NIST Open Machine Translation 2015
Evaluation Plan (OpenMT15)

1 INTRODUCTION

The 2015 NIST Open Machine Translation evaluation (OpenMT15) continues the ongoing series of evaluations of human language translation technology. NIST conducts these evaluations in order to support MT research and help advance the state of the art in MT technology. To do this, NIST:

- Defines a set of translation tasks,
- Collaborates with the data providers[1] to provide corpus resources to support research on these tasks,
- Creates and administers formal evaluations of MT technology,
- Provides evaluation utilities to the MT community, and
- Coordinates workshops to discuss MT research findings and results of task performance in the context of these evaluations.

These evaluations provide an important contribution to the direction of research efforts and the calibration of technical capabilities. They are intended to be of interest to all researchers working on the general problem of translating between human languages. To this end, the evaluations are designed to be simple, to focus on core technology issues, to be fully supported, and to be accessible to those wishing to participate.

The 2015 evaluation is a pilot evaluation focusing on exploring translation system limitations as well as measurement limitations on informal data genres. Highlights of OpenMT15 include:

- Evaluation on informal data genres SMS/Chat and Conversational Telephone Speech (CTS) for Arabic-to-English and Chinese-to-English,
- Inclusion of audio input track, and
- Explore common MT measurement techniques on these informal data genres.

Participation in the evaluation is invited for all researchers who find the tasks and the evaluation of interest. There is no fee for participation. However, participation in the evaluation requires participation in the follow-up workshop.[2] All OpenMT15 participants must attend the evaluation workshop and be prepared to discuss their system(s), results, and their research findings in detail. This workshop is restricted to the group of registered OpenMT15 participants, data providers and representatives of supporting government agencies.

To participate in the evaluation, sites must officially register with NIST[3] and agree to the terms specified in the registration form. For more information, visit the NIST OpenMT15 website.[4]

2 TRAINING CONDITIONS

MT R&D requires language data resources. System performance and R&D effort are strongly affected by the type and amount of resources used. Therefore, OpenMT15 has two different resource categories as conditions of evaluation. They differ solely by the specification of the data that may be used for system training and development. These evaluation conditions are *Constrained Training* and *Unconstrained Training*, as implemented in previous OpenMT evaluations.

As in previous OpenMT evaluations, training data are provided by the LDC. All participants are required to sign a license agreement[5] governing the use of LDC's data resources available for system development in preparation for OpenMT15. Participants must fully comply with all requirements that are (1) stated in this evaluation plan, (2) stated on the registration form, and (3) stated on the LDC license agreement, in order to retain rights to data obtained under the LDC license agreement.

2.1 CONSTRAINED TRAINING

Systems entered in the Constrained Training condition allow for direct comparisons of different algorithmic approaches. System development must adhere to the following restrictions:

Only data listed in the LDC data license agreement may be used for core MT engine development in the constrained training condition. OpenMT15 does not place a language specific restriction on the LDC data resources; that is, a site participating in Arabic to English may use Chinese to English data as long as that data is listed in the LDC data license.

[1] http://www.ldc.edu http://www.sdl.com/research/language-technology

[2] There is a registration fee associated with attending the evaluation workshop. This fee does not include travel or accommodation expenses.

[3] http://www.nist.gov/itl/iad/mig/upload/OpenMT15_Registration.pdf

[4] http://www.nist.gov/itl/iad/mig/openmt15.cfm

[5] http://www.nist.gov/itl/iad/mig/upload/OpenMT15_LDCAgreement.pdf

Resources that assist the core engine (such as segmenters, tokenizers, parsers, or taggers) are not subject to the same restriction. If such additional resources are used, they must be listed in the system description.

2.2 UNCONSTRAINED TRAINING

Systems entered in the Unconstrained Training condition may demonstrate the gains achieved by adding data from other sources. This training condition allows for more creativity in system development. System development must adhere to the following restrictions:

Data must be publicly available, at least in principle.[6] This ensures that research results are broadly applicable and accessible to all participants. Participants must specify in their system descriptions what data they used.

3 DATA SETS

3.1 EVALUATION DATA

The OpenMT15 evaluation data will consist of Arabic and Chinese sources. The Arabic data is primarily Egyptian dialect but may have other dialect(s) and/or other language(s). The Chinese data is primarily Mandarin Chinese but may have other dialect(s) and/or other language(s). Each language pair and genre combination has approximately 25K source words. A subset of the 25K (about 5K) reference will have HyTER network annotations.

