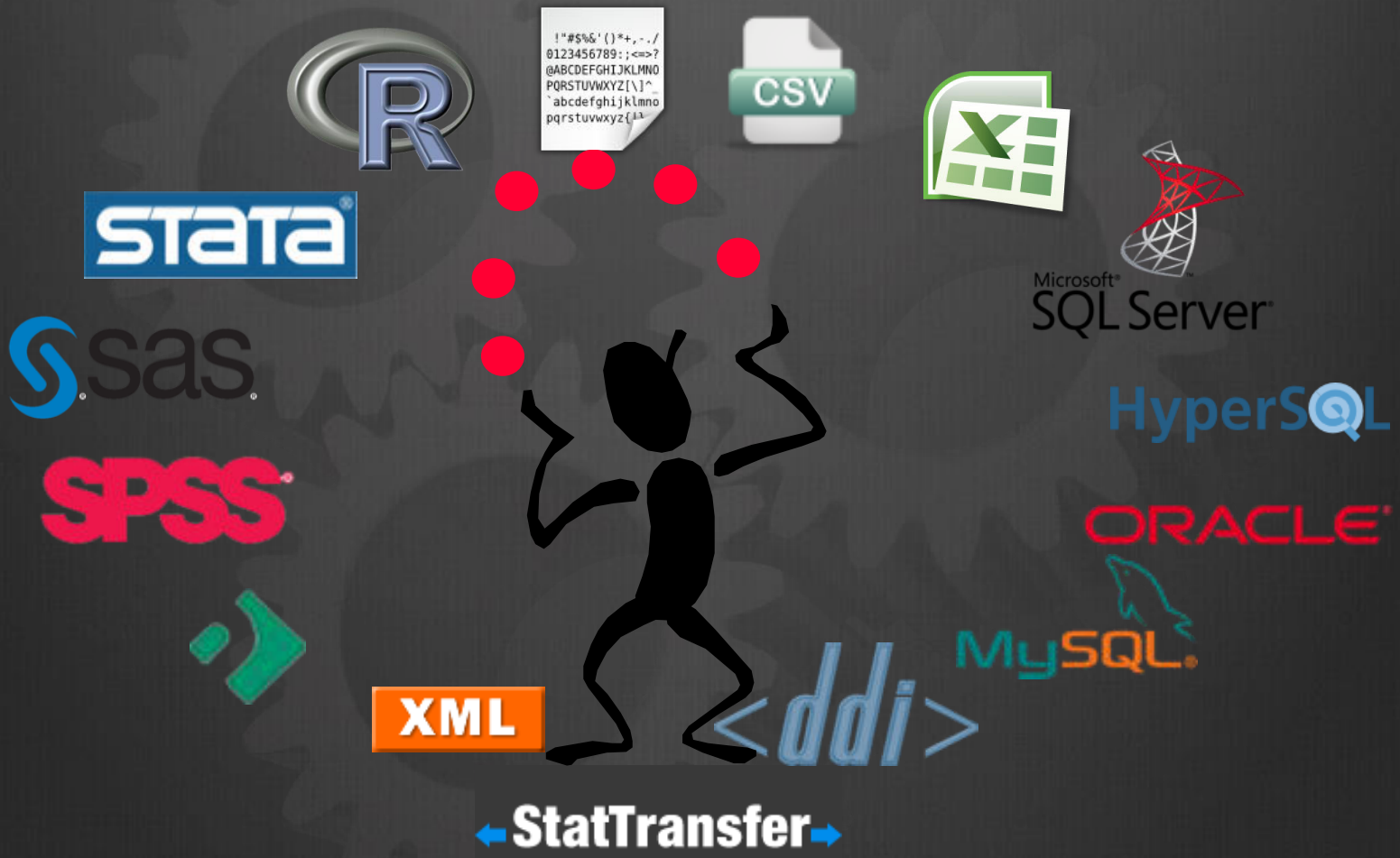
DataForge

EDDI 2013
5th Annual European DDI User Conference
December 3, 2013 – December 4, 2013
Paris, France

