

## LDC Data License Agreement for: LoReHLT 2018 Evaluation

In the remainder of this document the term **User** refers to \_\_\_\_\_ of

\_\_\_\_\_ and the term **User's Research Group** refers to:

\_\_\_\_\_.

**User** agrees, on behalf of **User's Research Group**, to receive media (CD-ROM, DVD, hard drive, web download, etc.) containing speech and/or text data from the Linguistic Data Consortium (LDC) named below under "Corpora/Data Received" and to use the material received under this agreement (the "Data") only for purposes of the LoReHLT 2018 Evaluation. **User** and **User's Research Group** may include limited excerpts from the Data in articles, reports and other documents describing the results of work performed in the LoReHLT 2018 Evaluation. **User** further agrees not to otherwise publish, retransmit, disclose, display, copy, reproduce or redistribute the Data to others outside of **User's Research Group**. **User** further agrees to comply with the conditions for using IARPA BABEL language packs set forth below.

**User** agrees that its use of any Twitter material contained in the Data is governed by the applicable terms and conditions of Twitter's Terms of Service, <https://twitter.com/tos>, Twitter's Developer Agreement, <https://developer.twitter.com/en/developer-terms/agreement> and the Twitter Privacy Policy, <https://twitter.com/privacy>.

**User** shall return the Data to LDC as follows: (1) if **User** withdraws from the LoReHLT 2018 Evaluation, **User** agrees to delete the Data and any files and software derived from it from any computer or media onto which it has been copied and to return all media to LDC **on or before June 28, 2018**; (2) if **User** fails to submit evaluation results fully and on time, **User** agrees to delete the Data and any files and software derived from it from any computer or media onto which it has been copied and to return all media to LDC **on or before September 30, 2018**; (3) if **User** fails to send a representative to the LoReHLT 2018 Evaluation workshop, **User** agrees to delete the Data and any files and software derived from it from any computer or media onto which it has been copied and to return all media to LDC **on or before October 31, 2018 (not applicable if no workshop is held)**; and (4) if **User** fulfills all requirements of the LoReHLT 2018 Evaluation, **User** agrees to delete the Data and any files and software derived from it from any computer or media onto which it has been copied and to return all media to LDC **on or before November 30, 2018**. **User** may retain the data available in LDC's general catalog by agreeing to license the data as a nonmember or by joining LDC.

All Data is provided "as is" and neither the Linguistic Data Consortium nor the University of Pennsylvania warrants the accuracy, completeness, currentness, merchantability or fitness for a particular purpose of the Data. In no event will the Linguistic Data Consortium or the University of Pennsylvania be liable for any loss or injury caused in whole or in part by its negligence or contingencies beyond its control in procuring, compiling, interpreting, editing, writing, reporting or delivering the Data, or for any errors, omissions or inaccuracies in the Data, regardless of how caused. In no event will the Linguistic Data Consortium or the University of Pennsylvania be liable to User, User's Research Group, or to any other individual or organization for any decision

made or action taken in reliance upon any part of the Data or for any consequential, direct, special or similar damages, even if advised of the possibility of such damages.

**User** designates \_\_\_\_\_ as the data contact person under this agreement. The data contact person shall receive all material distributed by LDC under this agreement. The data contact person is responsible for distributing such material to **User's Research Group**.

User shall send a signed copy of this agreement to LDC: (1) by e-mail to [ldc@ldc.upenn.edu](mailto:ldc@ldc.upenn.edu) or (2) by facsimile, Attention: Membership Office, fax number (+1) 215-573-2175.

### **Conditions for Using IARPA BABEL Language Packs**

LDC will provide **User** with the following dataset(s):

LDC2016S02 IARPA Babel Cantonese Language Pack  
LDC2016S06 IARPA Babel Assamese Language Pack  
LDC2016S08 IARPA Babel Bengali Language Pack  
LDC2016S09 IARPA Babel Pashto Language Pack  
LDC2016S10 IARPA Babel Turkish Language Pack  
LDC2016S13 IARPA Babel Tagalog Language Pack  
LDC2017S01 IARPA Babel Vietnamese Language Pack  
LDC2017S03 IARPA Babel Haitian Creole Language Pack  
LDC2017S08 IARPA Babel Lao Language Pack  
LDC2017S13 IARPA Babel Tamil Language Pack  
LDC2017S22 IARPA Babel Kurmanji Kurdish Language Pack  
LDC2017S19 IARPA Babel Zulu Language Pack  
LDC2018S02 IARPA Babel Tok Pisin Language Pack  
LDC2016E15 IARPA Babel Cebuano Language Pack  
LDC2016E16 IARPA Babel Kazakh Language Pack  
LDC2016E17 IARPA Babel Telugu Language Pack  
LDC2016E18 IARPA Babel Lithuanian Language Pack

Please see the README in each delivery archive for more details on each data set (hereinafter referred to as "the IARPA BABEL data").

