



D6.5: User Activity Data Community Translation Platforms

Philipp Koehn

Distribution: Public

CASMACAT
Cognitive Analysis and Statistical Methods
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D6.5



Project funded by the European Community
under the Seventh Framework Programme for
Research and Technological Development.



Project ref no.	ICT-287576
Project acronym	CASMACAT
Project full title	Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation
Instrument	STREP
Thematic Priority	ICT-2011.4.2 Language Technologies
Start date / duration	01 November 2011 / 36 Months

Distribution	Public
Contractual date of delivery	October 31, 2014
Actual date of delivery	January 7, 2015
Date of last update	January 7, 2015
Deliverable number	D6.5
Deliverable title	User Activity Data Community Translation Platforms
Type	Report
Status & version	Final
Number of pages	3
Contributing WP(s)	WP7
WP / Task responsible	UEDIN
Other contributors	
Internal reviewer	Barto Mesa
Author(s)	Philipp Koehn
EC project officer	Aleksandra Wesolowska
Keywords	

The partners in CASMACAT are:

University of Edinburgh (UEDIN)
Copenhagen Business School (CBS)
Universitat Politècnica de València (UPVLC)
Celer Soluciones (CS)

For copies of reports, updates on project activities and other CASMACAT related information, contact:

The CASMACAT Project Co-ordinator
Philipp Koehn, University of Edinburgh
10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom
pkoehn@inf.ed.ac.uk
Phone +44 (131) 650-8287 - Fax +44 (131) 650-6626

Copies of reports and other material can also be accessed via the project's homepage:
<http://www.casmacat.eu/>

© 2014, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Report

In workpackage 6, the CASMACAT workbench is integrated into community translation platforms. This allows us to collect user activity data. Since the integration will remain online for the foreseeable future, the amount of user activity data collected this way will increase and published on a regular basis.

The following Community Translation Platforms were integrated during the lifetime of the project:

- Global Voices
- TED Talks
- Wikipedia

Global Voices is the most mature of the integrations, with the most use.

The first release of community translation platform user activity data breaks down by language pairs as follows:

Language Pair	en-da	en-de	en-el	en-es	en-pt	es-de	es-en	fr-en	pt-de	Total
Documents	4	12	1	9	2	1	1	2	1	33

Data

The data is available in the CASMACAT data repository

<http://www.casmacat.eu/index.php?n=Main.Downloads>

under the link

<http://www.casmacat.eu/uploads/community-data-2014-11-06.tgz>