Table 1: Data volume for OpenMT15 test sets.

Language Pair	Genre	Volume (words)
Arabic-to-English	SMS/Chat	~25,000
	CTS	~25,000
Chinese-to-English	SMS/Chat	~25,000
	CTS	~25,000

Each language pair and genre has one gold standard reference. Additionally, HyTER network will be created for approximately 5,000 words for each language pair and genre.

3.2 DEVELOPMENT DATA

OpenMT15 participants will receive a development data set to validate their systems' capability. This set contains corpora with "BOLT Phase2/Phase3 DevTest" in their name as listed in the LDC data license. Sites should not use this data for training. It is up to the sites to divide this data into Dev-Tune and Dev-Test sets.

4 INPUT TRACKS

OpenMT15 will offer two input tracks: audio and text. If the audio track is chosen, participants are required to process both tracks.

4.1 AUDIO SOURCE INPUT

For the audio input track, system will process from the audio recording of the telephone conversations. Segmentation will be given.

4.2 TEXT SOURCE INPUT

For the text input track, system will process the SMS/chat messages as well as the human reference transcripts of the telephone conversations.

The Arabic SMS/chat data may contain a mixture Arabic script, Arabizi, and/or script from other languages, and the human reference transcripts of the Arabic CTS data contain Arabic script and/or script from other languages. The Chinese SMS/chat data may contain a mixture of Chinese characters, pinyin, and/or other languages, and the human reference transcripts of the Chinese CTS data contain Chinese characters and/or script from other languages.

5 PRIMARY AND CONTRASTIVE SUBMISSIONS

OpenMT15 allows participants to submit exactly one primary system submission for each language pair, input track, and training condition combination. There is no limit (within reason) on the number of contrastive submissions.

At the time of submission exactly one system must be identified as the primary system, for each given language pair, input track, and training condition combination. Only primary systems will be compared and contrasted across sites in NIST's reporting of results.

Contrastive systems are encouraged to test significant alternatives to the primary system. NIST discourages contrastive entries that represent mere tweaks and minor parameter setting differences.

[6] Data limited to government use, such as the FBIS data, is deemed to be not publicly available and not admissible for system development.

6 PERFORMANCE MEASUREMENT

OpenMT15 will use several automatic metrics and, time/resource permitting, will investigate several semi-automatic metrics as well as human assessments of system translations to understand measurement limitations on informal data genres. Unlike previous years, there will be no official primary metric. The results of the scoring techniques will be included as part of the public release of results.

6.1 METRICS

Metrics under consideration are:

- BLEU[7] – This technique scores a translation according to the N-grams that it shares with one or more reference translations of high quality. In essence, the more co-occurrences, the better the translation. An N-gram, in this context, is simply a *case sensitive* sequence of N tokens. (Words and punctuation are counted as separate tokens.) NIST will compute case-sensitive BLEU scores using NIST's publicly available *mteval* software[8].
- METEOR[9] – This technique scores a translation according to word-to-word matches between the system and reference translations but also includes a set of language specific weights tuned to the target language and the ability to incorporate stemming and synonymy.
- TER[10] – This technique scores a translation according to the number of transformations that must be performed on the system translation such that it has the same word ordering as the reference translation. NIST will compute case-sensitive TER scores using UMD's publicly available tercom software[11]. NIST may also compute human-targeted version of TER (HTER).
- HyTER[12] – This technique is similar to TER but makes use of large networks of reference translations.

6.2 HUMAN ASSESSMENTS

Human assessment will be performed on the translations of selected systems. The assessment will look at the operational usefulness of the translation. The assessment will use a five-point scale:

- **Usable (3)** – The translation is usable if it can be use as is, to understand the meaning of the source text or reference translation. A usable translation does not have to be grammatically perfect but should be easy to understand.
- **Editable (2)** –The translation is editable if it requires minor editing to understand the meaning of the source text or reference translation. It is up to you to determine what is minor (e.g., rewriting an entire sentence is not considered minor). In addition even without any edits, the translation is not misleading.
- **Characterizable (1)** –The translation is characterizable if one can only state the topic of what was said. The translation lacks supporting details conveyed in the source text or reference translation. Despite this, the translation is not misleading.
- **Unusable (0)** – The translation is unusable if it is not usable in any way because it is unintelligible or completely unrelated to the source text or reference translation.
- **Misleading (-1)** – The translation is misleading if it causes you to have a completely opposite idea or impression of what was said which, in certain situations, may lead to disastrous consequence.

