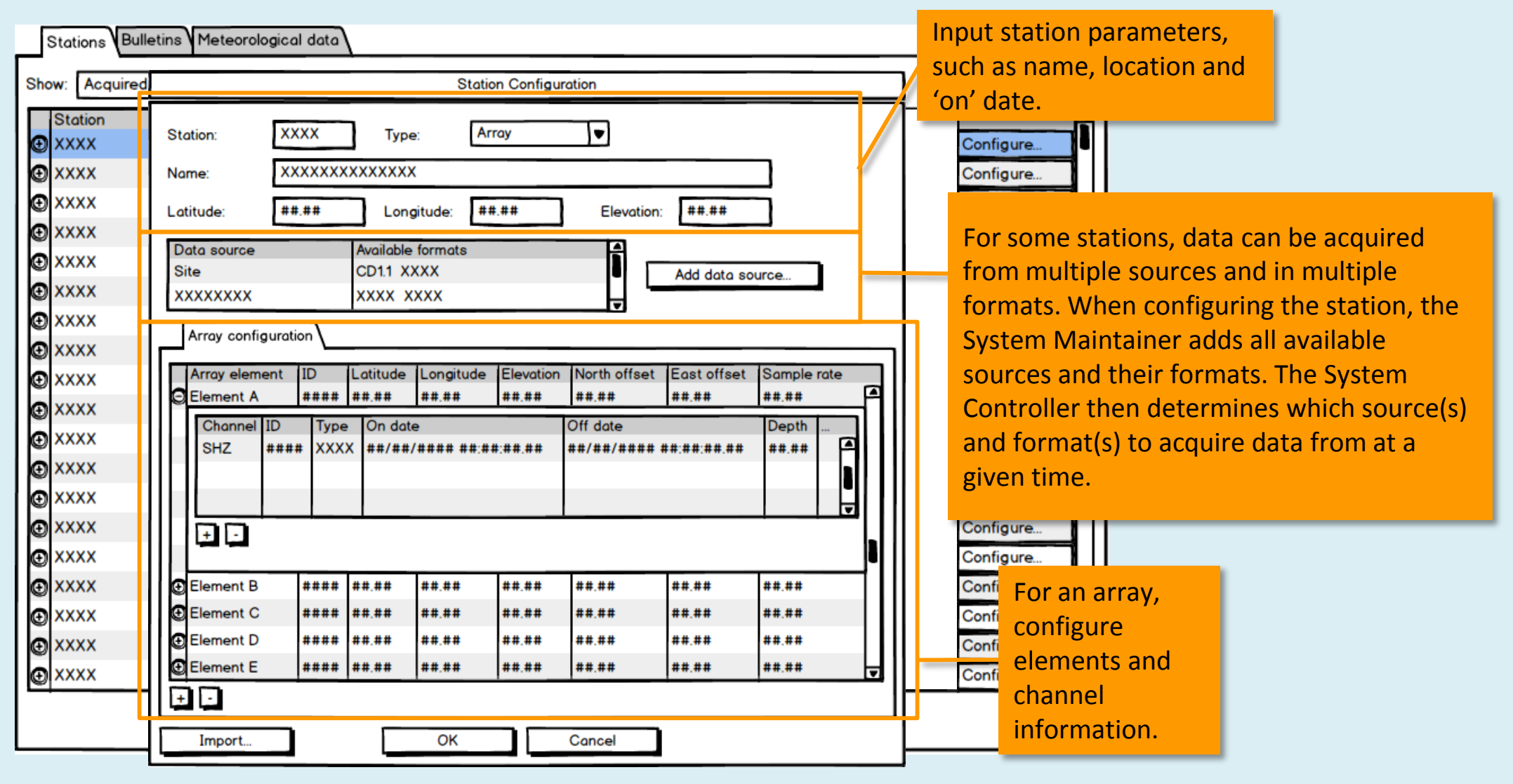
Key Improvements: Facilitating a Collaborative Software Development Model for the NDC Community

Software resulting from the re-engineering project is intended to be made available to NDCs under a software license compliant with treaty requirements. This software will expand the existing suite of tools the CTBTO provides to NDCs under the NDC-in-a-Box distribution and will integrate with these tools through data and product format conversion capabilities.
In addition, the existence of Component Interface Specifications should facilitate software contributions of individual processing components by NDCs, and even support collaborative software development models within the NDC community, such as those used in open-source projects and shown in the diagram at the right.
Such open source collaboration models (codified by sites such as GitHub) support making software freely available for others to use and modify, while still maintaining centralized control of the definitive copy.
A componentized architecture based on specified interfaces also offers a practical way forward toward fulfilling the treaty requirement (as specified in article 22.b) that the IDC integrate computer algorithms provided by Member States into its processing and that the resulting products be made available as products of the respective State Party.



