



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry

**Citation for published version:**

Germann, U, Barbu, E, Bentivoglio, M, Bogoychev, N, Buck, C, Caroselli, D, Carvalho, L, Cattelan, A, Cattoni, R, Cettolo, M, Federico, M, Haddow, B, Madl, D, Mastrostefano, L, Mathur, P, Ruopp, A, Samiotou, A, Sudharshan, V, Trombetti, M & van der Meer, J 2016, 'Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry' *Baltic Journal of Modern Computing*, vol 4, no. 2, pp. 397-397.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Baltic Journal of Modern Computing*

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry

U. GERMANN<sup>1</sup>, E. BARBU<sup>2</sup>, L. BENTIVOGLI<sup>3</sup>, N. BERTOLDI<sup>3</sup>,  
N. BOGOYCHEV<sup>1</sup>, C. BUCK<sup>1</sup>, D. CAROSELLI<sup>2</sup>, L. CARVALHO<sup>4</sup>,  
A. CATTELAN<sup>2</sup>, R. CATTONI<sup>3</sup>, M. CETTOLO<sup>3</sup>, M. FEDERICO<sup>3</sup>,  
B. HADDOW<sup>1</sup>, D. MADL<sup>1</sup>, L. MASTROSTEFANO<sup>2</sup>, P. MATHUR<sup>3</sup>,  
A. RUOPP<sup>4</sup>, A. SAMIOTOU<sup>4</sup>, V. SUDHARSHAN<sup>4</sup>,  
M. TROMBETTI<sup>2</sup>, J. van der MEER<sup>4</sup>

<sup>1</sup>University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom

<sup>2</sup>Translated srl, Via Nepal, 29, 00144 Rome, Italy

<sup>3</sup>Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo, Italy

<sup>4</sup>TAUS B.V., Oosteinde 9, 1483 AB De Rijp, Netherlands

ugermann@inf.ed.ac.uk

**Abstract.** *Modern MT* ([www.modernmt.eu](http://www.modernmt.eu)) is a three-year Horizon 2020 *innovation action* (2015–2017) to develop new open-source machine translation technology for use in translation production environments, both fully automatic and as a back-end in interactive post-editing scenarios. Led by Translated srl, the project consortium also includes the Fondazione Bruno Kessler (FBK), the University of Edinburgh, and TAUS B.V. *Modern MT* has received funding from the *European Union's Horizon 2020 research and innovation programme* under Grant Agreement No645487 (call ICT-17-2014).

## Project Description

*Modern MT* aims to improve the state of the art in open source machine translation software by developing cloud-ready software that offers

- A **simple installation** procedure for a ready-to-go, REST-based translation service.
- **Very fast set-up times** for systems built from scratch using existing parallel corpora (e.g., translation memories). The goal is to process incoming data at approximately the speed at which it is uploaded.
- **Immediate integration of new data** (e.g., from newly post-edited MT output). Rebuilding or retuning the system will not be necessary.
- **Instant domain adaptation** by considering translation context beyond the individual sentence, without the need for domain-specific custom engines.
- **High scalability** with respect to throughput, concurrent users, and the amount of data the system can handle.

A first version of the software is available at <https://github.com/ModernMT/MMT>.

*Modern MT* is also actively **collecting and curating parallel data** for internal use and public release from web crawls and contributions from translation stakeholders, to improve MT quality for everyone.