Towards a Domain Specific Software Architecture for Scientific Data Distribution
Anne Wilson, Douglas M Lindholm, LASP, CU Boulder

The Utility of a Reference Architecture: the Best of Best Practices

A Reference Architecture
- is a generic and somewhat abstract blueprint-type view of a system
  - includes the system's major components, the relationships among them, and the externally visible properties of those components.
- not usually designed for a highly specialized set of requirements. Rather, architectures tend to use it as a starting point and specialize it for their own requirements.

Consider a residential architecture. The implied model for a house is composed of the following components:
- A foundation and/or other subfloor structure to connect the house to the underlying environment, whether it is earth or a body of water
- Floors to stand on
- A reference architecture provides a template solution for an architecture for a particular domain.
- a common vocabulary with which to discuss implementations, often with the aim to stress commonality.

A reference software architecture
- is a software architecture whose structure and respective elements and relations provides templates for concrete architectures in a particular domain or family of software systems.
- often consists of a set of functions and their interfaces that can be defined at different levels of abstraction
- "The best of best practices"
- An iterative, community process

Examples
- NASA's Earth Science Data Systems Working Group (ESDWG)
- IPython: a software system to distribute scientific data to end users.
- 

NASA's Earth Science Data Systems Working Group (ESDWG)

**Motivation:** handle explosion in data volumes, increased system complexity, diverse data sources, greater user expectations,...increased collaboration.

**Uses case, scenario:** require sensor data, develop products, distribute products, develop predictions, manage remote sensor/instrument, data stewardship.

**Stakeholders:** system architect, data architect, program manager, project scientist, policy research, research scientist, NASA management...

Here we focus on RA use case class 3, "Distribute Products":
- Generated products are made available to end users.
- This includes discovery and access...tools, as well as actual delivery.
- "...and possibly other components as..."

**Domain Specific Software Architecture (DSSA)**

A process and infrastructure to develop:
- DSSA is a software architecture with reference requirements and a domain model. The infrastructure supports it.
- "...and a process to instantiate or refine it."
- DSSA focuses on the architectural level and modeling of the domain for additional insight and understanding. Requirements need to be evolved based on feedback.

Reference Architecture
- provides an unambiguous understanding of aspects of the domain

**Domain Model**
- a framework to capture application domain knowledge, the problem space
- Modeling tools: sequence, vocabulary, context/behavior diagrams, entity/relationship diagrams, data flow models, class diagrams...
- Analysis involves domain experts, customer inputs, others familiar with the application

**Goal:** provide an unambiguous understanding of aspects of the domain

Use cases/scenarios
- define functional, non-functional requirements
- define functional, and implementation: "how"

Domain: Scientific Data Access and Distribution

**For each component provide values for this template: name, rationale, constraints, rules of thumb, lessons learned, etc.**

**References:**
- Armitage, 1993
- By Biggerstaff
- Presented here are first steps towards a DSSA for scientific data distribution

**Conclusions**

Benefits to having a DSSA for scientific data distribution
- A reference architecture provides a documented high-level design for a family of applications and places constraints on implementation.
- This generalization helps reduce the time and effort needed to develop. By beginning at this level and providing specific, abstract, well documented models, the process of developing consistent, reusable services is greatly simplified.
- ...in essence, a DSSA for scientific data distribution will help developers provide specific, abstract, well documented, and deployable services that can be integrated into any component system.

PREPARE first steps towards a DSSA for scientific data distribution

**References:**
- [Armitage, 1993]
- [By Biggerstaff]
- [Law, Emily, 1993].

**Presented here are first steps towards a DSSA for scientific data distribution.**