A Domain Specific Language for SDTM and ADaM Datasets

ACRONET
Ippei Akiya

Disclaimer

- This presentation shows my personal opinion, therefore concept, method, technology and etc. which are introduced in this slide is NOT representative about my company position statement, business strategy, providing service and others.

Introduction

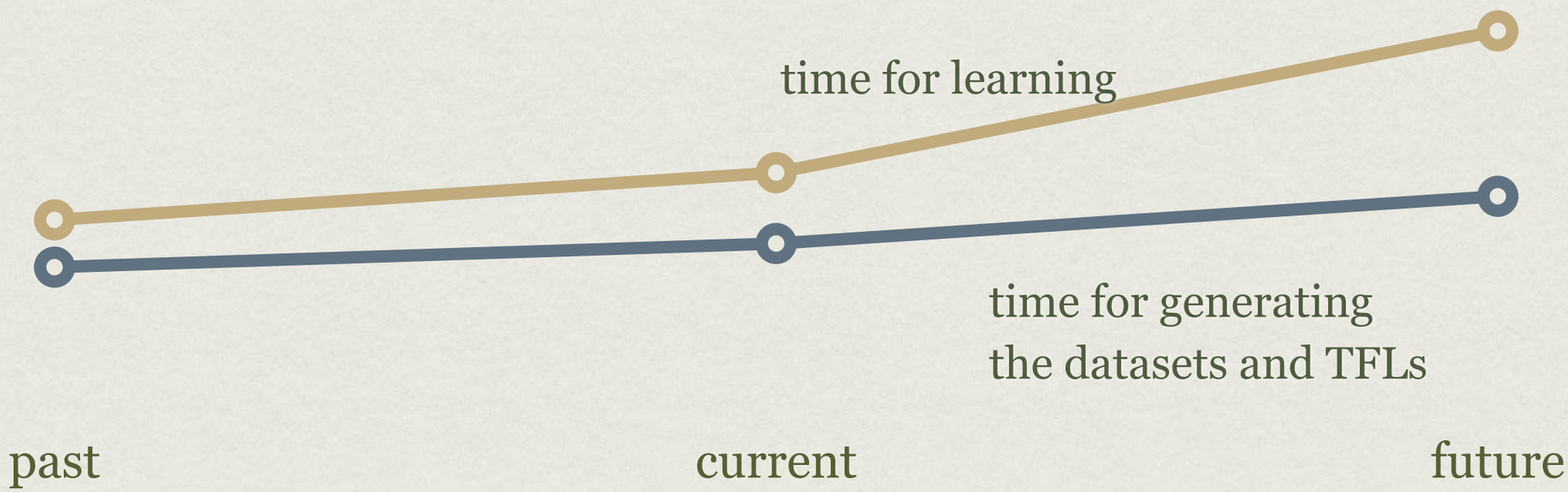
Recently, Role of a Clinical Programmer is expanding widely and deeply.

A Clinical Programmer is required to fully understand CDISC Standards and the guidances from the regulatory agencies.

Further, volume of these documents is continuously increasing.

Problem

As a result, a Clinical Programmer has to learn new standards and guidances, of cource additionaly new programming techniques.
In other words, a Clinical Programmer spends a lot of time for learning the standards and guidances compare to a time for generating the clinical datasets and TFLs.



Solution

- Metadata Registory (Metadata Repository)
  - Metadata Registory, similer to CDISC SHARE, is able to compress the time for learning the standards and the guidances.

- Domain Specific Language (DSL)
  - The specialized programming language for the CDISC dataset is able to compress the time for both learning the standards and writing the code.

Matcha Script

- Today's presentation is introduction about the project of a DSL for CDISC datasets.

What is a DSL?

- A domain-specific language (DSL) is a computer language specialized to a particular application domain. *(source: Wikipedia http://en.wikipedia.org/wiki/Domain-specific_language)*

- Examples for DSL
  - SQL
  - XQuary
  - BUGS
  - Matlab, SAS, and R
  - etc.

Goal of Matcha Script

- Reduce the code volume
- Reduce the time for writing code
- Metadata Driven
- Elevate a data quality

Scope of Matcha Script

What Matcha Script does:

- generate SDTM and ADaM datasets

Matcha Script does NOT:

- provide any functions for statistical analysis, generating TFLs, metadata definition, and etc.

