

Exceptional service in the national interest



SNL Elaboration Phase 1 (E1) Review for IDC Reengineering Phase 2

**IDC Reengineering as a
Collaborative Research & Development Environment**

Steve Kubica

2016 January 19

Proposition

- The re-engineered IDC software will facilitate the collaborative research and development of new algorithms and approaches for SHI data processing.
 - Framework for prototyping and testing new algorithm code
 - Shared baseline for collaborative development and evaluation
 - Cleaner path to deployment
- This proposition is realized by constructing a system that is ***scalable, modular, and extensible.***
 - Modern software engineering design principles provide the foundation.

Scalability

The re-engineered data center software will be scalable.

- Goal is to be able to run on anything from a high-end cluster of servers down to a laptop.
- This “NDC in a box” idea means any researcher can download and run the *same software* that is being used by the production system using commonly available hardware.

Modular

The re-engineered data center software will be modular.

- The modules within the system will clearly encapsulate responsibility and have well-defined interfaces.
- Layered interfaces will isolate algorithm code from being tied to specific underlying technologies.
 - Breaks the tie to a specific (expensive) database solution.
- The **Data Model** and an **Object Storage and Distribution** (OSD) mechanism will provide a clear and consistent interface for accessing and storing data.
 - Service interface for language independent data access.
 - Library interfaces also available for more stressing data access patterns.

Extensible

The re-engineered data center software will be extensible.

- A **Component Interface Specification (CIS)** will define how R&D algorithms can be plugged-in to this system.
 - No code changes required to the core system to add a new algorithm.
 - Add-on code can be developed outside the core system and be invoked dynamically at runtime.
 - Research code could be distributed separately as an add-on “toolbox” to be used by anyone with access to the core system.
- The **Processing Sequence Controller (PSC)** is a configurable mechanism that support the insertion of different workflows and different algorithmic sequence chains.

Open Source Collaboration Model

- The goal is to make the re-engineered system openly available for development within the community.
 - Plug-in toolboxes mean you don't *have* to modify the core system, but you are not prohibited from doing this if you need to or want to.
- 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.
 - The “fork-and-pull” model allows anyone to fork a copy and then offer changes back for consideration via pull-requests.
 - Note that Mikey Dickerson (head of new U.S. Digital Service) advocates all government funded software be made open source and freely available.

R&D Iterative & Collaborative Workflow