Key Processing Improvements: Graphical Tools for Configuring SHI Data Acquisition and Processing

Maintaining a correct and consistent configuration of SHI data acquisition and processing is essential for an effective operation of the IDC. The complexity and size of configuration information as well as the subtle ways in which small changes in configuration affect results make this a highly non-trivial task. At present, System Maintainers still often make changes in configuration by directly modifying database tables or configuration files without the use of supporting tools. In the re-engineered architecture, graphical user interface tools have been designed that allow the System Maintainer to configure all aspects of station data usage, processing components, analysis interfaces, data acquisition, and forwarding. The System User also has capabilities to view system configuration history. Showcased below are interfaces that allow the System Maintainer to configure station data, as part of the process of defining data acquisition configuration (left), and processing components configuration (right).



Station: XXXX Type: Array
Name: XXXXXXXXXXXXXX
Latitude: ##.## Longitude: ##.## Elevation: ##.##

Data source	Available formats
Site	CD11 XXXX
XXXXXXXX	XXXX XXXX

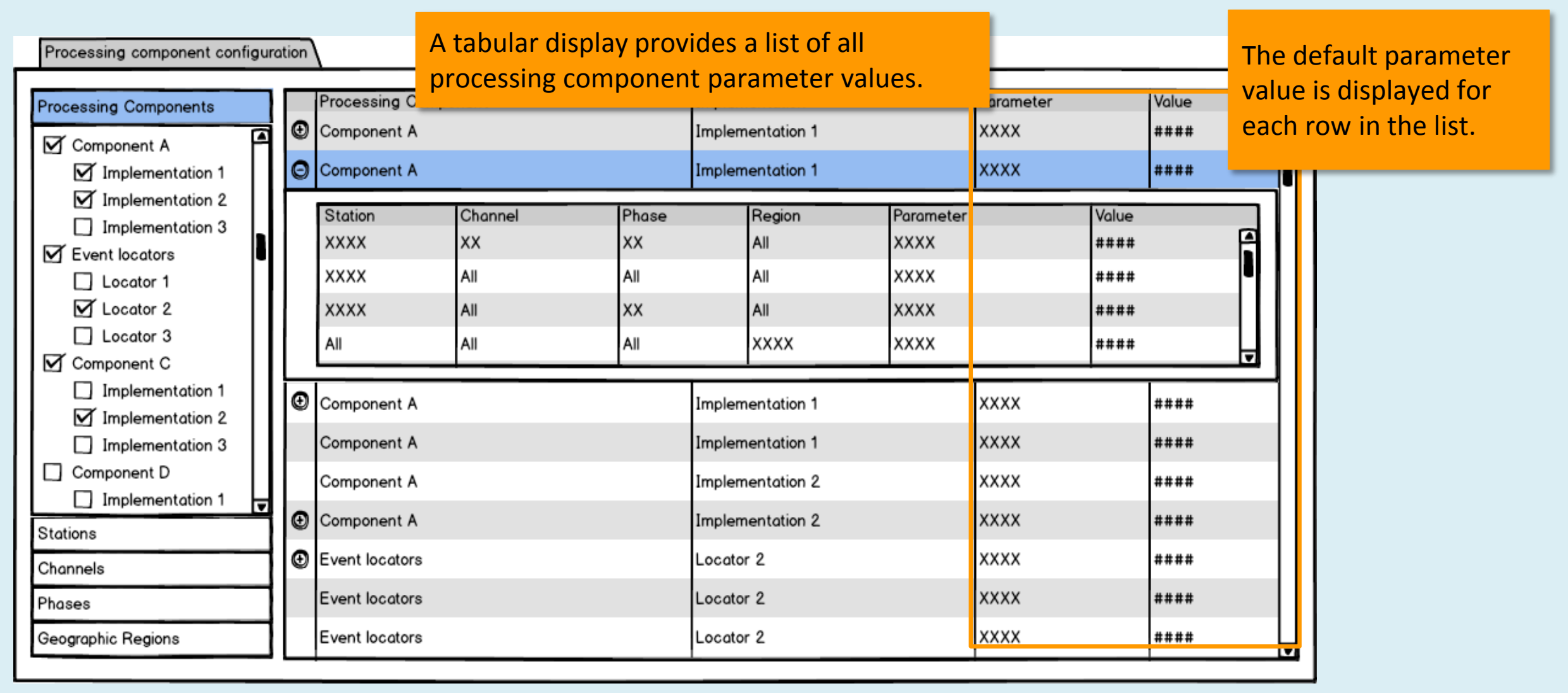
Array element	ID	Latitude	Longitude	Elevation	North offset	East offset	Sample rate
Element A	####	##.##	##.##	##.##	##.##	##.##	##.##
Element B	####	##.##	##.##	##.##	##.##	##.##	##.##
Element C	####	##.##	##.##	##.##	##.##	##.##	##.##
Element D	####	##.##	##.##	##.##	##.##	##.##	##.##
Element E	####	##.##	##.##	##.##	##.##	##.##	##.##

