



Maschinelle Übersetzungs-Plug-ins in CAT-Tools

MDÜ 04/2017

## Sicher (nicht) nutzen?

Dangerous MT TeMptations – or Secure Collaboration?

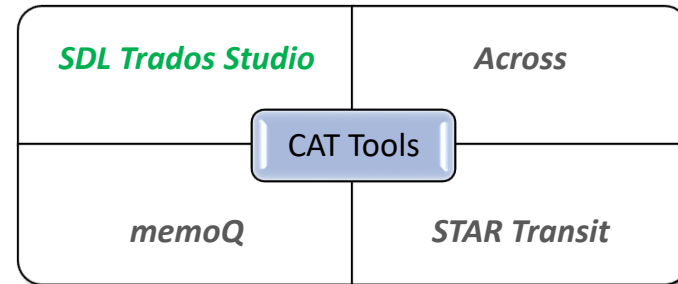
*ETUG Conference 2018*  
15.06.2018

Christine Bruckner

# MT Integration Modes with TM

In MDÜ article:  
focus on four CAT tools

Already existed in *Translator's Workbench*, see Heyn, Matthias (1996): *Integrating machine translation into translation memory systems*. <http://www.mt-archive.info/EAMT-1996-WS.pdf>, S. 111-123

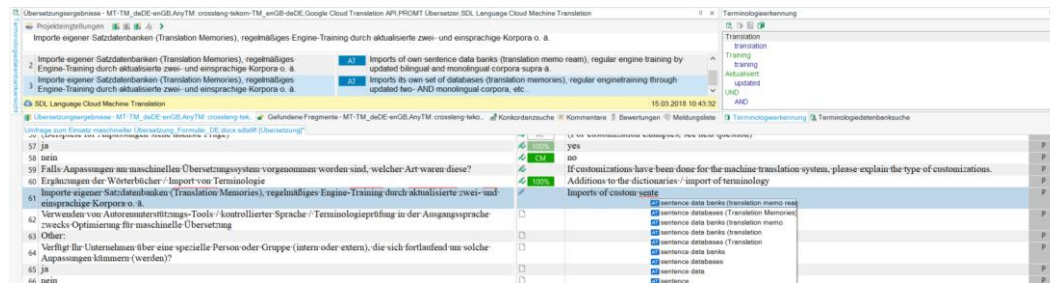


For overview of MT plugins available for *SDL Trados Studio*, see:

- [http://tradosstudiomanual.com/wp-content/uploads/2017/MT-plugin\\_overview.pdf](http://tradosstudiomanual.com/wp-content/uploads/2017/MT-plugin_overview.pdf) (Mats Linder)
- [SDL AppStore](#)

## Machine translation integration in CAT tools via:

- batch mode / via pre-translation
- interactive lookup during translation
  - Segment level
  - Sub-segment level (*AutoSuggest*, *MatchRepair*)



Adaptive MT options and interactive MT proposals (incl. AutoSuggest) in SDL Trados Studio 2017 SR-1 via SDL Language Cloud MT

# Information Security + Data Protection Aspects with Cloud MT Solutions

## Information Security:



## Data Protection / Privacy:



### Confidentiality:

- a) The “wrong” people can see information in transit.
- b) MT sites can use your data in ways you did not intend.

(Don de Palma, 2014: <http://www.tcworld.info/e-magazine/translation-and-localization/article/free-machine-translation-can-leak-data/>)

“data protection by design will become a legal obligation once the **GDPR** starts to apply in May 2018. Those who process **personal data** of individuals will have to take data protection into account “both at the time of the determination of the means for processing and at the time of the processing itself”, as Article 25 of the GDPR puts it. Data protection principles and safeguards must guide the process from the moment **technology and organisational practices** are designed and planned.”

[https://edps.europa.eu/press-publications/press-news/blog/crucial-moment-communications-privacy\\_en](https://edps.europa.eu/press-publications/press-news/blog/crucial-moment-communications-privacy_en)

➤ **Data encryption** 

➤ **Avoid use / storing of data input by third parties** 

➤ **Server location** 

➤ **Data Processing Agreements** 

➤ **Anonymization / Pseudonymization** 

# Examples from the Terms of Service “Jungle” of Cloud MT Providers

June 2018

## MyMemory

*„We collect any segment submitted and store it on a long term basis, whether it’s public or private. [...] The contributions to the archive, whether they are "Public Data" or "Private Data", are collected, processed and used by Translated to create statistics, set up new services and improve existing ones.”*

[...]

