



---

# Introduction to the Connecting Europe Facility **Automated Translation building block**

---

## **DG TRANSLATION**

Directorate-General for Translation

## **DG CONNECT**

Directorate-General for  
Communications Networks, Content  
and Technology

## **DIGIT**

Directorate General for Informatics

---

---

March 2016



---

## DISCLAIMER

This document is for informational purposes only and the Commission cannot be held responsible for any use which may be made of the information contained therein. References to legal acts or documentation of the European Union (EU) cannot be perceived as amending legislation in force or other EU documentation.

The document contains a brief overview of technical nature and is not supplementing or amending terms and conditions of any procurement procedure; therefore, no compensation claim can be based on the contents of the present document.

---

## Introduction to the Connecting Europe Facility Automated Translation building block

**Created by:** CEF Automated Translation

## DOCUMENT HISTORY

Version	Date	Modified by	Short Description of Changes
0.06	30/03/2016	CEF	Published



---

# Table of contents

		Page number
<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Context</b>	<b>9</b>
<b>3</b>	<b>Motivation</b>	<b>12</b>
<b>4</b>	<b>Technical Specifications</b>	<b>18</b>
<b>5</b>	<b>Implementations</b>	<b>20</b>
<b>6</b>	<b>Governance</b>	<b>22</b>
<b>7</b>	<b>Success stories</b>	<b>25</b>
<b>8</b>	<b>Definitions</b>	<b>27</b>



---

# Audience

This document describes the Automated Translation (AT) building block which is one of the Connecting Europe Facility (CEF) Digital programme's essential digital services. These essential digital services, called building block Digital Service Infrastructures (DSIs) will play a vital role in the flow of data across borders and sectors.

The document will help Service Providers who are interested to better understand the CEF Automated Translation Platform. In particular it provides information on the background, terminology, concepts, roles, components, connectivity, features and architecture of the CEF Automated Translation platform.

This document is intended for the following audiences:



## **SERVICE PROVIDERS**

European and national public administrations interested in integrating Automated Translation in their digital services to guarantee their services are supported in all official languages of the EU.



## **TRANSLATORS**

Translators in EU institutions and Public Administrations interested in finding out how they can use the CEF Automated Translation platform as a tool for their day-to-day work of producing high quality translations.

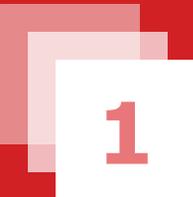


## **OTHER STAFF**

Non translating staff in EU institutions and Public Administrations interested in finding out how CEF Automated Translation platform can facilitate their day-to-day work and cross-border information exchange.

---

Whilst every effort has been made to ensure that the information contained in the document is correct, any comments on it should be submitted to the European Commission:  
[CEF-BUILDING-BLOCKS@ec.europa.eu](mailto:CEF-BUILDING-BLOCKS@ec.europa.eu)



**1**

# Introduction

What is CEF Automated Translation?

---

# Introduction to CEF Automated Translation

The CEF Automated Translation (CEF AT) building block helps European and national public administrations exchange information across language barriers in the EU. The main purpose of CEF AT is to make all Digital Service Infrastructures (DSIs) multilingual. For example, users of the Open Dispute Resolution Portal are able to access information and communicate with their counterparts using their own language. While CEF AT is mainly intended to be integrated into other digital services, it also offers useful stand-alone services for translating documents or snippets of text.

The technical implementation of a user interface for CEF AT guarantees confidentiality and security of all translated data. Unlike general-purpose web translators, CEF AT will be adapted to specific terminology and text types that are typical for the usage context (e.g. tender documents, legal texts, medical terminology). It enables multilingual operation of digital services and can be used to reduce the time and cost of translating documents.

Automated Translation is a building block of the Connecting Europe Facility (CEF). These building blocks are reusable specifications, software and services that will form part of a wide variety of IT systems in different policy domains of the EU.

The promotion of the use of common building blocks is a way to lower barriers for technical integration and provide tried and tested solution components that will speed up the delivery of Digital Public Services, that work across borders, in a cost efficient manner.

The CEF Automated Translation building block builds further on the existing Commission Machine Translation service (MT@EC) developed by the Directorate-General for Translation ([DGT](#) ) , who is responsible for the technical management of the Automated Translation building block DSI and more specifically of its machine translation service (eTranslation).

