

GALE Resource Distribution Task Specification
Version 2.0
April 1, 2008
Linguistic Data Consortium
<http://projects ldc.upenn.edu/gale/>

1. Introduction

The Resource Distribution objective is to disseminate the linguistic data collected and/or created by the LDC and GALE participants to GALE sites in a timely and efficient manner while conforming with agreements under which the LDC must operate. This data is distributed via HTTP download and via media (CD, DVD and hard drive).

2. Validation

Task-specific data quality checks for content and format are conducted before data is delivered for release. The Publications Group performs additional content checks. For media releases, each duplicated copy is automatically checked against the master copy and random manual checks are conducted on duplicated media. For web download releases, the Publications Group verifies content before preparing the download.

3. Distribution Methods

There are two principal methods for distribution depending on the size of the dataset. Each requires that recipient on the GALE distribution list and that all relevant license/user agreements be completed prior to distribution.

HTTP Distributions are downloadable from the LDC Members web site by authenticated GALE participants as soon as the LDC Membership Coordinator has been advised by the LDC GALE Project Coordinator that a dataset is ready for release. An automated notification is then sent to all relevant participants with the name and catalog number of the dataset(s) and download instructions.

Media Distributions are shipped by the LDC Membership Office to all relevant participants via expedited shipping (normally FEDEX for domestic shipments and DHL for overseas shipments).

PLEASE NOTE: The GALE license agreement provides that participants will receive media (CD-ROM, DVD, hard drive, web download, etc.) containing speech and/or text data as part of their participation in the GALE program and agrees to use the material received under this agreement only for purposes of the GALE program at a designated geographic site. After participation in the GALE program has ended, participants must delete the corpora from any computer or media onto which it has been copied and to return all media to LDC. Participants may retain data by agreeing to license the data by

joining the LDC. Participants further agree not to disclose, copy, or re-distribute the material to others outside of Participant's research group.