*Translated can entitle its partners of a usage license over "Public Data" or "Private Data" in order to improve the quality of our and/or 3rd parties' services (eg. machine translation suggestions, glossaries, language models, spell checkers...).*

*(<https://mymemory.translated.net/doc/en/tos.php>)*

## DeepL Pro

*“DeepL Pro never stores the texts you are translating, and the connection to our servers is always encrypted.*

*This means that your texts are not used for any purposes other than your translation, nor can they be accessed by third parties.*

*As a company based in Germany, all our operations comply with European Union Data Protection laws.”*

BUT:

- *Servers are located in Iceland*
- *“7.12 Customer is obligated to observe all legal requirements for the collection, processing and use of data which is transmitted to and processed by DeepL for the Customer in connection with the provision of its services under this Agreement. In particular, Customer shall immediately inform DeepL if Customer intends to transmit personal data to DeepL using the API. Customer guarantees not to collect, process or use any personal data in connection with the API without the express consent of the data subject or sufficient other legal authorisation. DeepL will reasonably co-operate with Customer in order to assist Customer in implementing such required legal authorisations.”*
- *(<https://www.deepl.com/pro-license.html>, June 2018)*

# Examples from the Terms of Service “Jungle” of Cloud MT Providers

June 2018

## SDL Language Cloud MT

- [SDL Language Cloud FAQs](#): “With SDL Language Cloud Machine Translation you can rest assured that your content is safe. SDL guarantees that your data is not saved or used outside of the scope or timeframe that is necessary to provide you with the service”.

- [SDL AdaptiveMT Server Hosting](#): „SDL AdaptiveMT engine servers are located in the United States and hosted on AMAZON AWS servers”

### [Terms and Conditions for Language Cloud Translation Services](#):

- “The data is **encrypted** when in transit between your device and the SDL system and our product is hosted in a third party data centre with industry leading security.”
- “If you submit personal data SDL will act only as a data processor as defined in Directive 95/46/EC and you will remain responsible for the personal data as data controller. “

## • Amazon Translate • Google MT • Microsoft MT

“It” looks like [Amazon has not yet caught up with the latest developments in data privacy](#). If you look under “Privacy” [right here](#), you can see that data submitted through Amazon Translate is used for purposes other than your own, which is so “last month” (when [Microsoft finally joined Google and DeepL in no longer doing this as long as the data comes through an API connection](#)).“ – Jost Zetsche, *The 286th Tool Box Journal*, April 2018

- [Amazon Translate servers in Europe \(Ireland\) available](#)
- [Google MT server location probably outside EU](#)
- [Microsoft MT server location outside EU](#)

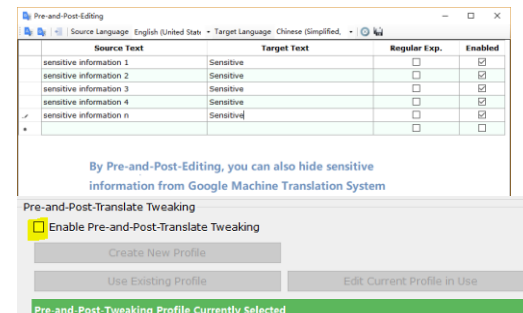
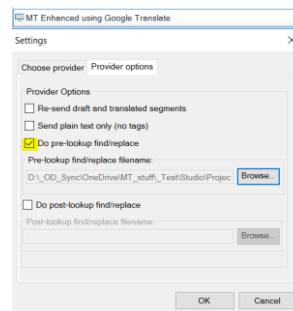
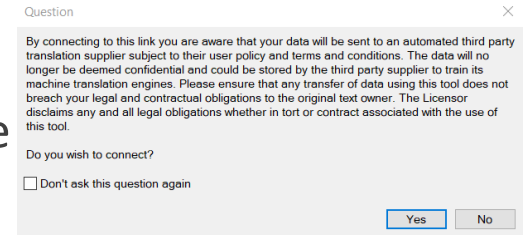
# Security Features in MT Plugins for SDL Trados Studio