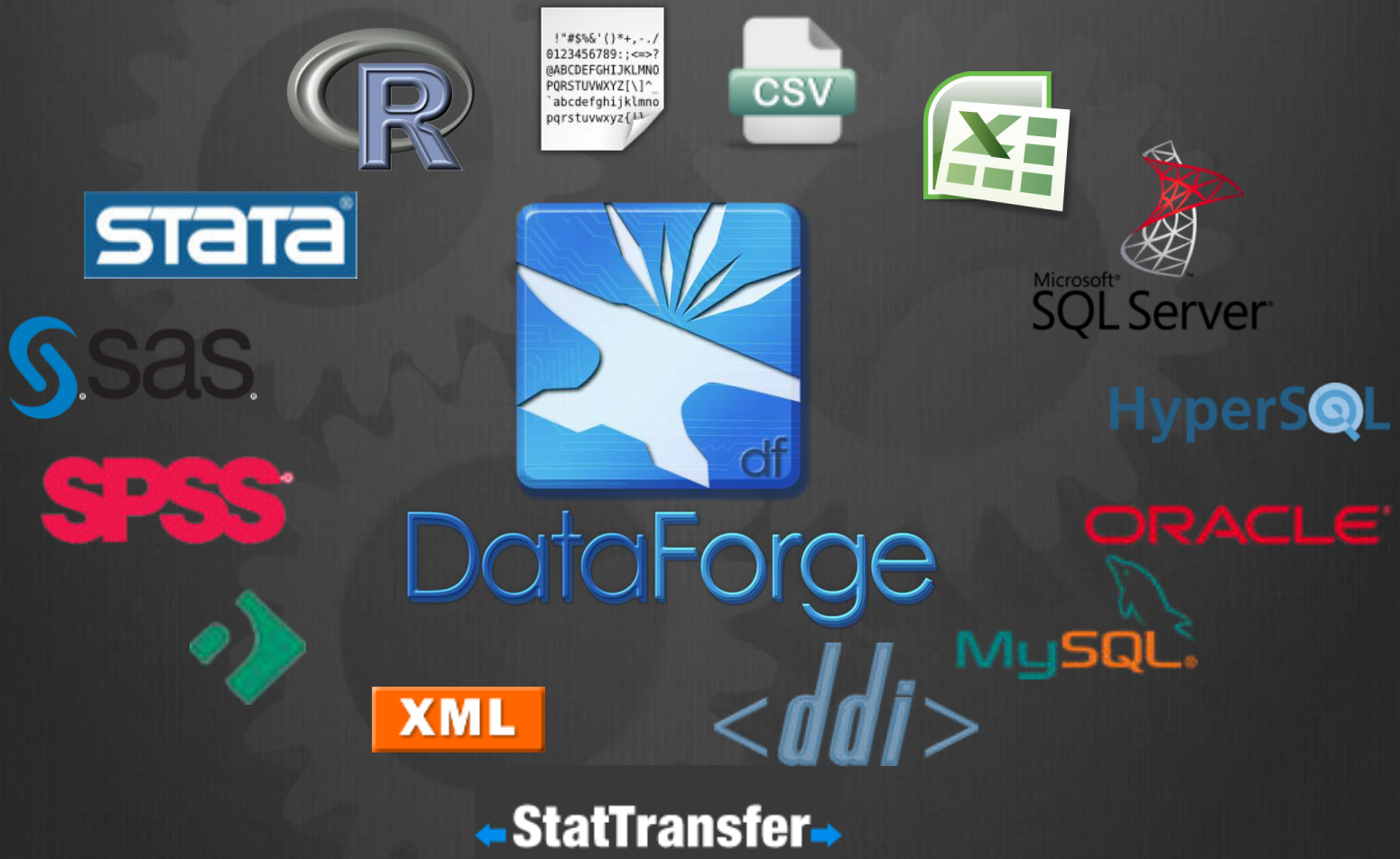
Metadata Technology North America
http://www.mtna.us
http://www.openmetadata.org
mtna@mtna.us