As a recipient of the IARPA BABEL data, **User** agrees to:

1. Insert the following statement in any product, report, publication, presentation, and/or other document that references the IARPA BABEL data: "This product contains or makes use of IARPA data <INSERT DATA SET NAME(S)>."

2. Accept the data "as is." **User** is solely responsible for any damage that may arise from its use of the IARPA BABEL data. **User** agrees to hold the US Government ("USG"), the source of the data, harmless and to indemnify the USG for all liabilities, demands, damages, expenses and losses arising out of its use for any purpose of the data. Unless prohibited by law, **User** assumes all liability for claims for damages against it by third parties which may arise from the use, storage, or disposal of the data, regardless of whether such liability is based on breach of contract, tort,

strict liability, breach of warranties, infringement of intellectual property, failure of essential purpose or otherwise.

3. Not further distribute the data.

LDC has entered into a License Agreement with IARPA to distribute the IARPA BABEL data. Under the terms of that License Agreement, LDC may be required to identify recipients of the IARPA BABEL data to IARPA.

**Corpora and/or Data Received:**

**MT-ONLY RESOURCES**

LDC2011T07 English Gigaword Fifth Edition  
LDC2008L03 Global Yoruba Lexical Database v. 1.0  
LDC2015T15 TS Wikipedia  
LDC2009L01 An English Dictionary of the Tamil Verb Second Edition  
LDC2015T11 2006 CoNLL Shared Task - Ten Languages  
LDC2015L01 SenSem Lexicons  
LDC2012T03 2009 CoNLL Shared Task Part 1  
LDC2011T12 Spanish Gigaword Third Edition  
LDC2009T25 Web 1T 5-gram, 10 European Languages  
LDC99T41 Spanish Newswire Text, Volume 2  
LDC95T9 Spanish News Text  
LDC94T4A UN Parallel Text (Complete)  
LDC2013T06 1993-2007 United Nations Parallel Text  
LDC2012T23 Russian-English Computer Security Parallel Text  
LDC94T5 ECI Multilingual Text  
LDC2008T01 Hungarian-English Parallel Text, Version 1.0  
LDC2010T24 Indian Language Part-of-Speech Tagset: Hindi  
LDC2008L02 Hindi WordNet  
LDC2010T16 Indian Language Part-of-Speech Tagset: Bengali  
LDC2010T10 NIST 2002 Open Machine Translation (OpenMT) Evaluation  
LDC2010T11 NIST 2003 Open Machine Translation (OpenMT) Evaluation  
LDC2010T12 NIST 2004 Open Machine Translation (OpenMT) Evaluation  
LDC2010T14 NIST 2005 Open Machine Translation (OpenMT) Evaluation  
LDC2010T17 NIST 2006 Open Machine Translation (OpenMT) Evaluation  
LDC2010T21 NIST 2008 Open Machine Translation (OpenMT) Evaluation  
LDC2010T23 NIST 2009 Open Machine Translation (OpenMT) Evaluation  
LDC2013T03 NIST 2012 Open Machine Translation (OpenMT) Evaluation  
LDC2013T07 NIST 2008-2012 Open Machine Translation (OpenMT) Progress Test Sets  
LDC2014T02 NIST 2012 Open Machine Translation (OpenMT) Progress Test Five Language Source

**EDL-ONLY RESOURCES**

LDC2017E03 TAC KBP Entity Discovery and Linking Comprehensive Training and Evaluation Data 2014-2016  
LDC2014T16 TAC KBP Reference Knowledge Base  
LDC2015E42 TAC KBP 2015 Tri-Lingual Entity Discovery and Linking Knowledge Base