## ➤ Warn before Use:

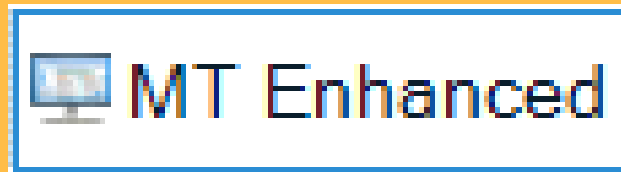
- (too few) warnings and links to Terms of Service of cloud MT providers in Studio environment

## ➤ Masking (= anonymization / pseudonymization) options:

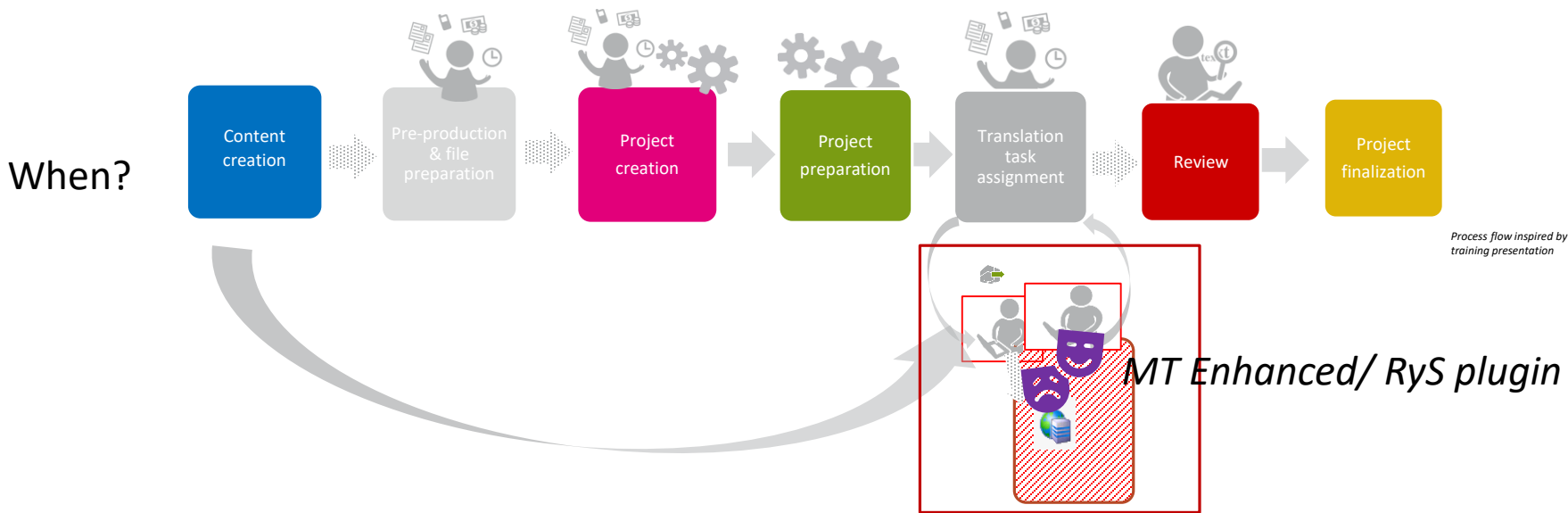
- in *MT Enhanced* plug-in (for Google MT and Microsoft MT)
- in *RyS Google Cloud Translate (NMT) (paid)*
- but not with SDL's core out-of-the-box MT engines (*SDL Language Cloud MT, Google Cloud Translation API, DeepL Pro* plugin etc.



