

Call for Participation: OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA) TC

Submitted by censign on Wed, 2019-11-13 15:51

Type:

Call for Participation

A new OASIS Technical Committee is being formed. The OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA) Technical Committee (TC) has been proposed by the members of OASIS listed in the charter below. The TC name, statement of purpose, scope, list of deliverables, audience, IPR mode and language specified in this proposal will constitute the TC's official charter. Submissions of technology for consideration by the TC, and the beginning of technical discussions, may occur no sooner than the TC's first meeting.

The eligibility requirements for becoming a participant in the TC at the first meeting are:

(a) you must be an employee or designee of an OASIS member organization or an individual member of OASIS, and

(b) you must join the Technical Committee, which members may do by using the Roster "join group" link on the TC's web page at [a].

To be considered a voting member at the first meeting:

(a) you must join the Technical Committee at least 7 days prior to the first meeting (on or before 09 December 2019) and

(b) you must attend the first meeting of the TC, at the time and date fixed below (16 December 2019).

Participants also may join the TC at a later time. OASIS and the TC welcomes all interested parties.

Non-OASIS members who wish to participate may contact us about joining OASIS [b]. In addition, the public may access the information resources maintained for each TC: a mail list archive, document repository and public comments facility, which will be linked from the TC's public home page at [c].

Please feel free to forward this announcement to any other appropriate lists. OASIS is an open standards organization; we encourage your participation.

-----?

[a] <https://www.oasis-open.org/apps/org/workgroup/lexidma/> [1]

[b] See <http://www.oasis-open.org/join/> [2]

[c] <http://www.oasis-open.org/committees/lexidma/> [3]

?CALL FOR PARTICIPATION?

OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA) Technical Committee Charter

The charter for this TC is as follows.

Section 1: TC Charter

(1)(a) Name of the TC

Lexicographic Infrastructure Data Model and API (LEXIDMA)

(1)(b) Statement of Purpose

This committee's high level purpose is to create an open standards based framework for internationally interoperable lexicographic work. This TC will be describing and defining standard serialization independent interchange objects based predominantly on state of the art in the lexicographic industry. Defining specific serializations, transaction models, standard interfaces, and web services based on the defined objects and object models is also in scope as far as it facilitates the high level purpose set out here. This TC aims to develop the lexicographic infrastructure as part of a broader ecosystem of standards employed in Natural Language Processing (NLP), language services, and Semantic Web.

Business Benefits

The key business benefit LEXIDMA deliverables aim for is to provide a simple, modular, and easy to adopt data model that will be attractive for all lexicographic industry actors across companies and academia as well as geographic locations. Adoption of that model will facilitate exchange of lexicographic and linguistic corpus data globally and also enable effective exchange with adjacent industries such as language services, terminology management, or technical writing. Semantic interoperability of lexicographic data should help the global lexicographic industry to surpass its current model of creating and curating lexicographic deliverables (such as prominently multi- and monolingual dictionaries) and corpora in linguistically a geographically demarcated silos and create a truly global market for lexicographic data exchange across and among languages and locales.

(1)(c) Scope

The following items belong to the Scope of Work and are expected to be refined as the TC gains additional insights into evolving and culturally diverse lexicographic use cases. Members will gather insights and requirements from consultations with the wider community of industry stakeholders, annual symposia, questionnaires, etc. and use these insights to produce concrete technical deliverables.

i) Define and maintain a serialization independent Data Model for globally applicable use cases in lexicography.

ii) Define and maintain XML, JSON, RDF, and other serializations, as industry or academic needs arise, of the said lexicographic data model.

iii) Define specific standard Application Interfaces (API) and abstract service architectures for various serializations of the lexicographic data model in concert with other related standards and formats (such as TEI, LMF, RDF, JSON-LD, XLIFF, ITS, TBX, etc.) and prominent data models in adjacent industries and verticals, such as terminology management, translation services, web publishing, etc.

iv) Define and describe lossless or nearly lossless mappings between the lexicographic data model and its native normative serializations (developed by this committee) with other common industry and academic serializations such as, prominently, Ontolex-Lemon and TEI Lex-0, define those mappings both in an abstract way and for specific serializations as the need arises.

v) Define and describe informative best practices and abstract services architecture recommendations with regards to usage of the LEXIDMA TC normative deliverables in the lexicographic industry and adjacent industries, terminology management, translation services, web publishing, etc.