The <Meta>Data Dance



The <Meta/>Data Dance

The <Meta>Data Dance



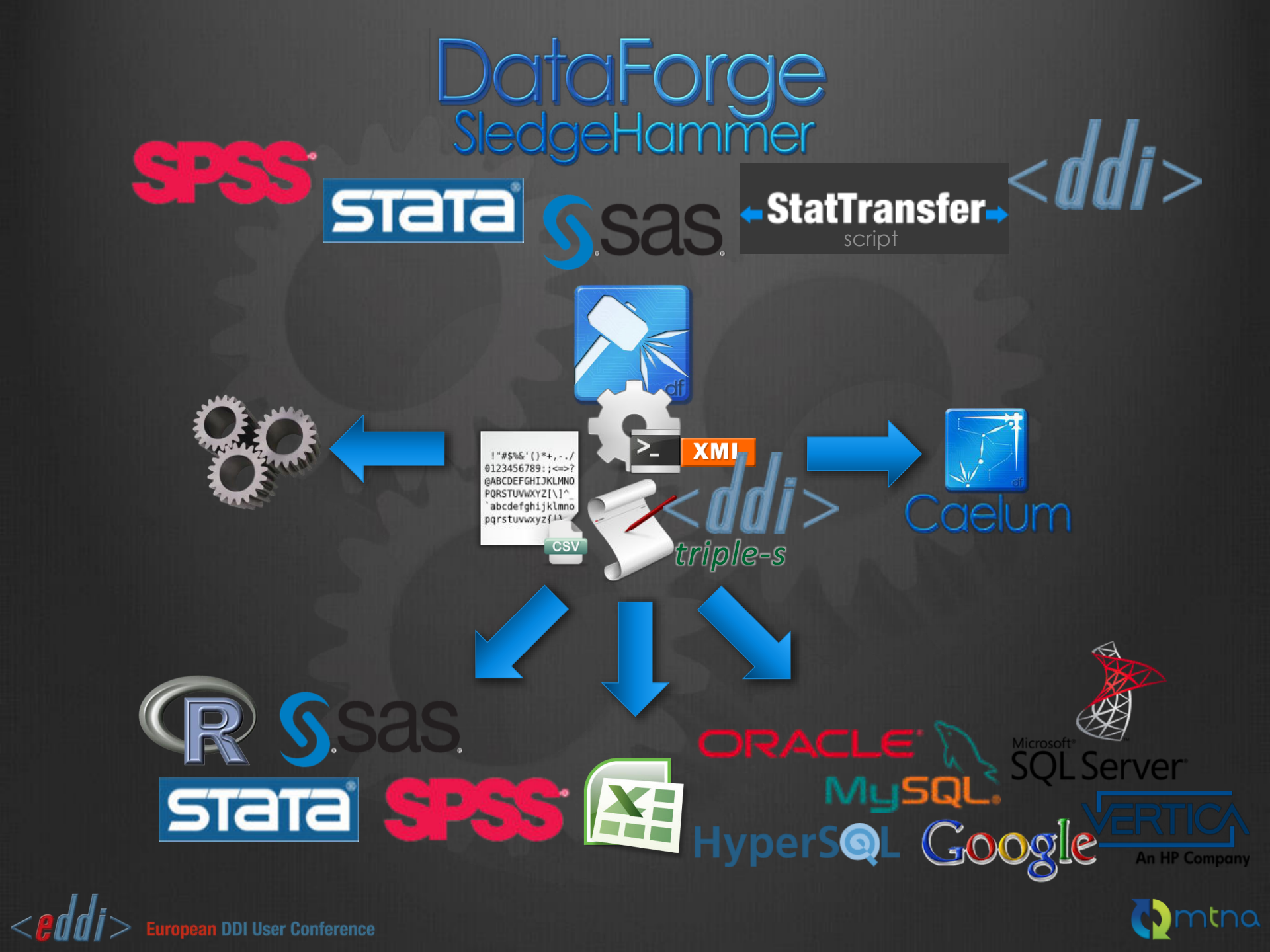
DataForge

- Tools to facilitate statistical/scientific data management
  - Complement existing tools (not replace)
  - Facilitate adoption of standards and promote best practices
  - Alleviate the need to master complexities of metadata standards, technologies
- Vision: web based services (cloud users), stand alone desktop /server products (agencies/enclaves), libraries (developers)
- Broad Target audience: Data Producers / Manager / Researchers / Users
- DataForge Today (lean model):
  - Simple desktop based tools
  - **SledgeHammer** - Data/metadata transformation
    - LavaCore - Java library
  - **Caelum** - Reporting / Publication
  - **Asmurex**: ASCII Multi-Record Exctraction tool
- DataForge Tomorrow:
  - DataForge Online, Web Services, Desktop w/UI, Java Library