Implementation of the EU policy directly related to Automated Translation and its overall administration is the responsibility of the Directorate-General for Communications Networks, Content and Technology ([DG CNECT](#) ) of the European Commission.

The Innovation and Networks Executive Agency ([INEA](#) ) is responsible for the implementation of the CEF Telecom programme grants in cooperation with the Commission.



*More information about the CEF Telecom policy background, its Work Programmes and related information is available on the [Digital Agenda website](#).*



*More information about Automated Translation is available on the [CEF Digital Single Web Portal](#).*

# Introduction to CEF Automated Translation

The harmonious co-existence of many cultures and languages in Europe is one of the cornerstones of the European project, and is embedded in Article 3 of the [Treaty on the Functioning of the European Union](#) .

The treaty aims to protect the cultural, religious and linguistic diversity in Europe. It states that:

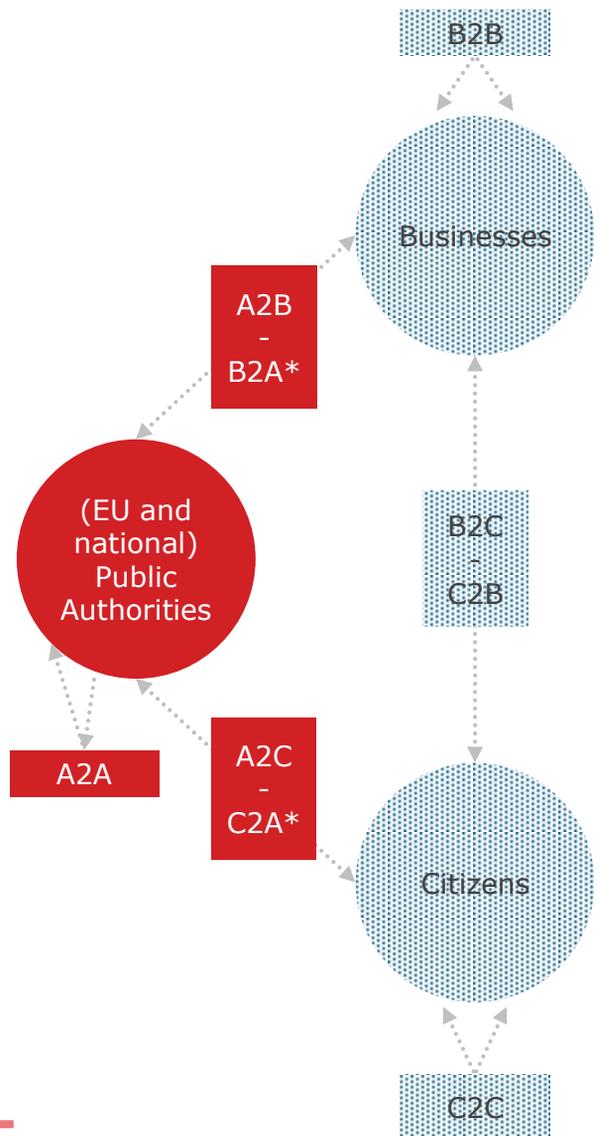
**Citizens have a right to address the official EU bodies in any of the EU's official languages and to receive a reply in that language.**

Automated Translation is a key enabler for cross-border electronic services and transactions and a central building block of the [Digital Single Market](#) .

The goal of the CEF Automated Translation building block is to set-up a platform enabling multilingualism of public digital services and public administration, facilitating the interactions with and between actors in the EU Member States.

More precisely, the scope of CEF AT is three-fold:

- Integration of AT in digital services so that public administration, citizens and businesses in the EU can use these services in their own language.
- AT services to public administrations to support the work of translators
- AT services to public administrations to facilitate cross-border information exchange.



 Primary scope

 Out of scope

\* Through web-portals

A = Public Administration

B = Business

C = Citizen



---

# Benefits of CEF Automated Translation

Below is an overview of the expected **benefits** to the individual Projects and Policy Domains:

## BENEFITS

### Increase speed



Increasing processing speed for inbound information so the receiving administration quickly understands the information without having to wait for a translation and "routes" it to the right person or department.