(1)(d) Deliverables

The following are high priority technical goals that should be addressed by development of one or more deliverables on OASIS standards track or as committee notes within 24 months from TC initiation:

i) Serialization independent Data Model for Lexicography (DMLex)

ii) XML serialization of DMLex

iii) JSON serialization of DMLex

iv) RDF serialization of DMLex

v) Informative Ontolex-Lemon mapping

vi) Informative TEI Lex-0 mapping

Work on the following may start during the work on addressing of the above high priority goals deliverables or later on given the general sense of urgency for those within the lexicographic industry:

vii) Reference architecture

viii) APIs with various bindings

(1)(e) IPR Mode

This TC will operate under the Non-Assertion IPR mode as defined in the OASIS Intellectual Property Rights (IPR) Policy.

(1)(f) Audience

The expected audience for the work of the LEXIDMA TC includes but is not limited to:

* Lexicographers

* Terminologists

- * Multilingual content and software architects and strategists, multilingual content publishers
- * NLP services architects and developers
- * Owners and managers of lexicographic content
- * Software providers for lexicography, corpus management, etc. including producers of language technology components
- * Technical communicators employing lexicographic tools or linguistic corpora in the process of multilingual publishing of their content
- * Translation service providers and freelance translators who need to use lexicographic tools or products in order to deliver their services

(1)(g) Language

English (UK spelling)

Section 2: Additional Information

(2)(a) Identification of Similar Work

Ontolex-Lemon <https://www.w3.org/2016/05/ontolex/> [4]

TEI Lex-0 <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html> [5]

ISO/TC 37/SC 4 Language resource management ? Lexical markup framework (LMF) ? [multipart]

LEXIDMA TC aims to establish informal liaisons with Ontolex-Lemon and TEI Lex-0 communities as well as formal liaisons with ISO/TC 37 and its subcommittees, in particular SC 2, SC 3, SC 4, and SC 5.

ISO fast tracking through TC 37 or one of the TC 37 SCs will be considered.

(2)(b) First TC Meeting

First TC meeting is planned as a webconference to be held on 16th December 2019 1600 UTC and the GoToMeeting webconferencing facility will be provided by IJS.

(2)(c) Ongoing Meeting Schedule

The TC aims to hold monthly webconferences to be hosted by IJS on their GoToMeeting facility. Meeting frequency will be adjusted when deliverables go for public reviews, OASIS approval etc. A limited number of face-to-face meetings is likely to resolve public review comments and similar. Any such meetings will be announced well in advance to allow membership to make travel plans.

(2)(d) TC Proposers

Simon Krek, Jožef Stefan Institute (IJS), simon.krek@ijs.si [6]

Tomaž Erjavec, Jožef Stefan Institute (IJS), tomaz.erjavec@ijs.si [7]

Iztok Kosem, Jožef Stefan Institute (IJS), iztok.kosem@ijs.si [8]

Miloš Jakubiček, Individual Member, milos.jakubicek@sketchengine.co.uk [9]

Ilan Kernerman, Individual Member, ilan@kictionaries.com [10]

David Filip, Trinity College Dublin (ADAPT), david.filip@adaptcentre.ie [11]

Patrick Durusau, Individual Member, patrick@durusau.net [12]

(2)(e) Primary Representatives' Statements of Support

?I, Simon Krek (simon.krek@ijs.si [6]), as OASIS primary representative for Jo?ef Stefan Institute, confirm our support for the proposed LEXIDMA TC charter and endorse our participants listed above.?

?I, Dave Lewis (dave.lewis@adaptcentre.ie [13]), as OASIS primary representative for Trinity College Dublin (ADAPT), confirm our support for the proposed LEXIDMA TC charter and endorse our participants listed above.?

(2)(f) TC Convener

David Filip, Trinity College Dublin (ADAPT), david.filip@adaptcentre.ie [11]

(2)(g) OASIS Member Section

N/A

(2)(h) Anticipated Contributions

ELEXIS Consortium (elex.is) plans to submit their lexicography exchange data model as the initial input for the DMLex deliverable.

(2)(i) FAQ Document

None

(2)(j) Work Product Titles and Acronyms

Data Model for Lexicography (DMLex), Version 1.0

Associated TC:

Lexicographic Infrastructure Data Model and API (LEXIDMA)

Deadline:

Wed, 2019-11-13 - Mon, 2019-12-16

Links:

[1] <https://www.oasis-open.org/apps/org/workgroup/lexidma/>

[2] <http://www.oasis-open.org/join/>

[3] <http://www.oasis-open.org/committees/lexidma/>

[4] <https://www.w3.org/2016/05/ontolex/>

[5] <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

[6] <mailto:simon.krek@ijs.si>

[7] <mailto:tomaz.erjavec@ijs.si>

[8] <mailto:iztok.kosem@ijs.si>

[9] <mailto:milos.jakubicek@sketchengine.co.uk>

- [10] <mailto:ilan@kictionaries.com>
- [11] <mailto:david.filip@adaptcentre.ie>
- [12] <mailto:patrick@durusau.net>
- [13] <mailto:dave.lewis@adaptcentre.ie>