

DataForge: A DDI-Enabled Toolkit for Researchers and Data Managers

Arofan Gregory

Pascal Heus

J Gager

Metadata Technology North America

An Observation...

- DDI is a complex standard
 - It has to be, to support the management of sometimes complex data
- The organizations who use DDI have the capacity to handle the complexity
 - Training staff in the standard
 - Implementing IT tools
 - Organizing and migrating metadata

What About Researchers?

- It is unrealistic to expect researchers to expend the same effort to learn and use a standard
- But unless researchers are using DDI, the work has to be done by the archives and libraries where they deposit their data
- Most research projects have lots of different proprietary tools, databases, and formats
 - The data is not easy to re-use across software packages

A Solution to This problem



DataForge

- DataForge is a simple tool for performing useful tasks for researchers and data managers
- It does not require any knowledge of DDI
- Simple, command-line interface

Two Packages



SledgeHammer

For mining metadata out of proprietary formats, expressing it in standard formats, generating summary statistics, and creating imports and set-ups



Caelum

For generating reports and codebooks in PDF and HTML

Sledgehammer Functionality

- DataForge can read SAS Script plus ASCII, SPSS, and Stata files, DDI plus ASCII, and StatTransfer plus ASCII
- The metadata is mined out of these formats and can be exported as DDI 1.0/2.1, 2.5, and 3.1
- Also supports Triple-S (in future, SDMX support is planned)

Sledgehammer Functionality (2)

- Can generate summary statistics from the data (include min, max, average, standard deviation, missing count, weighted/unweighted frequencies)
- Can generate scripts for reading data into SAS, SPSS, and Stata
- Can generate SQL for relational data bases (MySQL, Oracle, MS-SQL, Vertica)
 - Creates database schema
 - Loads ASCII data
- Can run as an interactive command line, or in batch mode

Caelum

- Provides a simple XSLT-based tool for generating codebooks and quality reports from DDI metadata
- Outputs include HTML and PDF
- Runs with a single command line
- “Template” transformations can be modified
- Custom XSLT can be used

Demo

Planned Release and Licensing

- DataForge tools are available as freeware, to be released in the spring of 2012 (IASSIST is the target)
- We are also starting a beta program for a professional release of Sledgehammer
 - We are looking for interested beta-testers
- Long-term plans are to make DataForge tools available as software-as-a-service (SaaS)
 - Currently only stand-alone
 - Will be integrated with the OpenMetadata.org site