### Reduce Cost



Reducing cost since human translators in the receiving administration will only receive requests to translate the specific pages of an incoming document deemed important.

### Improve Information Exchange



Facilitating easier information exchange within and between interest groups, between judicial collaborators etc., based on their expertise and not on the knowledge of the working language(s) of the group.



2

## Context

What is the background of CEF Automated Translation?

---

# Context

## MT@EC

CEF Automated Translation builds on the existing machine translation system of the European Commission, MT@EC, developed by DGT under the Interoperability Solutions for European Public Administrations (ISA) programme.

MT@EC is operational since June 2013 and is currently used by EU institutions and bodies as well as Public Administrations in the EU Member States and EEA countries. It is also available in European online services funded or supported by the Commission, such as IMI and SOLVIT.

MT@EC is currently based on MOSES open-source technology that is a Statistical Machine Translation (SMT) system developed with co-funding from EU research and innovation programmes. MT@EC translation engines are trained by DGT using the vast Euramis translation memories, comprising over 900 million sentences in 24 languages produced by the translators of EU institutions over the past two decades. MT@EC is, as such, particularly suited for the needs of EU documents and policy.

The MT@EC service is available via a web user interface accessible by individual users authenticated through ECAS. Other IT applications (like TED or IMI) also submit translation requests to MT@EC making use of web services (i.e. mechanisms allowing machine-to-machine interactions).

The MT@EC service provides the ability to translate formatted documents and plain text between any pair of EU official languages while preserving the structure and format of those documents. Several documents may be translated into multiple target languages in a single submission.



More information about MT@ EC is available on the [DGT web site](#).



More information about SMT open-source technology on which MT@EC is based can be found on the [MOSES website](#)



## CEF Automated Translation

CEF Automated Translation will carry on the MT@EC functionalities and will scale-up the existing infrastructure supporting MT@EC service by:

- Promoting and supporting effective pooling of even larger set of language resources from various sources to a) continuously improve the quality of the automated translation service and b) be able to build Machine Translation engines which are customised for specific domains.
- Reviewing and adapting the functional architecture of MT@EC to improve the quality and performance of the service (in so far the underlying engines) and to be modular enough in order to serve the needs of current and future users (CEF DSIs as a first priority). The emphasis will be put on developing further the use of web services (i.e. mechanisms allowing other IT systems to automatically call a service provided by the CEF AT platform). Ultimately, it is foreseen that CEF AT would also be used by other public pan-European services and Public Administrations across the EU, Iceland and Norway.
- Developing an Engines factory to automate the building of machine translation which will comprise also a domain adaptation facility.
- Acquiring additional infrastructure resources (possibly taking advantage of new emerging cloud offers) where parts of the service could be moved to improve its performance and its efficiency.

The aim is to build a platform that is much more flexible and scalable to accommodate increasing load and performance requirements as well as to offer custom solutions to different groups of users (i.e. other DSIs and online services) at a much larger scale than is currently possible through MT@EC. The present document focuses on the development of the "core service platform" in the sense of the service to which other DSIs, public online services and users can connect to request automated translation to serve their needs.



3

## Motivation

Why should you reuse Automated Translation?

---

# Motivation – Goals

The EU now has more than 500 million citizens, 28 Member States, 3 alphabets and 24 official languages.

The harmonious co-existence of many languages in Europe is a powerful symbol of the EU's aspiration to be united in diversity, one of the cornerstones of the European project. A successful multilingualism policy can strengthen the life chances of citizens: it may increase their employability, facilitate access to cross-border services and rights' preservation, and contribute to solidarity through enhanced intercultural dialogue and social cohesion.

With respect to the linguistic diversity in the EU, Automated Translation is a key enabler to achieve the Digital Single Market. A success factor for implementing Automated Translation is to clearly understand the motivation for its use, its scope and (business) needs.