LDC2016E63 TAC KBP 2016 Evaluation Source Corpus  
LDC2014E30 TAC 2014 KBP Spanish Source Corpus  
LDC2017T17 TAC KBP Chinese Cross-lingual Entity Linking - Comprehensive Training and Evaluation Data 2011-2014  
LDC2016T26 TAC KBP Spanish Entity Linking - Comprehensive Training and Evaluation Data 2012-2014  
LDC2015E19 TAC KBP English Entity Linking - Comprehensive Training and Evaluation Data 2009-2013  
LDC2018T03 TAC KBP Comprehensive English Source Corpora 2009-2014  
LDC2017E25 TAC KBP 2017 Evaluation Source Corpus  
LDC2017E51 TAC KBP 2017 Evaluation Core Source Corpus  
LDC2017E52 TAC KBP 2017 Entity Discovery and Linking Evaluation Gold Standard Entity Mentions and Knowledge Base Links

### **SF-ONLY RESOURCES**

LDC2017E07 LORELEI Situation Frame Exercise Annotation  
LDC2016E119 LORELEI IL3 Dev Speech Database  
LDC2016E120 LORELEI IL3 Eval Speech Database  
LDC2016E121 LORELEI IL3 Eval Speech Annotation Database  
LDC2016E108 LORELEI\_Mandarin\_Speech\_Database.zip  
LDC2016E109 LORELEI\_Turkish\_Speech\_Database.zip  
LDC2016E110 LORELEI\_Hausa\_Speech\_Database.zip  
LDC2016E111 LORELEI\_Russian\_Speech\_Database.zip  
LDC2016E113 LORELEI\_Amharic\_Speech\_Database.zip  
LDC2016E123 LORELEI\_Arabic\_Speech\_Database.zip  
LDC2016E124 LORELEI\_Farsi\_Speech\_Database.zip  
LDC2016E125 LORELEI\_Hungarian\_Speech\_Database.zip  
LDC2016E126 LORELEI\_Somali\_Speech\_Database.zip  
LDC2016E127 LORELEI\_Spanish\_Speech\_Database.zip  
LDC2016E128 LORELEI\_Vietnamese\_Speech\_Database.zip  
LDC2016E129 LORELEI\_Yoruba\_Speech\_Database.zip  
LDC2016E112 LORELEI IL4 Speech Database  
LDC2017E33\_LORELEI\_IL2\_Dry\_Run\_Speech\_Data.tgz  
LDC2017E50\_LORELEI\_US\_English\_Speech\_Database.zip  
LDC2017E84\_LORELEI\_Akan\_Speech\_Database.zip  
LDC2017E85\_LORELEI\_Wolof\_Speech\_Database.zip  
LDC2017E86\_LORELEI\_Swahili\_Speech\_Database.zip  
LDC2017E87\_LORELEI\_Tamil\_Speech\_Database.zip  
LDC2017E88\_LORELEI\_Hindi\_Speech\_Database.zip  
LDC2017E89\_LORELEI\_Tagalog\_Speech\_Database.zip  
LDC2017E90\_LORELEI\_Thai\_Speech\_Database.zip  
LDC2017E91\_LORELEI\_Indonesian\_Speech\_Database.zip  
LDC2017E92\_LORELEI\_Bengali\_Speech\_Database.zip  
LDC2017E93\_LORELEI\_Zulu\_Speech\_Database.zip

### **RESOURCES FOR ALL TASKS (MT, EDL, SF)**

LDC2015E14 Urdu Language Pack from REFLEX Program  
LDC2018E37 LoReHLT Turkish Representative Language Pack

