Scope for IDC Reengineering Phase 2

July 2014

1. Background

The following describes the scope of the IDC Re-engineering Phase 2 program. The purpose of the program is to specify a software architecture and guide the further development and sustainment of the IDC seismic, hydro-acoustic, and infrasound (SHI) processing software.

IDC Reengineering Phase 1 started in 2011 with the goal to significantly modernize SHI automatic and interactive data processing software. The following is in scope of IDC Re-engineering Phase 1:

- Replacement of the SHI network processing software, with a system based on Bayesian inference and machine learning techniques (NET-VISA), that reduces analyst workload.
- Automatic testing of IDC applications through the introduction of a Continuous Automatic Testing Framework, to validate software changes and to maintain and improve software quality.
- Development and design of a new SHI automatic processing pipeline system, based on open-source software.
- Design and development of an integrated waveform data quality control data model and API.
- Redesign of the infrasound automated processing and interactive review software.
- New software for distributing data and products to member states and for acquiring data from SHI and radionuclide stations.
- Physical data model enhancements focusing on simplified management of archived data, increased security, and implementing referential integrity best practices.

While Phase 1 of IDC Re-engineering focused on enhancements to individual components of the system, Phase 2 addresses the task of specifying a unified architecture for all SHI software, across processing stages, to pave the way for further software development and sustainment in the future. A well-documented architecture would be valuable for preparing medium to long-term budget and staffing plans for the Commission. Furthermore, it provides a framework for the Member States to prioritize the software needs of the PTS when considering voluntary financial or in-kind contributions.

A key objective of Phase 2 is an architecture sufficient to provide a basis for a cost estimate for the development or enhancement of the software components and sub-systems. Implementation of software components based on this architecture will be addressed in a third phase of IDC Reengineering. Phase 3 will begin with the creation of an implementation plan that will include priorities and timelines for developing software components complying with the architecture created in Phase 2. It is foreseen the Phase 3 will start in 2017, as Phase 2 comes to an end. Following terminology of the Rational Unified Process software development methodology (e.g. http://en.wikipedia.org/wiki/Rup) Phase 2 of IDC Reengineering covers the Inception and Elaboration stages, while Phase 3 will cover the Construction and Transition stages.

Phase 2 of the IDC Reengineering project contributes to Strategic Goal 1 (Operation and sustainment of the verification system) of the Mid Term Strategy (MTS), in particular to Enhanced Sustainment of the IDC Infrastructure and PTS Internal Systems and Network, as laid out in paragraph 38 of the MTS. It is also part of the technology refreshment program described in paragraph 15. The IDC Re-engineering Phase 2 is also expected to provide the basis for establishment of an IT Disaster Recovery System, specified in paragraph 41 of the MTS.

2. Business Case

In the last 16 years, since the delivery of the first version of the SHI acquisition, processing and product dissemination software by the 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. This project aims to develop a modern, model-based component architecture 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.

3. Goals

The goals of the system architecture to be developed in Phase 2 are to:

1. Unify 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 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, 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.

4. Objectives

The following objectives support the goals of Phase 2.

1. The proposed software implementation is faithfully represented and documented in a hierarchical UML-style model that guides software operation, maintenance and development, and is continuously updated to be synchronized with the current system.
2. All waveform technologies are fully integrated in the processing and analysis toolsets.
3. The software architecture supports analysis based on scientifically-sound fit-for-purpose geophysical models, wave propagation algorithms and inference methods .
4. The system architecture is:
   - 4.1. Modular, with well-defined interfaces that enable software developers from different organizations to build applications that can be integrated directly.
   - 4.2. Layered, with a clear understanding of responsibilities between layers and with minimal coupling between layers.
5. The system architecture encourages code reuse and uses encapsulation, with data accessed through a conceptual data layer that simplifies access, hides the implementation details of the storage mechanism, and minimizes the impact to client applications of changes in the physical data layer.
6. The system architecture builds on the results of Phase 1, taking advantage of ongoing developments in automatic processing, advanced scientific algorithms, and interactive analysis, product distribution, and automated testing.
7. Analysis provenance is captured and easy to access; system configuration history is maintained and accessible.
8. The software architecture explicitly recognizes the need to provide a platform for training new analysts and processing engineers as well as improving the skills of existing ones.
9. The architecture encourages the development of open software using open-source components and the interoperability with 3rd-party software through standard interfaces.
10. External data sources can be accommodated and integrated into SHI analysis to support activities and analysis requested by the Member States.
11. Software is developed in a consistent style using a minimal set of prevalent, modern programming languages.
12. Interactive analysis software tools are implemented following modern best practices in UI design.
13. New computer technologies can be incorporated efficiently.
14. Analysis and system monitoring may be performed remotely.
15. System alerts and notifications are performed in a consistent way to provide all responsible and otherwise authorized parties with access to timely information about system status and events of interest.

16. System performance is monitored and reported to authorized users, including Member States.
17. System Design explicitly meets requirements for an IT disaster recovery system.