# Demo: Masking via *MT Enhanced Plugin*



# Data Masking for Cloud MT in the Translation Process





# SDL *projectAnonymizer*



projectAnonymizer

Von SDL Community Developers

★★★★★ 127 

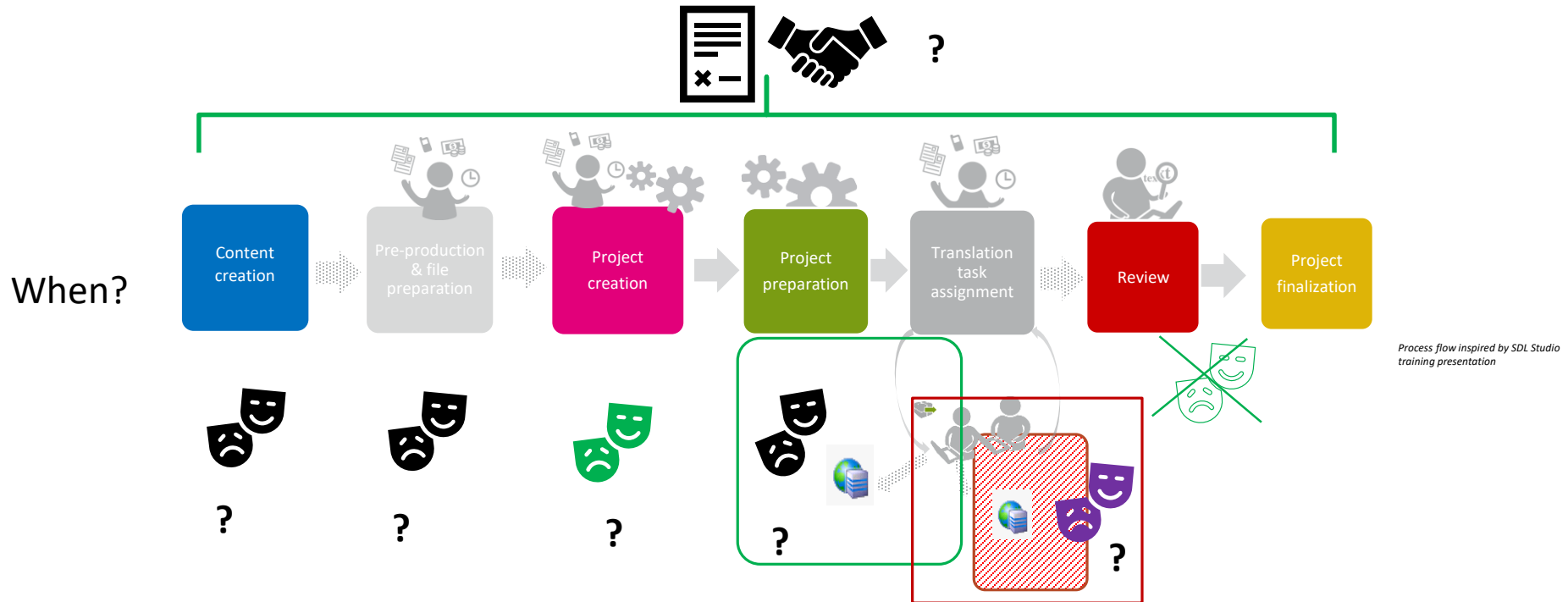
(0 Bewertungen)

**\*\*Updated to Version 2.1\*\***

The *projectAnonymizer* application adds the ability to protect data during a translation project by converting it to tags. These tags can be encrypted if it's important the data is securely controlled and they can be decrypted after the project is complete. The process is managed through a new batch task that is added to your menu in Studio.