Language Design

- Matcha Script is designed:
  - Fewer code
  - Simple syntax and expression
  - Higher readability
  - Running on Java Virtual Machine
  - Parallel Processing

Eliminate a code for…

- long text value over 200 bytes [SDTMIG 4.5.3.2]
- supplemental Qualifiers in separate datasets [SDTMIG 8.4.2]
- insert new analysis parameter in ADaM BDS.
- etc…

Metadata Driven

- Create define.xml or RDF in first
- Macha Script generates class definition about datasets and variables automatically

Code Example:

```
def dm = new dataset()
def dm_rec = new record(“DM”)
        dm_rec.USUBJID = “xxx-yy-001”
        dm_rec.AGE = agecal(dm_rec.BIRTHDT, “2014-04-15”)

dm.add(dm_rec)
```

Built-in Functions

- Age Calculation
- Date diff Calculation
- Native support ISO8601
- Execute OpenCDISC Validator
- Basic Mathematical and Statistical Functions

Matcha IDE

Photo by http://www.flickr.com/photos/kiri_no_hana/

Matcha IDE

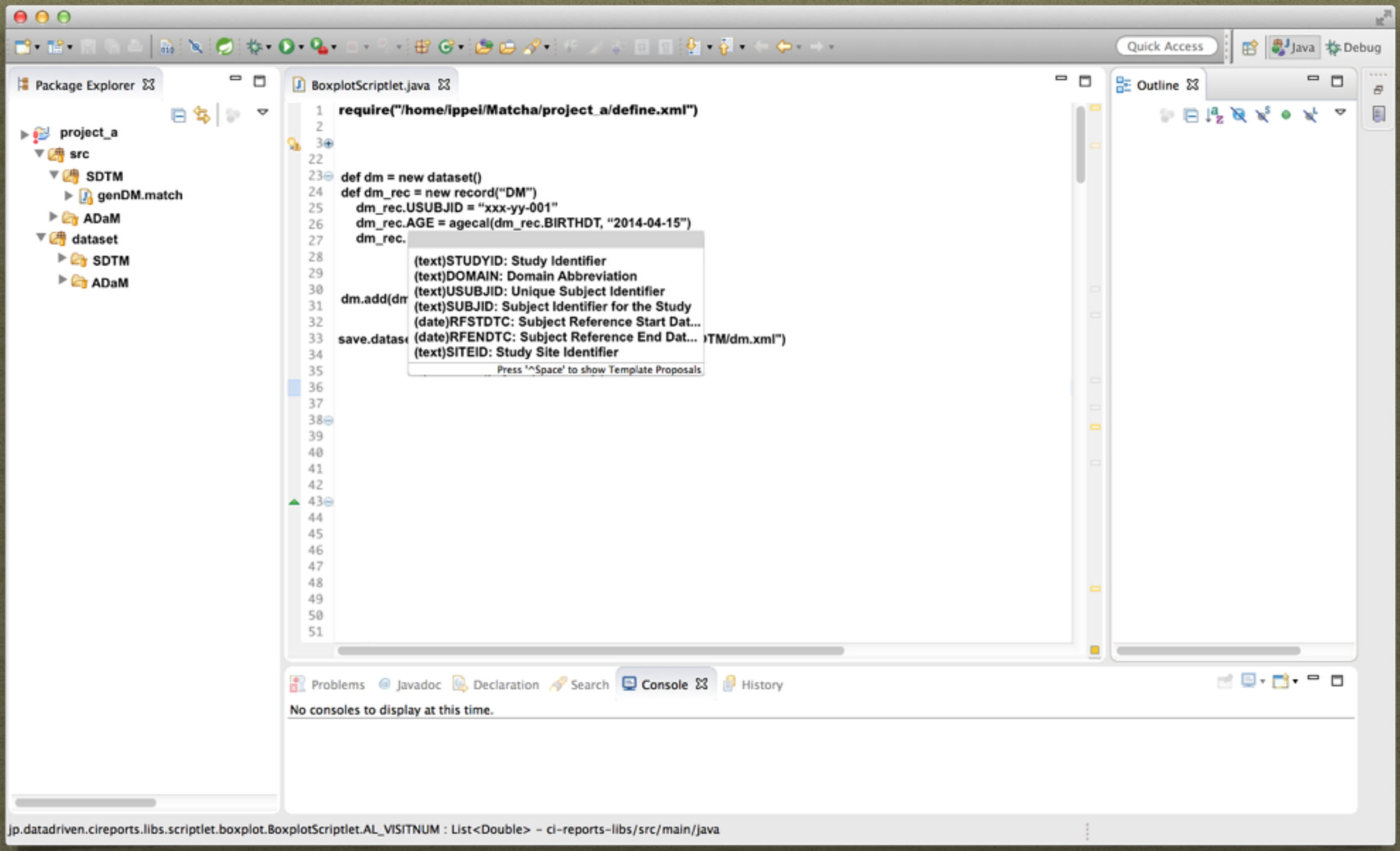
- If new language and script engine will be developed, it is not enough to realize an accerated data implementation.

- Integrated Development Environment(IDE) assists to write a code.

Main Features

- Code completion
- Code formatter
- Version Control (SVN or Git client)
- Display Data Standards via Metadata Resistory

Screen Image



Matcha Script Project

- Project Goal: Develop and maintain Matcha Script
- Deriverables: Open Source and free for use

- Project Web Site: http://matcha-lang.org/
- GitHub: https://github.com/i-akiya/MatchaScript

- The first beta version will be released Mar 2015

Questions?

Photo by Janine