7 EVALUATION RULES

The OpenMT15 evaluation is an open evaluation where the test data is sent to the participants who will process and submit the output to NIST. As such, the participants have agreed to process the data in accordance with the following rules:

- Investigation of the evaluation data prior to submission of all systems' output is not allowed. Both human and automatic probing is prohibited to ensure that all participating systems have the same amount of information on the evaluation data.
- No training or tuning to the development data is allowed.
- Those who participate in the audio track must process the audio data before the text track. Once the text track data is released, no submission for the audio track will be accepted.

---

[7] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu (2001). "Bleu: a Method for Automatic Evaluation of Machine Translation". This report may be downloaded from URL http://domino.watson.ibm.com/library/CyberDig.nsf/home (keyword RC22176).

[8] ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v13a-20091001.tar.gz

[9] Michael Denkowski and Alon Lavie, "Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems", Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation, 2011.

[10] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul, "A Study of Translation Edit Rate with Targeted Human Annotation," Proceedings of Association for Machine Translation in the Americas, 2006.

[11] http://www.cs.umd.edu/~snover/tercom/

[12] Dreyer, Markus, and Daniel Marcu. "Hyter: Meaning-equivalent semantics for translation evaluation." Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2012.

8 EVALUATION PROCEDURES

The OpenMT15 evaluation process includes a number of mandatory steps; please see the schedule in Section 14 for the dates for each of these:

1. Register to participate. Each site electing to participate in the evaluation must register with NIST.
2. Sign LDC's data license agreement and return it to LDC. **Even if not selecting any training data, participants must sign the agreement to receive the evaluation data, which are listed on the agreement.**
3. Sign up for an evaluation account. See Section 10.
4. Download the dry run source data from NIST according to instructions given in Section 10.
5. Perform the dry run translation. Each site must run its translation system(s) on the entire dry run set for each language pair attempted.
6. Upload the dry run translations to NIST according to instructions given in Section 10.
7. Download the evaluation source data from NIST. Source data will be available for download at the beginning of the evaluation period. Inspection and manipulation of the evaluation data are prohibited.
8. Perform the evaluation translation. Each site must run its translation system(s) on the entire dry run set for each language pair attempted.
9. Upload the evaluation translations to NIST according to instructions given in Section 10.
10. View the submission status. The status of the submission will be posted in the participant's evaluation account.
11. Submit a system description (see Section 12).
12. Be ready to prepare an oral or poster presentation for the workshop; NIST will contact selected presenters in ample time before the workshop.
13. Attend the evaluation workshop. NIST sponsors a follow-up evaluation workshop where evaluation participants and government sponsors meet to review evaluation results, share knowledge gained, and plan for the next evaluation. A knowledgeable representative from each participating site is required to attend this workshop and be ready to describe their technology, research, and findings. Attendance at the workshop is restricted to evaluation participants and government sponsors of MT research.

Handling of late and debugged submissions: Scores on submissions received at NIST after the submission deadline, as well as submissions that were debugged beyond formatting errors after an initial submission, will not be listed in the official public release of results. The respective sites will be listed in the release as having participated with a late and/or debugged submission. Such submissions will be scored as time permits and may be reported at the evaluation workshop.

9 NIST OPENMT DATA FORMAT

Translation systems must be able to process the input source files and produce the translations that meet the OpenMT data format. There are two input tracks in OpenMT15 (see Section 4). The input and output data formats are described below.

9.1 TEXT SOURCE FILE FORMAT

The format for the text input track will follow a format similar to one used in previous years with one main difference: *this year each document resides in its own file while in previous years all the documents reside in one file*.

NIST has defined a set of XML tags that are used to format MT source, reference, and translation files for evaluation. All NIST OpenMT source, reference, and translation files have an *xml* extension; their format is defined by the current XML DTD.[13] NIST requires that all submitted translation files are well-formed and valid against the above-mentioned DTD.

