

# A Functional Description of the NEC TM System

Version 1.00

---

<b>Revision Control</b>	<b>4</b>
<b>Introduction</b>	<b>4</b>
Contributors	5
<b>NEC TM Objectives</b>	<b>5</b>
<b>NEC TM Actors</b>	<b>7</b>
<b>Logical View</b>	<b>9</b>
External Calls Layer	10
API Layer	11
Service Layer	12
Data Layer	13
MT Suppliers	14
Users	15
Anonymization Toolkit	16
<b>Business View / Usage Case</b>	<b>17</b>
<b>Service View</b>	<b>20</b>
Fuzzy Match Services	20
Search translation memory segments	20
Batch search translation memory segments	21
User Management Services	22
Add/update user details	22
Delete user	23
Get user details	24
Add/update user scope	24
Delete user scope	25
Authorization	26
Domain Management Services	27
Add/update domain	27
Delete domain	27
Get domain details	28
Import Services	29
Import translation memory segments from TMX file	29
Add new translation memory unit	30
Export Services	30
Export translation memory segments to TMX file	30
Data Services	31
Language Resource Management	31

---

User Registration & Management	32
External MT Supplier Management	33

---

## Revision Control

Author	Description
Alex Helle	Initial Version of Specification (Sept 10th 2018)
Amando Estela	First Revision (Sept 17th 2018)
Carmen Herranz-Carr and Carolina Herranz-Carr	Second Revision (Sept 27th 2018)
Alex Helle	Third Revision (Sept 28th 2018)

---

## Introduction

This document is a high level description of the functional specifications of the NEC TM Platform. This document is created by Pangeanic and will be presented to other consortium members as a discussion document in the kick-off conference planned for 20-22 September.

This document aims to improve the consortium's understanding of the core functionalities of the NEC TM System and how they can build connectors that interface with these services.

## Contributors

The following consortium members were involved in the development of this document.

Member	Organisation
Alex Helle	Pangeanic
Amando Estela	Pangeanic
Carmen Herranz-Carr	Pangeanic
Carolina Herranz-Carr	Pangeanic

## NEC TM Objectives

NEC TM Data aims to promote better data-sharing practices in translation contracts between public administrations and translation service providers working on public sector texts, and so facilitating a single digital market. The central objective of the NEC consortium is to organise unexploited national bilingual assets that can be used as open data and general data for machine learning, in order to lower translation costs at a national level and across member states. Therefore the NEC TM Data platform will provide a solid framework for public administrations to adopt as policy for general data sharing from public translation contracts and real adoption scenarios at a national level. NEC TM Data promotes better data-sharing practices using open standards in translation contracts between public administrations and translation service providers.

The consortium will deliver a pan-European data-sharing awareness program, engaging national public administrations by providing a solid legal framework for data sharing. Specifically, its points of actions are:

- To provide a solid framework for public administrations to adopt as policy for general data sharing from public translation contracts and adoption solutions at a national level. The NEC TM Data platform's objective is to organise unexploited national bilingual assets that can be used as open data and general data for machine learning, in order to lower translation costs at a national level and across member states. It will gather translation memories from previous national contract awards from Member States and help them to centralise these language assets with the fast-performing NEC TM database, following industry best practices.
- To inform public administrations at the national level about translation technologies available for language resources, as well as to lobby for TM gathering to be enshrined in national translation contracts. Given that the majority of translated data generated in public contracts is currently not returned to public administrations and therefore its potential is not being maximised, NEC TM Data will ensure that public administrations make better use of the translated data generated in public contracts.
- To provide the centralised infrastructure for efficient data sharing, TM matching, TM retrieval, and domain categorisation of resources generated in 10 Member States (including the participating 3). This will enable the development of NEC TM, which will be an open source software developed from Pangeanic's translation memory database [ActivaTM](#).

---

While the NEC TM system will be accessible to any MT Supplier and to any EU Consumer of machine translation services, the initial release will provide services for the following regional and national public administration.

Public Administration	Description
SEAD	Provision of multilingual Publish Administration content (Source Language Spanish)