<http://multifarious.filkin.com/2018/05/27/data-protection/>

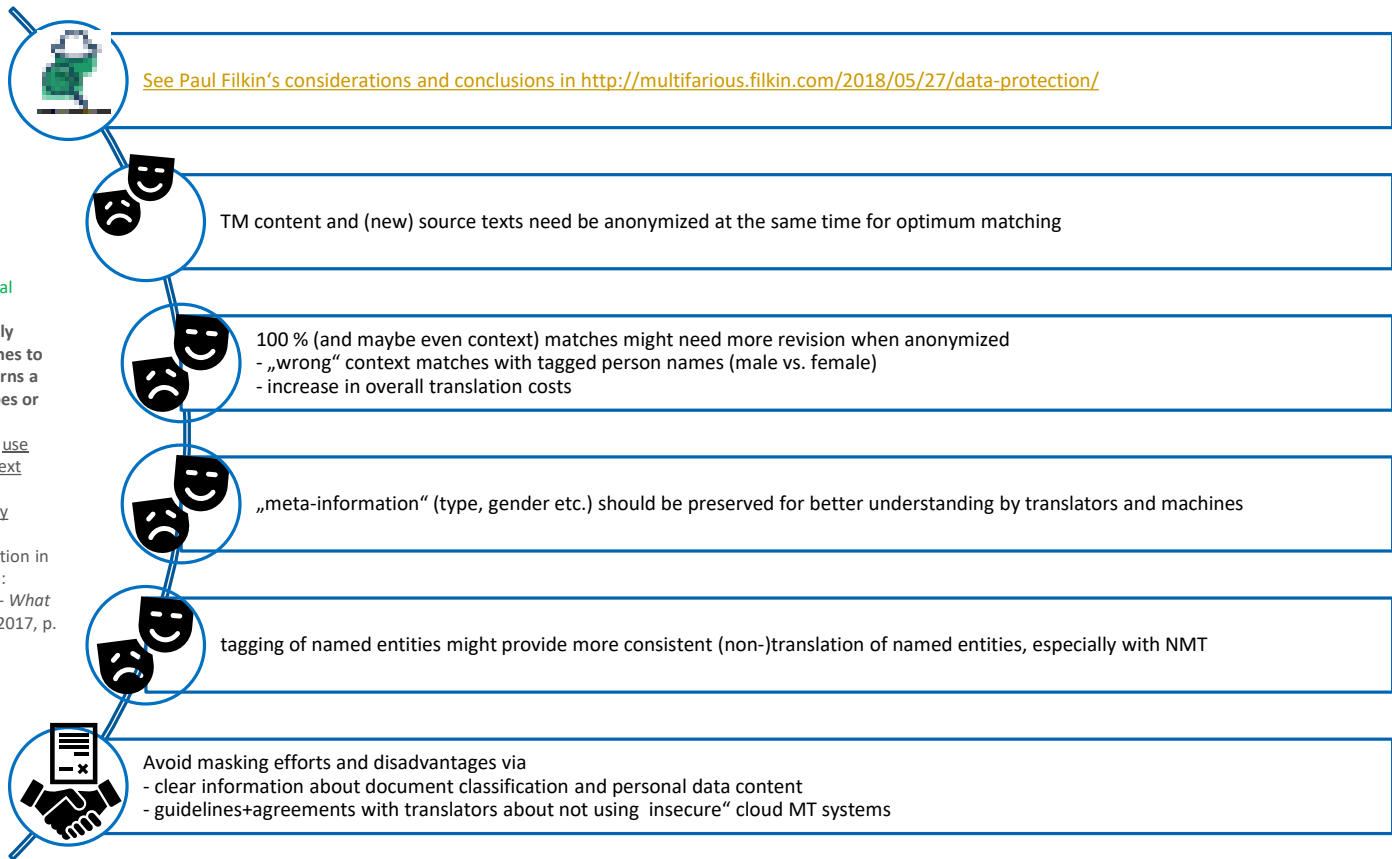
# Data Masking in the Translation Process



## How?

- Manually in source texts
- semi-automatically via search+replace strings+regex file ([projectAnonymizer](#), *MT Enhanced/ RyS plugin, etc.*)
- automatically via *Named Entity Recognition (NER)* tools
- “For many real-life use cases it isn’t enough to just substitute names with pseudonyms, or mask some of the values. With a little additional background knowledge it is often possible to identify the individuals you thought had been anonymized.” (<https://blogs.sap.com/2018/01/26/going-beyond-masking-how-to-anonymize-large-data-sets/>)

# Some Conclusions on Data Masking



“Offline MT [...] does not pose any special problems with regards to personal data processing. [...] The user should generally avoid online MT services where he wishes to have information translated that concerns a third party (or is not sure whether it does or not).

[...] This means users may be advised to use online MT services only for translating text from their own language into another language, and not vice versa (where they cannot be sure of the content)“.

(Kamocki, Dr. Tauch (2017): Data protection in Machine Translation under the GDPR, in: Porsiel, Jörg (ed.): *Machine Translation - What Language Professionals Need to Know*, 2017, p. 71 ff.)

?

# Discussion

---

- ❖ How does your organization deal with access to *cloud* MT solutions (via web interface or via plugin to CAT / TMS)?
- ❖ How does your organization deal with personal data in *documents* (GDPR compliance), especially when translation is outsourced?
- ❖ Has your organization established any measures / guidelines regarding personal *meta-data* in translation memories fields / termbase entries etc.?

Thank you!

[Christine.Bruckner@CATTMaTTers.de](mailto:Christine.Bruckner@CATTMaTTers.de)

---