

Report to DDI Alliance Scientific Board: SDTL Working Group

March 3, 2021

Current activities:

The SDTL Working Group is charged with maintaining Structured Data Transformation Language, which was added to the DDI Alliance suite of standards in December 2020.

Structured Data Transformation Language (SDTL) is an independent intermediate language for representing data transformation commands from statistical analysis packages (e.g., SPSS, Stata, SAS, and R). SDTL was designed to be integrated into DDI metadata files to provide machine-actionable descriptions of the provenance of variables and data files. SDTL is compatible with all versions of DDI, and it has applications in the DDI-Cross Domain Integration standard. SDTL was created by the Continuous Capture of Metadata (C²Metadata) Project, which was funded by the US National Science Foundation.

An open comment period for SDTL was held in July and August 2020. The DDI Alliance membership voted to adopt SDTL in November, and the SDTL product page was launched in December. An introduction to SDTL was published in the IASSIST Quarterly in December (<http://hdl.handle.net/2027.42/156015>).

In preparation for the comment period, a full range of documentation for SDTL was developed and provided through the COGS system. The SDTL standard was already managed in COGS, and an SDTL User Guide was added. The User Guide includes both an overview of SDTL and a chapter on “SDTL Best Practices and Conventions.”

SDTL Working Group meets monthly. [Notes](#) from WG meetings are available on the DDI Alliance Confluence system.

Since SDTL 1.0 was released, there have been a few small additions to SDTL (i.e., elements for date-time constants) and several adjustments to the documentation. These changes are all documented on the [SDTL Working Group Confluence site](#).

Public presentations:

- A webinar introducing SDTL for the benefit of DDI Alliance members was presented on November 2, 2020. A [recording](#) is available.
- SDTL was included in a session on “FAIR Provenance” at the CODATA/GO FAIR Convergence Symposium on December 2, 2020.

- A paper proposal has been submitted for the International Digital Curation Conference, which will be held April 19, 2021.
- A paper proposal has been submitted for the Provenance Week 2021, July 19-22.

Future activities:

We expect that there will be minor adjustments to the SDTL standard as it is applied in new ways. Up to now, the development of SDTL has been driven by the C²Metadata Project, which ends in September 2021. We hope that the next phase will involve integration of SDTL and C²Metadata Project tools into data production workflows.

A subgroup of the SDTL Working Group is developing a grant proposal for extending SDTL to cover data created by statistical analysis commands, such as regression and factor analysis. The C²Metadata Project limited its scope to data transformation commands, but statistical analysis can produce new variables, like predicted values from regression models and indexes based on weights derived from factor analysis. The new project is being developed in partnership with the STATO ontology of statistics at the University of Oxford e-Research Centre.

SDTL Working Group

George Alter Lead
Carson Hunter
Jeremy Iverson
Chifundo Kanjala
Dara O'Neill
Ornulf Risnes
Dan Smith
Thomas Thelen
Jim Todd