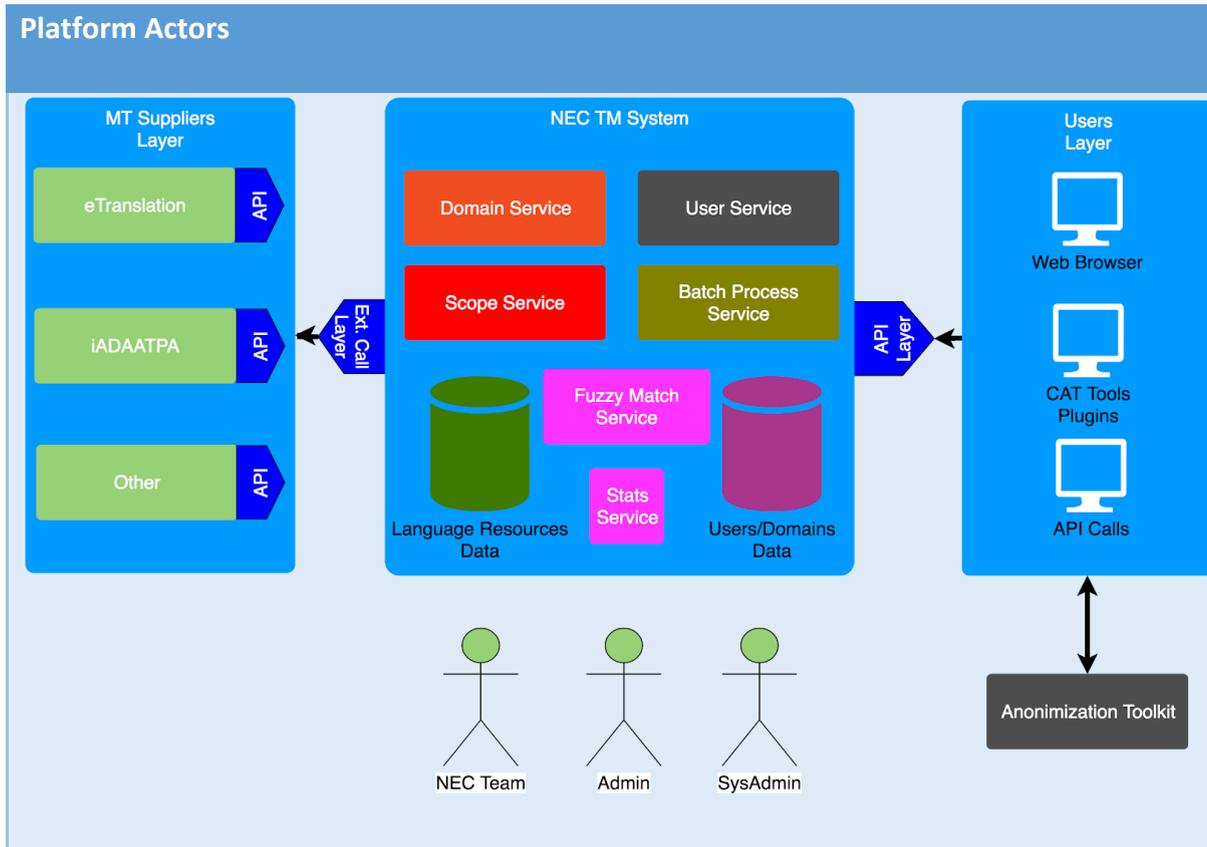
---

**NOTE:** The NEC TM platform will be available for cloud and on-site deployment and can be operated at a national or regional level with deployments operating independently of each other if required.

---

# NEC TM Actors

The following section introduces and defines the roles of actors that will engage with Us and leverage the facilities provided by the NEC TM platform.