A source file contains one single `srcset` element, immediately beneath the root `mteval` element. The `srcset` element has the following attributes:

- `setid`: The dataset.
- `srclang`: The source language. One of: Arabic, Chinese.

The `srcset` element contains one `doc` elements, which have the following attributes:

- `docid`: The document name.
- `genre`: The data genre. One of: `sms`, `chat`, `ctstext`, `ctsaudio`.

Each `doc` element contains several segments (`seg` elements). Each `segment` has a single attribute, `id`, which must be enclosed using double quotes.

One or more segments may be encapsulated inside additional elements, such as (but not limited to) `hl`, `p`, or `poster`. Only the native language text that is surrounded by a `seg` start-tag and its corresponding end-tag is to be translated.

OpenMT15 sample source file:

[13] ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-xml-v1.7.dtd

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE mteval SYSTEM "ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-xml-v1.7.dtd">
<mteval>
        <srcset setid="sample_document_1" srclang="Arabic">
                <doc docid="sample_document_1" genre="sms">
                        <seg id="1">ARABIC SENTENCE #1</seg>
                        <seg id=”2”>ARABIC SENTENCE #2</seg>
                        ...
                </doc>
        </srcset>
</mteval>
```

The source files will be named as `<base>.arz.su.xml` and `<base>.cmn.su.xml` for Arabic and Chinese source files, respectively.

9.2 AUDIO SOURCE FILE FORMAT

The format for the audio input track will include audio and segmentation files. The audio files are FLAC-compressed MS-WAV (RIFF) files. The audio content is 2 interleaved channels per file (representing the two independently recorded sides of the telephone conversation), comprised of 16-bit PCM samples at 8000 samples per second.

While the entire conversation will be provided, only portions of the recording will be evaluated. The segmentation file specifies the time regions within each audio recording to be evaluated and has the same format as the text source file as described in Section 9.1 but the content of the segment will be empty and the segment element will include four additional attributes:

- `begin`: indicates the begin time of the segment measured in seconds from the beginning of the file which is time 0
- `end`: indicates the end time of the segment measured in seconds from the beginning of the file which is time 0
- `speaker`: indicates the id of the speaker
- `channel`: indicates the channel in the audio file in which this segment resides.

The audio files will be named as `<base>.flac`

9.3 REFERENCE FILE FORMAT

A reference file contains one or more `refset` elements, immediately beneath the root `mteval` element. Each `refset` element contains the following attributes:

- `setid`: The dataset.
- `srclang`: The source language. One of: Arabic, Chinese.
- `trglang`: The target language, English.
- `refid`: The current reference.

Each `refset` element contains one document, which, in turn, contains the segments. The document elements and their subsequent child elements is exactly the same as described in Section 9.1 above for the source file.

OpenMT15 sample reference file:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE mteval SYSTEM "ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-xml-v1.7.dtd">
<mteval>
        <refset setid=" sample_document_1" srclang="Arabic" trglang="English" refid="reference01">
                <doc docid="sample_document_1" genre="sms">
                        <seg id="1">ENGLISH REFERENCE TRANSLATION #1</seg>
                        <seg id="2">ENGLISH REFERENCE TRANSLATION #2</seg>
                        ...
                </doc>
        </refset>
</mteval>
```

The reference files will be named as `<base>.eng.su.xml`.

9.4 TRANSLATION (TEST) FILE FORMAT

A translation file contains one `tstset` elements, immediately beneath the root `mteval` element. Each `tstset` element contains the following attributes:

- `setid`: The dataset.
- `srclang`: The source language. One of: Arabic, Chinese.

- `trglang`: The target language, English.
- `sysid`: A name identifying site and system (see Section 10.3.2 for requirements).

The content of each `tstset` element is exactly the same as described previously for the source file format and the reference file format.