The main goals addressed by Automated Translation are:

## **Machine Translation**

*Provide an automated translation system that is aware of the correct terminology and style for different contexts.*

## **Cross-border Information Exchange**

*Allow the exchange of information and access to online content and services across borders in any of the supported languages.*

## **Increased Accessibility**

*Allow easy access to the machine translation service for both people and machines.*

## **Continuity and Quality of Service**

*Guarantee continuous service of high quality, with due consideration for the confidentiality and other legal aspects related to trust when exchanging information.*

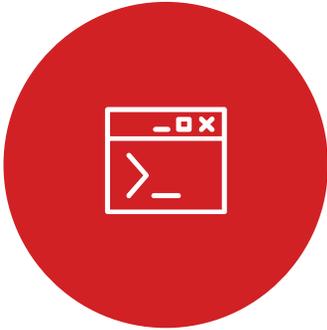


---

# Motivation – Use Cases

The CEF Automated Translation solution is offered mainly as a service, in two modes:

## For System Suppliers



Providing automated translation capabilities for digital services by integrating CEF AT through an Application Programming Interface (API).

1. You are a system supplier of a DSI or any other public online service.
2. You would like to give the possibility to users of your online service to use it in their own language.



3. Integrate the Automated Translation Services within the system/service you supply by connecting through the Application Programming Interface (API) provided by web services to automatically request machine translations to MT@EC.



4. Users of your online service will be able to use your online service in their own language.

## For Public Administrations



Offering a web user interface for direct use by human users.

1. You are a person working in a EU institution or agency or public administration in an EU country, Iceland or Norway
2. You need a quick, raw translation of a text in a language you can't read



3. Directly access the MT@EC service and select the type of translation you want:

- Translate documents — upload one or more documents (this can even be done simultaneously)
- Translate text — type or copy and paste your snippet of text



4. The result of your translation is stored in your personal workspace or sent to you by email.

---

# Motivation – Services

The CEF Automated Translation solution is offered as a service.

The service provides the ability to translate formatted documents and plain text between any pair of EU official languages, as well as Norwegian (Bokmål), while preserving to the greatest extent possible the structure and format of those documents.

## FEATURES

- User web interface is available in 24 languages
- Can translate multiple documents to multiple languages in one go
- Accepts several input formats: doc(x), xls(x), ppt(x), odt, rtf, txt, html, tmx, xliff, Pdf
- Source document format/formatting maintained [not for pdf]
- Specific output formats for translation: tmx and xliff
- Automatic language detection is supported for the translation of text snippets for texts at least 30 characters long
- The interface will support the selection of specific domains (subject areas) of translation
- Monitor progress of translation requests
- Translation can also be returned by email or stored on the interface.
- Indication of quality for language pairs
- Feedback mechanism

## ACCESS

### System integration

System suppliers interested in integrating the Automated Translation services within their online services can contact: [CEF-BUILDING-BLOCKS@ec.europa.eu](mailto:CEF-BUILDING-BLOCKS@ec.europa.eu).

### Direct access

The Automated Translation service is available for:

1. Staff working for EU institutions or agencies using their ECAS account credentials.
2. Staff working for a public administration in an EU country, Norway and Iceland should follow these steps:
  - a) Sign up for your personal ECAS account and password (using only your professional email address).
  - b) Email [DGT-MT@ec.europa.eu](mailto:DGT-MT@ec.europa.eu) to tell us you have an ECAS account. Say what your job involves and which public administrative body you work for. Don't forget to include your full signature.
  - c) We'll create your MT@EC account (and notify you).



---

# Motivation – Services

## SUPPORT

If you come from an EU institution and you cannot access the service, please contact your local IT helpdesk.

For problems related to the output of machine translation (access to the file, formatting, visualisation, etc.) please check the troubleshooting guide included on the Help pages of MT@EC, , *available in all 24 EU languages*. If you cannot find an answer there, contact [DGT-MT@ec.europa.eu](mailto:DGT-MT@ec.europa.eu).

For other problems related to the operation of the service contact [DGT-MT@ec.europa.eu](mailto:DGT-MT@ec.europa.eu).