Channel	ID	Type	On date	Off date	Depth
SHZ	####	XXXX	##/##/#### ##:##:##.##	##/##/#### ##:##:##.##	##.##

Input station parameters, such as name, location and ‘on’ date.

For some stations, data can be acquired from multiple sources and in multiple formats. When configuring the station, the System Maintainer adds all available sources and their formats. The System Controller then determines which source(s) and format(s) to acquire data from at a given time.

For an array, configure elements and channel information.



Processing Component		Parameter	Value
Component A	Implementation 1	XXXX	####
Component A	Implementation 1	XXXX	####
Component A	Implementation 1	XXXX	####
Component A	Implementation 1	XXXX	####
Component A	Implementation 2	XXXX	####
Component A	Implementation 2	XXXX	####
Event locators	Locator 2	XXXX	####
Event locators	Locator 2	XXXX	####
Event locators	Locator 2	XXXX	####

Station	Channel	Phase	Region	Parameter	Value
XXXX	XX	XX	All	XXXX	####
XXXX	All	All	All	XXXX	####
XXXX	All	XX	All	XXXX	####
All	All	All	XXXX	XXXX	####

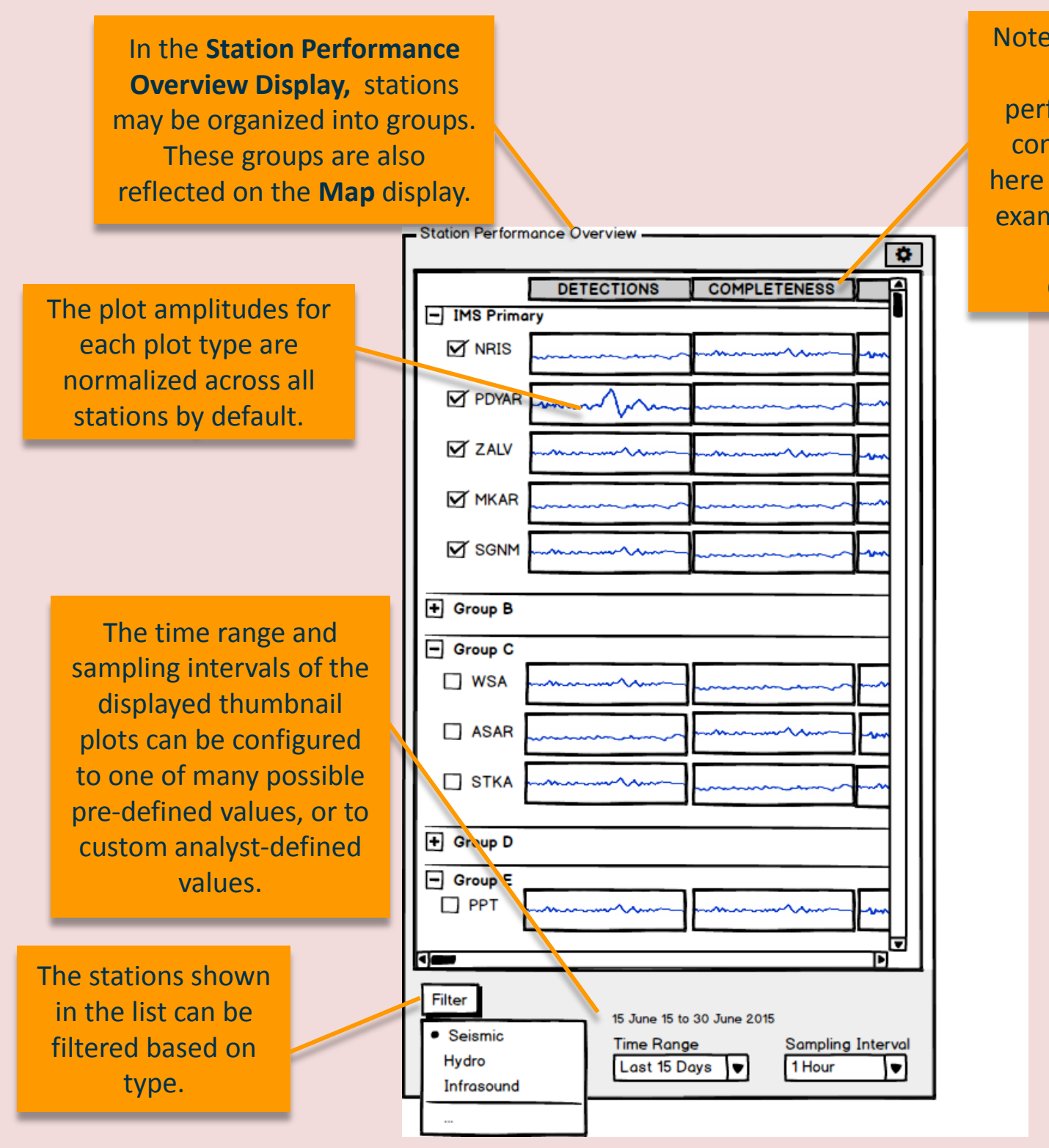
A tabular display provides a list of all processing component parameter values.