DataForge

Sledgehammer

- Data Liberation
  - Unlock from proprietary formats, turn into open data
  - Produce metadata + convert data into standard ASCII formats
- Open <Meta>Data Integration
  - Generating scripts for loading into stats packages, databases, big data engines, cloud
  - Use with open systems / environments


DataForge
SledgeHammer
SPSS
STATA
sas
StatTransfer
script
<ddi>
XMI
CSV
<ddi>
triple-s
Caelum
R
sas
STATA
SPSS
ORACLE
MySQL
Microsoft SQL Server
HyperSQL
Google
VERTICA
An HP Company

DataForge SledgeHammer

- Challenges
  - Read proprietary formats
  - Data integrity
  - ASCII optimization
  - Cross package formatting issues
  - SQL Database limits (columns, indexes, views)
  - Performance (Timings, size, memory usage)
  - …

DataForge SledgeHammer

- Key Features
  - Input: SPSS, Stata, SAS/Syntax+ASCII, ASCII+DDI-C / DDI-L, ASCII+SSS 1.1/2.0, ASCII+StatTransfer
  - Data Out: Fixed w/optimization, CSV, Delimited
  - Metadata Out: DDI-C 1.2.2/Nesstar/2.1/2.5, DDI-L 3.1-RP/3.1-SU/Colectica, DDI-L 3.2-RP, DDI-L 3.2 SU, Triple-S 1.1/2.0
  - Summary Stats: Min / Max /Valid/Invalid /Mean, StdDev /Variance / Frequencies, *Weighted *, Save in CSV**.
  - Scripts Generators: SAS, SPSS, Stata, R, Mysql, *Oracle*, MS-SQL*, HSQL*, Google BigQuery*, HP/Vertica** + *dimension/lookup tables, indexes/foreign keys**
  - *Command line interface*
  - See User's Guide at http://goo.gl/9xukB