4

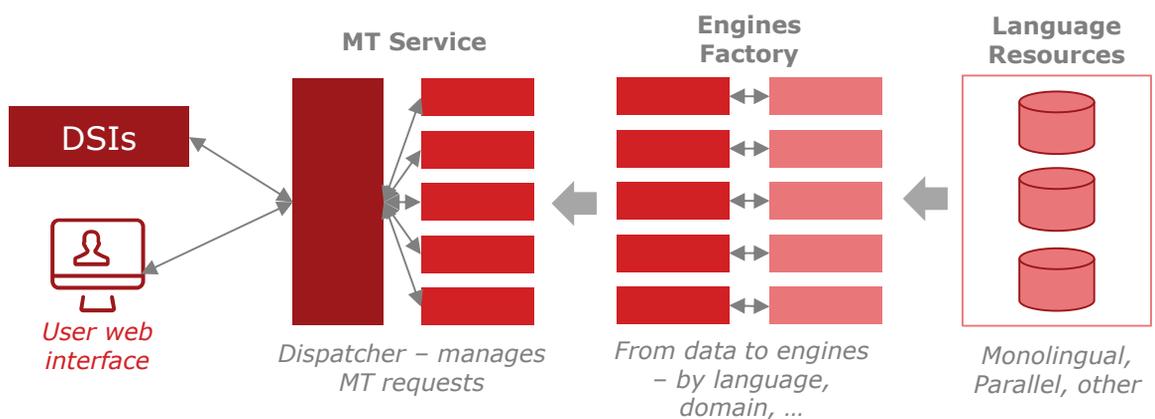
## Technical specifications

What are the technical foundations of CEF Automated Translation?

# Technical specifications

The MT@EC service uses statistical machine translation technology based on [Moses](#) .

Moses is an open-source project. It is an implementation of the statistical (or data-driven) approach to machine translation (MT). This is the dominant approach in the field at the moment, and is employed by most online machine translation systems. In statistical machine translation (SMT), translation systems are trained on large quantities of parallel data (from which the systems learn how to translate small segments), as well as even larger quantities of monolingual data (from which the systems learn what the target language should look like). Parallel data is a collection of sentences in two different languages, which is sentence-aligned, in that each sentence in one language is matched with its corresponding translated sentence in the other language. It is also known as a bitext.



## SECURITY

To ensure security, MT@EC is hosted in the European Commission Data Centre. The MT@EC service is accessible either over the sTESTA network, or over a secure https Internet connection. In both cases documents are encrypted (by https), while in the case of sTESTA they are also transported from the user to the translation system and back over this non-public network, which is separate from the internet.



5

## Implementations

What are the existing CEF Automated Translation implementations?

---

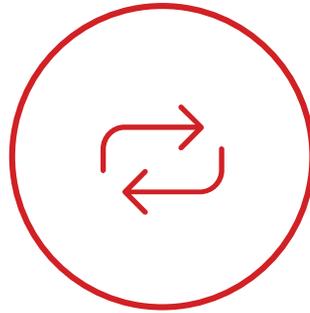
# Your options

While considering a CEF Automated Translation project, two alternative implementation scenarios can be considered:



### **human-to-machine use**

Individuals directly using the Automated Translation services through the user web interface