The default parameter value is displayed for each row in the list.

Key Processing Improvements: Monitoring Mission Performance

Being able to analyze performance of the verification system, to determine potential problems and investigate their cause, is an important function of the CTBTO. A variety of tools are in use today to help accomplish this function. One of the merits of the re-engineered architecture is to propose a unified and extensible user interface concept that brings together a multitude of existing performance monitoring capabilities into a coherent whole.

The design covers tools for:
- Verifying that *Station Processing* components are correctly processing raw waveform data from each station and appropriately identify signal detections, through metrics such as raw waveform data availability and quality trends, signal detection metrics, station noise estimation through power spectral density and others.
- Verifying that *Network Processing* is correctly deriving events by associating related signal detections from the network of stations and producing events of reasonable quality. Interfaces are provided for viewing event statistics, comparison of event bulletins, magnitude of completeness, continuous threshold monitoring, and network modelling.



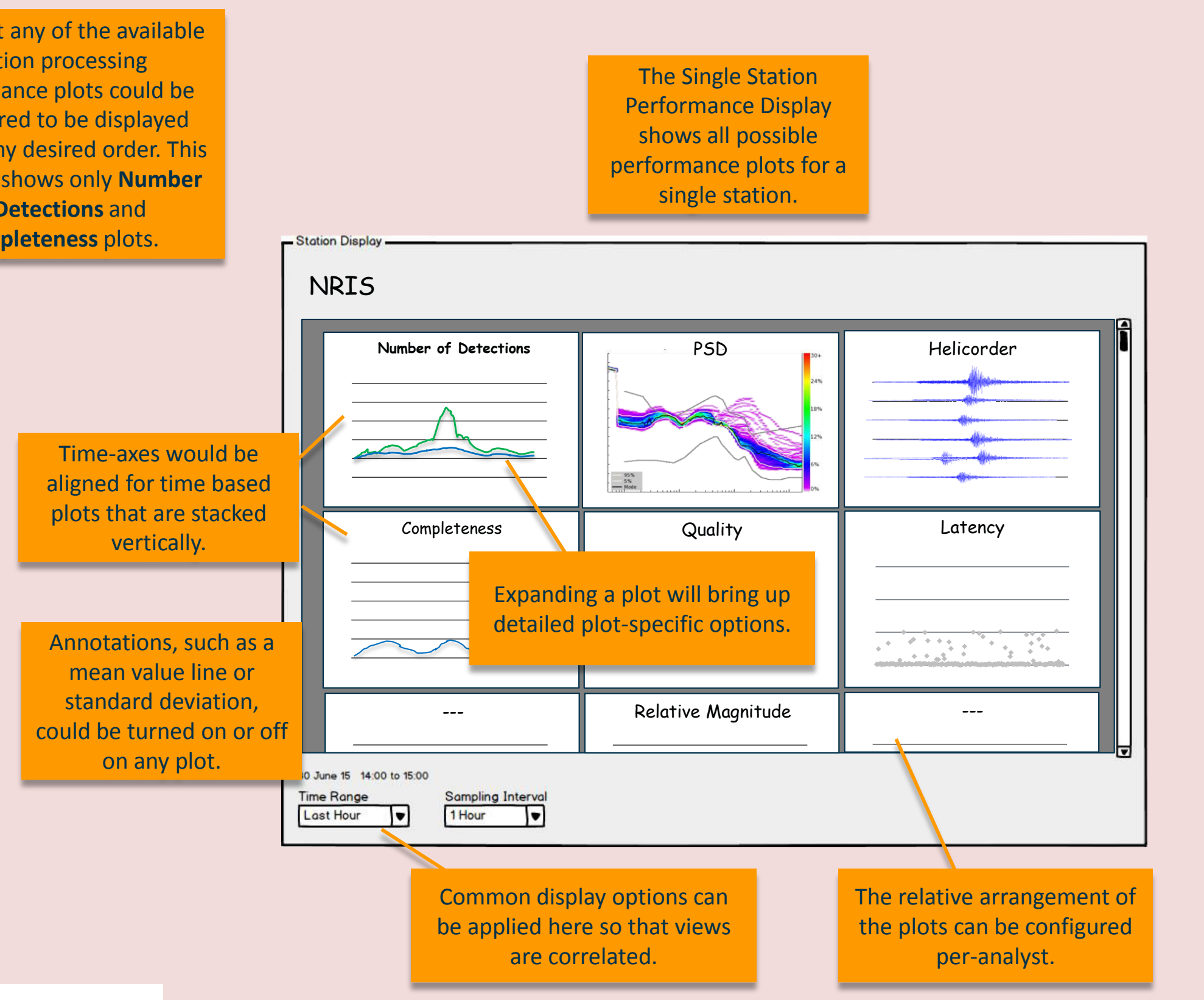
In the **Station Performance Overview Display,** stations may be organized into groups. These groups are also reflected on the **Map** display.

Note that any of the available station processing performance plots could be configured to be displayed here in any desired order. This example shows only **Number of Detections** and **Completeness** plots.

The plot amplitudes for each plot type are normalized across all stations by default.

The time range and sampling intervals of the displayed thumbnail plots can be configured to one of many possible pre-defined values, or to custom analyst-defined values.

The stations shown in the list can be filtered based on type.



The Single Station Performance Display shows all possible performance plots for a single station.