OpenMT15 sample translation (test) file:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE mteval SYSTEM "ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-xml-v1.7.dtd">
<mteval>
    <tstset setid=" sample_document_1" srclang="Arabic" trglang="English" sysid=" NIST_ara2eng_primary_cn">
        <doc docid="sample_document_1" genre="sms">
            <seg id="1">ENGLISH SYSTEM TRANSLATION #1</seg>
            <seg id="2">ENGLISH SYSTEM TRANSLATION #2</seg>
            ...
        </doc>
    </tstset>
</mteval>
```

The translation files will be named as `<base>.eng.su.xml`. Note that the translation files must preserve the value of the id attributes (`setid, docid, id`) of the corresponding source files so that the translations can be referenced back to the source and reference translations.

10 EVALUATION INSTRUCTIONS

10.1 EVALUATION ACCOUNT

All evaluation activities will be stored in an evaluation account for each participant. Therefore, participants must sign up for an evaluation account. To sign up for an account, go to https://openmteval.nist.gov. Registered participants must use the email address provided in the registration's "main contact" field as the login to prevent unauthorized sign-up.

After successfully confirming the account, participants will be able to perform various tasks such as

- download the source data using the "Get the data" link at the top of the page
- upload your submission using the "Upload submission" link at the top of the page
- check submission status using the "Status" link also at the top of the page

10.2 SOURCE DATA DOWNLOAD

The evaluation source data (dry run and official) will be available as download via the evaluation account. Source data will be available for download at the beginning of the evaluation period. Inspection and manipulation of the evaluation data are prohibited. Sites who participate in the audio track will have the audio data mailed to them because the size of the audio is too large for web download.

10.3 SUBMISSION REQUIREMENTS

Participants may submit output from multiple systems for a given language pair, training condition, and input track. One system must be declared as primary at the time of submission and all other as contrastive. Each configuration (language pair, training condition, and input track) is considered as a single experiment and is identified by an experiment identifier (EXP-ID). See Section 10.3.1 for the format of the EXP-ID. All experiments are to reside in a single submission file. See Section 10.3.3 for the format of the submission file.

If more than one submission is made, the last submission replaces all previous submissions. Submissions that fail validation will be returned to participants for correction. Late and/or debugged submissions will be documented and scored but will not be compared to other on-time submissions in NIST's reports.

10.3.1 Experiment Identifier (EXP-ID)

The system output files are organized by experiment identifier (EXP-ID) directory. The EXP-ID directory has the following format:

`<year>_<evaltype>_<site>_<langpair>_<train>_<input>_<systype>`

Where:

- `year`: The evaluation name. This year it's `openmt15`
- `evaltype`: The evaluation being performed, referring to DryRun or Main.
  - One of: `dryrun`, `eval`
- `site`: The unique ID ***assigned by NIST*** to the site upon registration
- `langpair`: The language pair attempted in this submission
  - One of: `ara2eng`, `chi2eng`
- `train`: The training condition, referring to Constrained or Unconstrained training
  - One of: `cn`, `un`
- `input`: The input modality, referring to Audio or Text
  - One of: `audio, text`

- `systype`: The type of system of the particular submission. A primary submission must always be present.
  - One of: `primary`, `contrastX` where `X` is a positive integer `1..N`

Example of a well-formed EXP-ID directory: `openmt15_eval_nist_ara2eng_cn_audio_primary`

10.3.2 File Naming

The system output files must comply with the following naming convention:
`<base>.eng.su.xml`
Where:

- `base`: The base filename of the test file and should match the base of the input test file.

10.3.3 Submission File

All EXP-ID directories must reside inside a submission file. The submission file has the following format:

`<year>_<evaltype>_<site>_<subnum>.tgz`

Where:

- `year`, `evaltype`, and `site`: Same as described in Section 10.3.1 above
- `subnum`: The submission number. The initial submission must be `01`. If more than one submission is made, the last submission replaces all previous submissions. Subsequent submissions are to be numbered consecutively (`02`, `03`, etc.).

10.3.4 Submission Instructions

The submission for each language pair must be compressed as follows:

- Create an experiment directory for each experiment
  - `mkdir openmt15_eval_nist_ara2eng_cn_audio_primary`
  - `mkdir openmt15_eval_nist_ara2eng_cn_audio_contrast1`
- Place the system output files in the corresponding experiment directory
  - `cp <system output files> openmt15_eval_nist_ara2eng_cn_audio_primary`
  - `cp <system output files> openmt15_eval_nist_ara2eng_cn_audio_contrast1`
- Create a submission directory