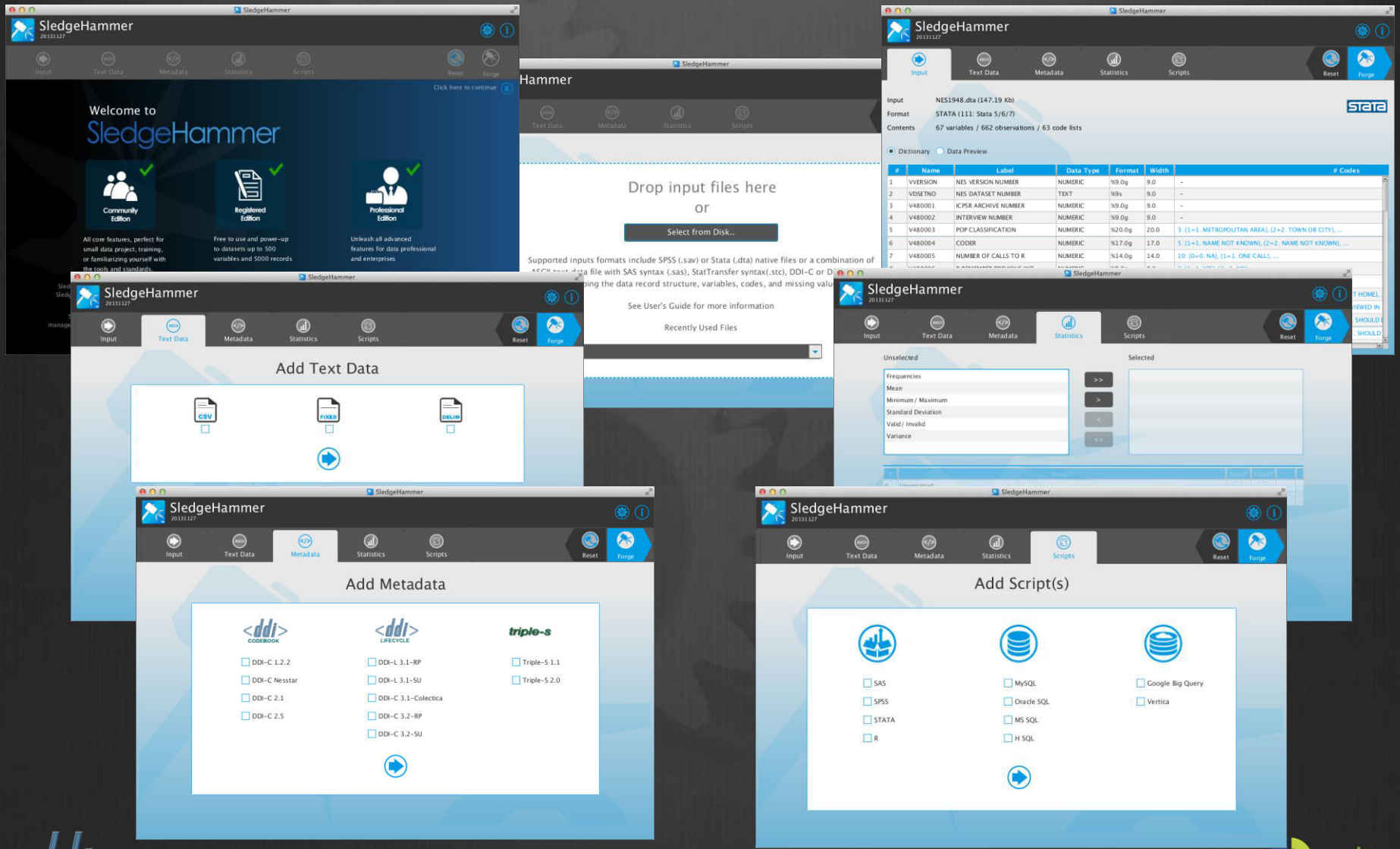
DataForge SledgeHammer

- Key changes since IASSIST 2013
  - **Desktop User Interface**
  - SAS (Win+DSRead), handling of missing values, DDI 3.2, lots of tweaks
- Editions
  - Freeware
    - Community: 100 variables / 1,000 obs.
    - Registered (free): 500 variables / 5,000 obs. (or contact us)
    - Available today (Candidate Release)
  - Pro/Enterprise (1Q2014)
    - No data size limits
    - Weighted stats, save stats, classification auto-harmonizer, batch mode, add database formats & star schema
    - Subscription: $29/3-month or $99/year , Perpetual: $279
  - LavaCore:
    - Currently internal only, planning for OEM

DataForge SledgeHammer

- Future Features (demand driven)
  - Support additional input/output formats
  - Additional ID generators
- DataWeaver (data transformation)
  - Produce data subset
  - Synthetic data generators
  - Disclosure control
- DataForge Online (2014)
  - Free and pay as you go services
- DataForge Services (available today)
  - Community support in Google groups
  - MTNA: pay as you go / annual support, custom developments
  - IDMS: data/metadata management / processing

DataForge
SledgeHammer



Products

SledgeHammer

Data/metadata Management (free/paid)

Caelum

Documentation XML Transform (free)

DataWeaver

Data subsets, derivation, transforms (internal/project)

Asmurex

ASCII multi record Extraction (free)

SledgeHammer--BI

Integration with Business Intelligence (internal/projects)

Survey Manager

DDI 3.1 Editor for simple projects or training (OSS)

DMP Editor

Simple editor for Data Management Plan (OSS)

CMS

Classification Management System (2014Q3 / paid)

Services

Expertise

Expert Advisory and Technical Assistance

- Establishment of corporate data management strategies
- Adoption of metadata standards and related best practices
- Development of technical specifications for data management infrastructure
- Statistical and scientific data management technologies

Development

Application Development

- Statistical and Scientific data management solutions
- Solutions include production, archiving, dissemination, analysis
- Web-based data delivery solutions (web services, portals, secure environments)
- Standards-based data management tools/utilities

Training

Training in and Support for:

- Data Documentation Initiative (DDI)
- Statistical Data and Metadata Exchange (SDMX)
- Generic Statistical Business Process Model (GSBPM)
- Generic Statistical Information Model (GSIM)
- General data and metadata management support

Technologies

Leverage New Technologies

- Help introduce new state-of-the-art technologies to the community and customers
- Deployment, hosting, and management of cloud-based infrastructures

http://www.openmetadata.org/dataforge/sledgehammer

DEMO