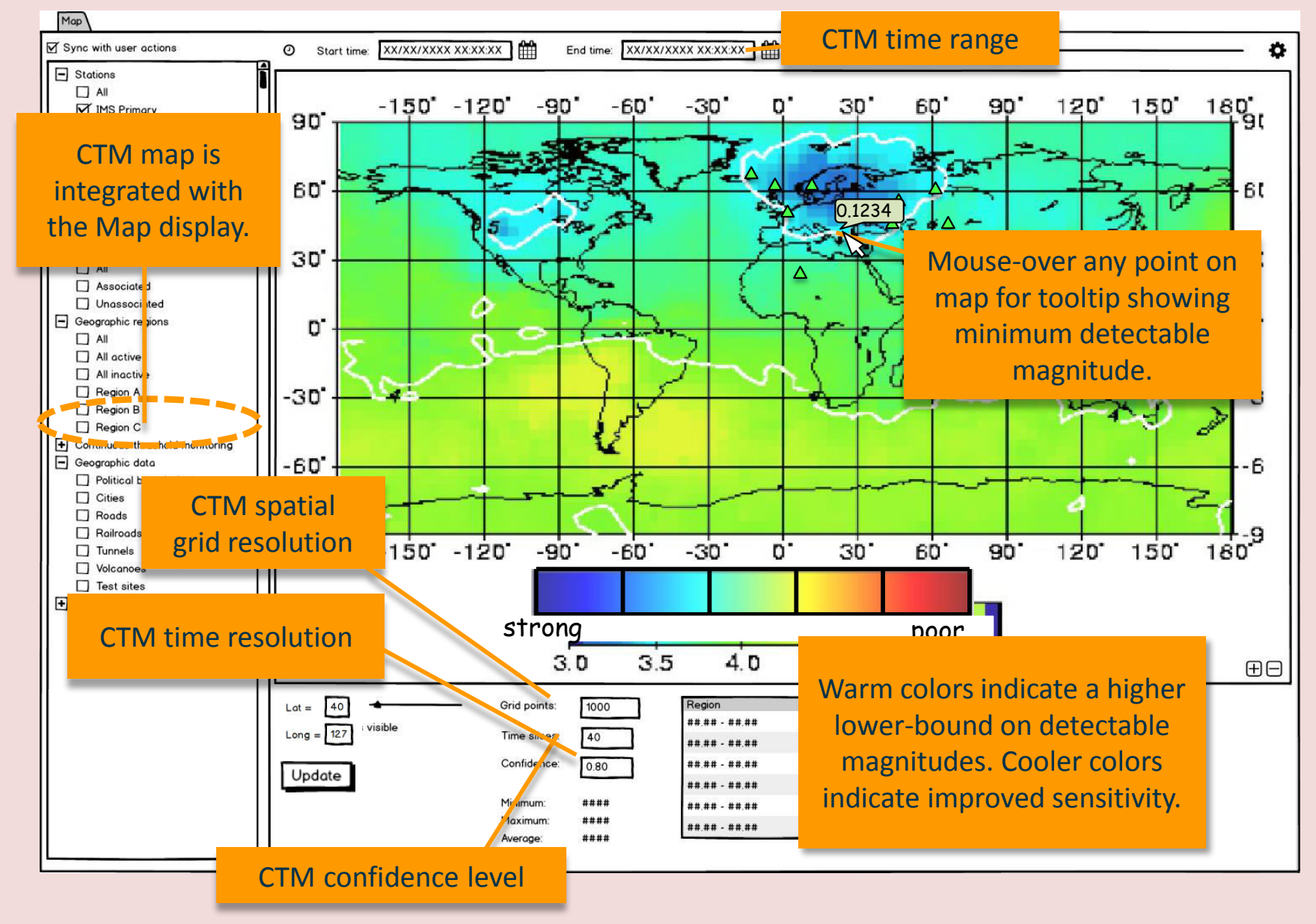
Time-axes would be aligned for time based plots that are stacked vertically.

Expanding a plot will bring up detailed plot-specific options.

Annotations, such as a mean value line or standard deviation, could be turned on or off on any plot.

Common display options can be applied here so that views are correlated.

The relative arrangement of the plots can be configured per-analyst.



CTM time range

CTM map is integrated with the Map display.

Mouse-over any point on map for tooltip showing minimum detectable magnitude.

CTM spatial grid resolution

CTM time resolution

Warm colors indicate a higher lower-bound on detectable magnitudes. Cooler colors indicate improved sensitivity.

CTM confidence level

Continuous Threshold Monitoring (CTM) shows the network detection capability for the International Monitoring System using real-time station noise measurements. The CTM display shown at the left is part of the toolset for analyzing mission performance. It includes capabilities to modify CTM parameters and run CTM on the fly. The CTM display is integrated with the Map and the Station Performance Overview display, the latter shown above at the left. The user can control which stations go into the CTM calculation by selecting stations on the **Station Performance Overview** or the **Map** displays.

What’s Next: IDC Re-engineering Phase 3

Following the completion of RP2, the IDC will engage in a project to implement software based on the RP2 architecture. As outlined in this poster’s section on “RP2 Status”, this development and implementation project, named Phase 3 of IDC Re-engineering (RP3), is expected to start in early 2019. Until that time the IDC will engage in preparation activities aimed at increasing its capacity to develop, test and integrate new software into its operational environment. These preparation activities are centered around:
- Improving IDC testing and release processes, through automation, following DevOps principles, to allow more rapid and more efficient development and release cycles.
- Developing frameworks for the evaluation of processing components, to ensure that components newly developed under RP3 are equivalent or superior in performance with current ones.
- Devising means to run software based on the existing database schemas and the new data model during a transition period.

The views expressed here do not necessarily reflect the views of the United States Government, the United States Department of Energy, the National Nuclear Security Administration, or Sandia National Laboratories. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525.