5. Deliverables

The deliverables of Phase 2 are as follows:

1. System requirements document (SRD): a high-level list of declarative-style requirements covering all aspects of SHI data acquisition, automatic and interactive processing, and SHI data and product dissemination. Each requirement is associated with a status, which indicates whether it is satisfied in the current SHI system, and a priority, which indicates its importance to future development.
2. System specification document (SSD): A refined version of the SRD with all requirements, both explicit and implicit, in the SRD expanded, clarified and specified.
3. Use Case Model (UC): a document describing in detail all use cases derived from the SSD, including system actors and relationships between use cases.
4. Use Case Realization (UCR) document: A description of analysis classes, their attributes and responsibilities, their assignment to various layers in the architecture and the way in which they interact in order to realize the architecturally significant use cases described in the UC.
5. System Architecture Document (SAD) containing: a specification of the main logical and physical components of the system and their interfaces, including in particular interfaces and components for SHI automatic and processing applications and the data access layer.
6. A definition of the technologies to be used to implement the architecture (e.g. programming languages, class libraries, third party data access components and middleware)
7. A proposal of how physical components of the system shall be distributed on hardware hosting the system, as well as requirements for the hardware.
8. A document outlining the principal elements of the physical data model that are required to support the architecture specified in the SAD.
9. A prototype of the core system architecture that validates the important design features specified in the SAD, and confirms that the design can be implemented with minimum risk.

6. In Scope

The new architecture will cover the following components of IDC software:

- Continuous Data Acquisition: (1) IMS waveform data are automatically forwarded to the IDC; (2) data are reformatted, archived, and made available for processing.

- Data Forwarding: (1) IMS waveform data are automatically forwarded to states parties (states parties control what data to receive).
- Auxilliary Data Acquisition: (1) IMS data from the Auxilliary seismic network are requested and integrated into the processing stream. (2) Auxiliary seismic stations may be switched to Primary in the event of a failure of a Primary station, as authorized according to the Draft IDC Operational Manual.
- Data Import: Other data from, e.g. Cooperating National Facilities (CNF), or National Technical Means (NTM) are acquired and imported into the processing environment according to the protocol for integrating these data with IMS data.
- Data Processing: (1) Data are processed to detect and characterize events (e.g. lat, lon, depth, time, magnitude).
- Interactive Analysis: 1) automatically detected events are reviewed and corrected; 2) missed events are manually added.
- Event Screening: (1) event features are computed to allow, for example, screening of obvious non-explosions and other screening based on user (i.e. Member State) specified criteria.
- Special Event Analysis: Special analysis of events of interest may be analyzed using data from CNF and NTM sources integrated with IMS data.
- Data and Product Dissemination: (1) various event lists (automatic, analyst reviewed, screened), waveform data, station quality products.
- Monitoring and control tools used by the Processing Engineers.

Some software, e.g. that for data acquisition and product dissemination, have been recently upgraded and are not targeted for redevelopment. On the other hand, the software for automatic data processing, interactive analysis, event screening, monitoring, and control tools are based on largely outdated software technologies and these will be the subject of focused attention in the software architecture design.

7. Out of Scope

The following is out of scope of the project:
- Specification of a hardware architecture to host the system.
- IMS Stations State of Health, except the information that is part of the CD protocol.
- Software architecture to support station and network operation, including GCI, IRS.
- Specification of architecture to cover radionuclide software. (However there is some overlap in the software systems that should be considered, including, for example, the dissemination of radionuclide products, the acquisition of meteorological data, and the fusion of radionuclide and SHI results).

8. Assumptions

- The minimum sensor network (size and complexity) is specified in the Treaty..
- The available SHI stations that may be imported for special processing will increase over time. The number of openly-available stations currently numbers in the 1000's. A substantial number of additional stations may be provided by Member States through National Technical Means. Stations contributing data as Cooperating National Facilities represent yet another source of data. Each of these data sources will be subject to different protocols regarding their use in the products of analysis.
- It is assumed that the current requirement of acquiring data from Auxiliary seismic stations through explicit data requests will remain valid in the long term.
- The system will be supported by an international team, all with limited term appointments.
- Most of the software development will be contracted, potentially to several geographically dispersed teams.
- A full-time CTBTO software engineer will be assigned to the project for its full duration.
- Project stakeholders (as listed in section 10) will be available to review and provide input on the project deliverables.
- Components of the core system architecture shall be specified to a sufficient level of detail (in terms of protocols and interfaces) to allow the CTBTO to use different components developed according to the same specifications by other organizations (e.g. software donated by Member States as contribution in kind or software developed by external contractors).

9. Constraints

- The system must initially implement the procedures, services, products and metrics described in the Draft IDC Operations Manual.
- Stakeholders include the Member States, and their experts will be invited to provide input on software requirements.
- The system must run on commodity hardware.
- No system transition down time. The old/new system must continue to function consistently and reliably as components are modified/replaced, in a staged manner, during the period of modernization
- The modernized system must be sustainable and adaptable to an evolving mission for next ~20 years.

10. Authority and Responsibility

The following are the internal stakeholders for the project:

- For data acquisition, Chief of IDC/Operations Section

- For automatic data processing and data product distribution, Chief of IDC/Automatic Processing Systems Section.
- For data analysis, Chief of Monitoring and Data Analysis Section.
- For distribution of and training on stand-alone system (NDC-in-a-box), Chief of Capacity Building and Training Section
- For software maintenance, Chief of IDC/Software Applications Section.
- For hardware maintenance, Chief of IMS/Network and Systems Support Section