### **machine-to-machine use**

Integration of Automated Translation services in your existing online services



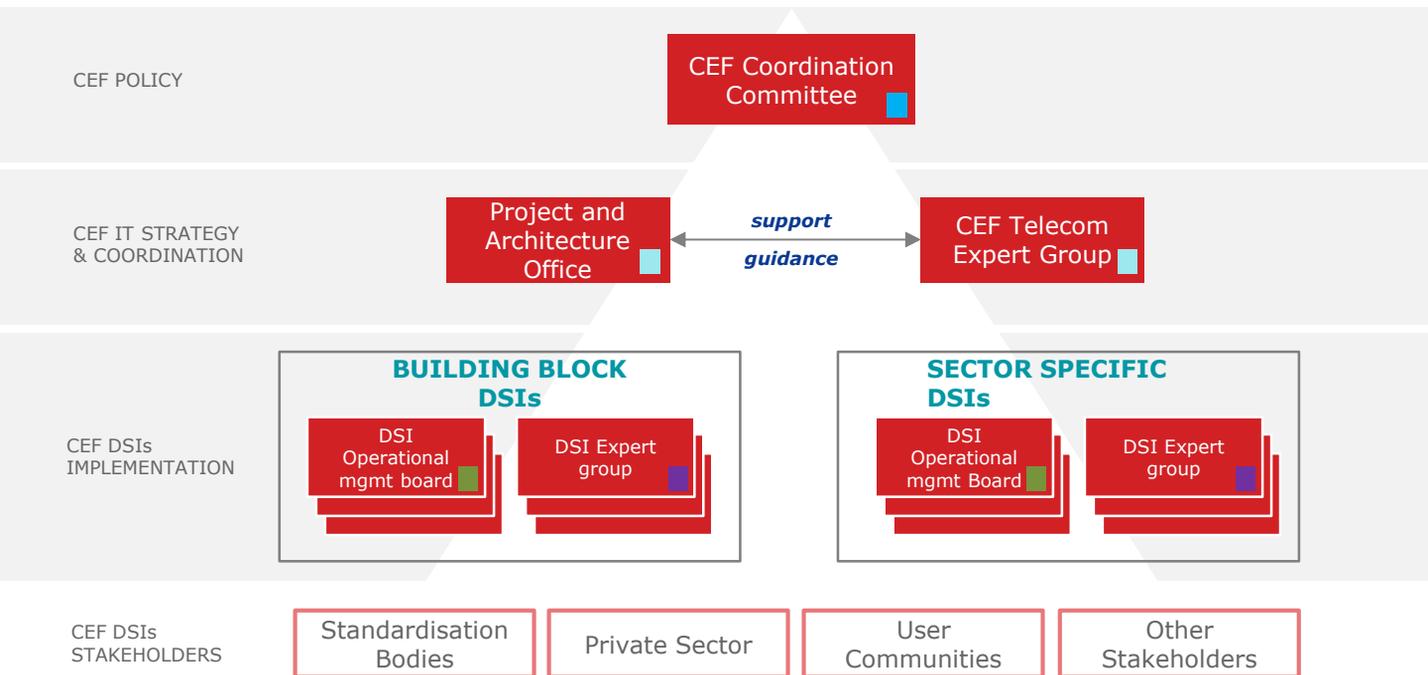
# Governance

What is the CEF IT governance model?

# Governance

CEF has a [governance model](#)  place to ensure that the reusable building blocks correspond to user needs and expectations and that stakeholders can influence their evolution over time.

As shown below, this model, which is also used for the Automated Translation platform, is multi-layered, designed according to guiding principles stemming from the CEF regulations.



## SCOPE OF DECISIONS

-  Policy implementation decisions
-  Strategic decisions
-  Tactical decisions
-  Operational & technical implementation decisions

*CEF's governance model*



---

# Governance

The Automated Translation Operational Management Board (OMB) meets frequently (ideally on a monthly basis) to discuss all types of operational matters (change requests, release planning, etc.). This governance body brings together:

- The Automated Translation DSI (Policy) Owner, accountable for its policy and its translation into the functionality of the building block: this role is currently played by DG CNECT.
- The DSI Solution Provider, accountable for the delivery side of the Automated Translation building block and related core services: this role is currently played by DGT.
- The Directorate Generals of the Commission that have decided to reuse, or are interested in reusing Automated Translation for core services under their responsibility (especially CEF DSIs).

The OMB works in close collaboration with, and is advised by the Language Resource Board. This is the DSI Expert Group comprising two representatives by country, the National Anchor Points, one for language technologies and one for public services. Consultation processes ensure that standardisation bodies, the private sector, user communities and other stakeholders are involved as required.

In case of issues, these can be escalated to the 'CEF Telecom Expert Group' (also composed of Member State Representatives). The CEF Telecom Committee is mostly involved in the formulation of the CEF Telecom Work Programmes and less in their implementation.



## Success stories

---

## Success Stories

MT@EC has been operational since June 2013 and is currently being used by EU institutions and bodies.

It is also available in European online services such as:

---

Service	Description/URL/use
IMI	Internal Market Information System – an intranet for more than 7000 public administrations in the EU Member States <a href="http://ec.europa.eu/internal_market/imi-net/index_en.html">http://ec.europa.eu/internal_market/imi-net/index_en.html</a> ➔ free-text message exchange
SOLVIT	An on-line problem solving network concerning misapplication of Internal Market law by public authorities. <a href="http://ec.europa.eu/solvit/">http://ec.europa.eu/solvit/</a> ➔ free-text message exchange
nLex	A common gateway to National Law <a href="http://eur-lex.europa.eu/n-lex/">http://eur-lex.europa.eu/n-lex/</a> ➔ translate results of queries
TED	Tenders Electronic Daily is the online version of the 'Supplement to the Official Journal' of the EU (European public procurement). <a href="http://ted.europa.eu/">http://ted.europa.eu/</a> ➔ translate contract notices

---





8

# Definitions

---

# Definitions

A **Building Block** offers basic capabilities and services that can be combined with other building blocks or sector-specific applications to deliver architectures, solutions and sector-specific services.