LDC2018E36 LoReHLT Hausa Representative Language Pack  
LDC2018E38 LoReHLT Uzbek Incident Language Pack  
LDC2018E39 LoReHLT Uzbek Representative Language Pack  
LDC2018T04 LORELEI Amharic Representative Language Pack - Monolingual and Parallel Text  
LDC2018E11 LoReHLT Amharic Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E88 LORELEI Arabic Representative Language Pack Monolingual Text  
LDC2018E12 LoReHLT Arabic Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018T11 LORELEI Somali Representative Language Pack - Monolingual and Parallel Text  
LDC2018E20 LoReHLT Somali Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E92 LORELEI Farsi Representative Language Pack Monolingual Text  
LDC2018E14 LoReHLT Farsi Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E94 LORELEI Russian Representative Language Pack Monolingual Text  
LDC2018E19 LoReHLT Russian Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E96 LORELEI Spanish Representative Language Pack Monolingual Text  
LDC2018E21 LoReHLT Spanish Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E98 LORELEI Hungarian Representative Language Pack Monolingual Text  
LDC2018E16 LoReHLT Hungarian Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E40 LoReHLT Mandarin Incident Language Pack  
LDC2016E100 LORELEI Mandarin Representative Language Pack Monolingual Text  
LDC2018E18 LoReHLT Mandarin Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E102 LORELEI Vietnamese Representative Language Pack Monolingual Text  
LDC2018E26 LoReHLT Vietnamese Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2016E104 LORELEI Yoruba Representative Language Pack Monolingual Text  
LDC2018E28 LoReHLT Yoruba Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E42 LoReHLT IL4 Incident Language Pack  
LDC2018E41 LoReHLT18 IL3 Incident Language Pack for Year 1 Eval Unsequestered Data  
LDC2018E43 LoReHLT IL5 Incident Language Pack Training Data  
LDC2018E44 LoReHLT IL6 Incident Language Pack Training Data  
LDC2017E59 LORELEI Bengali Representative Language Pack Monolingual Text  
LDC2018E13 LoReHLT Bengali Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2017E61 LORELEI Hindi Representative Language Pack Monolingual Text  
LDC2018E15 LoReHLT Hindi Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2017E63 LORELEI Swahili Representative Language Pack Monolingual Text  
LDC2018E22 LoReHLT Swahili Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2017E65 LORELEI Indonesian Representative Language Pack Monolingual Text

LDC2018E17 LoReHLT Indonesian Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2017E67 LORELEI Tagalog Representative Language Pack Monolingual Text  
LDC2018E23 LoReHLT Tagalog Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2017E69 LORELEI Tamil Representative Language Pack Monolingual Text  
LDC2018E24 LoReHLT Tamil Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E02 LORELEI Thai Representative Language Pack Monolingual Text  
LDC2018E25 LoReHLT Thai Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E04 LORELEI Zulu Representative Language Pack Monolingual Text  
LDC2018E29 LoReHLT Zulu Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E06 LORELEI Akan Representative Language Pack Monolingual Text  
LDC2018E10 LoReHLT Akan Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools  
LDC2018E08 LORELEI Wolof Representative Language Pack Monolingual Text  
LDC2018E27 LoReHLT Wolof Representative Language Pack Translation, Annotation, Grammar, Lexicon and Tools

**IARPA BABEL LANGUAGE PACKS (For all tasks: MT, EDL, SF)**

LDC2016S02 IARPA Babel Cantonese Language Pack  
LDC2016S06 IARPA Babel Assamese Language Pack  
LDC2016S08 IARPA Babel Bengali Language Pack  
LDC2016S09 IARPA Babel Pashto Language Pack  
LDC2016S10 IARPA Babel Turkish Language Pack  
LDC2016S13 IARPA Babel Tagalog Language Pack  
LDC2017S01 IARPA Babel Vietnamese Language Pack  
LDC2017S03 IARPA Babel Haitian Creole Language Pack  
LDC2017S08 IARPA Babel Lao Language Pack  
LDC2017S13 IARPA Babel Tamil Language Pack  
LDC2017S22 IARPA Babel Kurmanji Kurdish Language Pack  
LDC2017S19 IARPA Babel Zulu Language Pack  
LDC2018S02 IARPA Babel Tok Pisin Language Pack  
LDC2016E15 IARPA Babel Cebuano Language Pack  
LDC2016E16 IARPA Babel Kazakh Language Pack  
LDC2016E17 IARPA Babel Telugu Language Pack  
LDC2016E18 IARPA Babel Lithuanian Language Pack

**EVALUATION DATA**

LDC208ETBD LoReHLT IL5 OTAL Test Data  
LDC2018ETBD LoReHLT IL6 OTAL Test Data  
LDC2018ETBD LoReHLT IL9 Incident Language Pack  
LDC2018ETBD LoReHLT IL10 Incident Language Pack  
LDC2018ETBD IL9 Speech Data  
LDC2018ETBD IL10 Speech Data

Organization: \_\_\_\_\_

Name: \_\_\_\_\_

Mailing address: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Phone: \_\_\_\_\_

Email: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

For LDC:

Christopher Cieri  
Executive Director