  - `mkdir openmt15_eval_nist_01`
- Place all the experiment directories in the submission directory
  - `mv openmt15_eval_nist_ara2eng_cn_audio_primary openmt15_eval_nist_01`
  - `mv openmt15_eval_nist_ara2eng_cn_audio_contrast openmt15_eval_nist_01`
- Tar and zip the submission directory
  - `tar zcfv openmt15_eval_nist_01.tgz openmt15_eval_nist_01`
- Upload your submission using your evaluation account

11 DRY RUN EVALUATION

All participants are encouraged to participate in a dry run evaluation to demonstration evaluation readiness for the official evaluation. The purpose of the dry run is to iron out all of the bugs in the evaluation pipeline on both sides of the evaluation (participants and NIST). The results of the dry run will not be used in any official findings.

12 SYSTEM DESCRIPTIONS

Participants are required to submit system descriptions of the MT systems used for their submissions. Please use NIST's template[14] for system descriptions. System descriptions should be submitted in text format, and the file name should reflect the site ID.

13 GUIDELINES FOR PUBLICATION OF RESULTS

NIST Multimodal Information Group's MT evaluations follow an open model to promote interchange with the outside world. At the conclusion of the evaluation cycle, NIST will create a report that documents the evaluation. The report will be posted on the NIST web space and will identify the participants and the scores from various metrics achieved for each language pair and training condition. Results from the human assessments may also be posted.

The report that NIST creates should not be construed or represented as endorsements for any participant's system or commercial product, or as official findings on the part of NIST or the U.S. Government.

[14] http://www.nist.gov/itl/iad/mig/upload/OpenMT15_SysDescTemplate.txt

13.1 Rules Governing Publication of Evaluation Results

The rules governing the publication of NIST OpenMT15 evaluation results are similar to those used in previous years.

- Participants must refrain from publishing results and/or releasing statements of performance until the official OpenMT15 results are posted by NIST. Statements of performance may not claim winning or be perceived as a ranking amongst other participants.
- Participants are free to publish results for their own system, but participants must not publicly compare their results with other participants (ranking, score differences, etc.) without explicit written consent from the other participants. Publications should point to the NIST report as a reference.[15]
- NIST does not approve, recommend, or endorse any proprietary product or proprietary material. No reference shall be made to NIST, or to reports or results furnished by NIST in any advertising or sales promotion which would indicate or imply that NIST approves, recommends, or endorses any proprietary product or proprietary material, or which has as its purpose an intent to cause directly or indirectly the advertised product to be used or purchased because of NIST test reports or results.
- All publications must contain the following NIST disclaimer:

  NIST serves to coordinate the NIST OpenMT evaluations in order to support machine translation research and to help advance the state-of-the-art in machine translation technologies. NIST OpenMT evaluations are not viewed as a competition, as such results reported by NIST are not to be construed, or represented, as endorsements of any participant's system, or as official findings on the part of NIST or the U.S. Government.
- Linguistic resources used in building systems for OpenMT15 should be referenced in the system description. Corpora should be given a formal citation, like any other information source. LDC corpus references should adopt the following citation format:

  Author(s), Year. Catalog Title (Catalog Number). Linguistic Data Consortium, Philadelphia PA.

  For example:

  Xiaoyi Ma et al, 2005. Arabic News Translation Text Part 1 (LDC2004T17). Linguistic Data Consortium, Philadelphia PA.

14 Schedule

- February 1, 2014: Initial evaluation plan available
- February 1, 2014 – December 31, 2014: Registration period (early registration highly encouraged)
- February – December 31, 2014: Training data available from LDC with incremental releases of new genres until the end of May 2014
- November 3, 2014 – January 31, 2015: Dry run period
- February 2 – 6, 2015: Main evaluation period for audio track
  - Input available for download on February 2, 11:59am ET
  - Output due at NIST February 6, 11: 59am ET
- February 9 – 13, 2015: Main evaluation period for text track
  - Input available for download on February 9, 11:59am ET
  - Output due at NIST February 13, 11: 59am ET
- February 13, 2015: System description due (this will be used to help select systems for human assessment)
- February 23 – May 23, 2015: Human assessment period
- June 2015 (TBD): Workshop in the Washington DC area
- August 2015 (TBD): Official public release of results

[15] This restriction exists to ensure that readers concerned with a particular system's performance will see the entire set of participants and tasks attempted by all researchers.