**Source:** TOGAF® 9.1 and Regulation (EU) No 283/2014

A **Digital Service Infrastructure (DSI)** enables networked services to be delivered electronically, typically over the internet, providing cross-border interoperable services for citizens, businesses and/or public authorities.

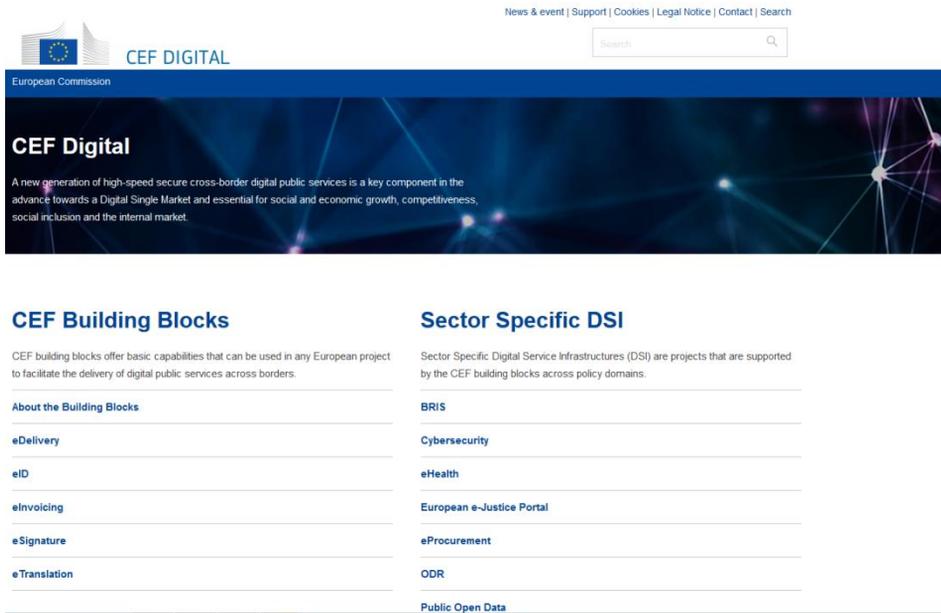
**Source:** Regulation (EU) No 283/2014

The CEF **Automated Translation (AT) building block** helps European and national public administrations exchange information across language barriers in the EU. The main purpose of AT is to make all Digital Service Infrastructures (DSIs) multilingual.

**Source:** Join-up

**Statistical machine translation (SMT)** is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

Visit the catalogue of building blocks on the CED Digital Single Web Portal   
<https://ec.europa.eu/cefdigital>



The screenshot shows the top of the CEF Digital website. At the top right, there are links for "News & event", "Support", "Cookies", "Legal Notice", "Contact", and "Search". Below these is a search bar with the text "Search" and a magnifying glass icon. The main header features the European Commission logo and the text "CEF DIGITAL". Below the header is a large blue banner with the text "CEF Digital" and a sub-headline: "A new generation of high-speed secure cross-border digital public services is a key component in the advance towards a Digital Single Market and essential for social and economic growth, competitiveness, social inclusion and the internal market." Below the banner are two columns of content. The left column is titled "CEF Building Blocks" and lists: "About the Building Blocks", "eDelivery", "eID", "eInvoicing", "eSignature", and "eTranslation". The right column is titled "Sector Specific DSI" and lists: "BRIS", "Cybersecurity", "eHealth", "European e-Justice Portal", "eProcurement", "ODR", and "Public Open Data".

---

**DGT**  
Directorate-General for Translations

**DG CONNECT**  
Directorate-General for  
Communications Networks, Content  
and Technology

---

Contact us

 [CEF-BUILDING-BLOCKS@ec.europa.eu](mailto:CEF-BUILDING-BLOCKS@ec.europa.eu)