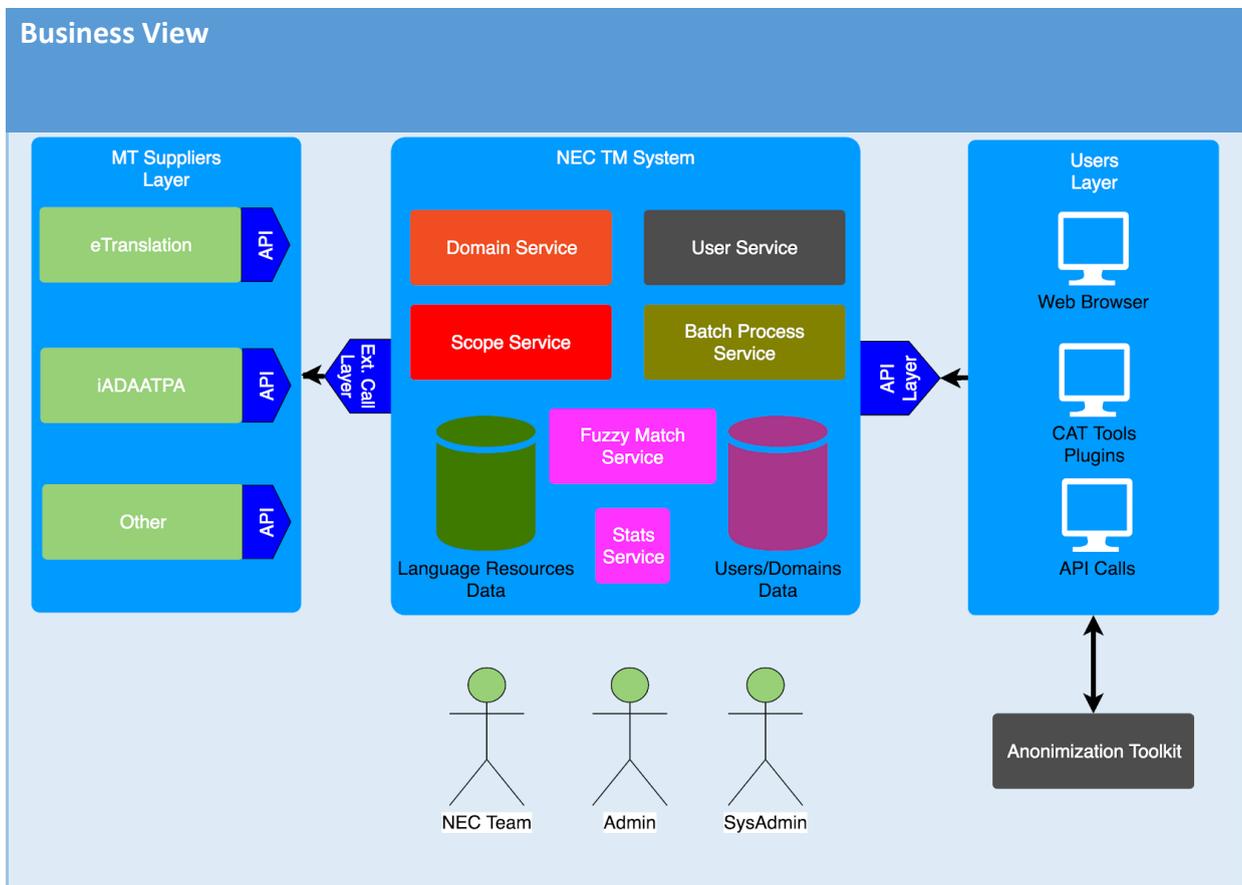


Actor	Definition/Description
NEC TM System	This is the instance installed by the NEC Team. If the installation is on the central administration level, it will be called "Central NEC TM", if the installation is on the national administration level, it will be called "National NEC TM".
NEC Team	This is a NEC TM administrator who installs an instance of NEC TM and makes the initial configuration of the NEC TM System.
Admin	This is an account created by NEC Team to administrate the instance of NEC TM. If the account is on the Central NEC TM, it will be called CentralAdmin. If the account is on the National NEC TM, it will be called NationalAdmin.

SysAdmin	This is an account created by NEC Team to system administrate (backup, monitor, restart...) the NEC TM instance. If the account is on the Central NEC TM, it will be called CentralSysAdmin. If the account is on the National NEC TM, it will be called NationalSysAdmin.
Users	These are users that consume translation services provided by the NEC TM System through API calls, the Web Browser or the Plugins from the CAT Tools.

## Logical View

This section describes the Logical View of the NEC TM system. This represents the component layers and services of the platform.



## External Calls Layer

This layer is responsible for the interaction between MT services and the NEC TM System through the **External Calls Layer**. Interaction with the layer is via RESTful interfaces in the External Services that are called from the NEC TM System.

Component	Description
eTranslation (MT)	This is the API services provided by <b>eTranslation</b> that supply MT translation to the NEC TM Platform.
iADAATPA (MT)	This is the API service provided by <b>iADAATPA</b> that supplies MT translation to the NEC TM Platform.
Other (MT)	These are other API services that supply MT translation to the NEC TM Platform.
<b>External Calls Layer</b>	This layer provides an interface between eTranslation, iADAATPA and other MT Suppliers and the NEC TM Platform. The NEC TM Platform will initiate RESTful calls to the Endpoints (Access Points) for each MT Supplier to access machine translation services for a User. These Endpoints (Access Points) are defined in the <b>MT Supplier Account</b> .

## API Layer

This layer is responsible for the interaction between Users and the NEC TM platform. Interaction with the layer is via REST interfaces which allow access to the **Services and Data in the NEC TM System** and routing to machine translation services provided by **MT Suppliers**.

Component	Description
<b>Web Browser</b>	This is a NEC TM website UI to facilitate the access to the services and data for NEC TM.
<b>CAT Tools Plugins</b>	These are a series of plug-ins/add-ins/connectors for CAT tools that provide connection to the NEC TM Fuzzy Match Service and if the user has permissions to update the TM, it is possible to add finished translations into NEC TM.
<b>API Calls</b>	This is a series of API Calls from third party applications or other instances of NEC TM installations.
<b>Users Layer</b>	This layer exposes all of the services and data provided by NEC TM. This is a full services layer with access to Domain services, User services, Batch Process Service, Scope Service, Fuzzy Match Service and MT services. These RESTful services will provide the core functionality that can be designed and implemented into a plug-in/add-in/connector.

**NOTE:** An Anatomical View of an API Calls presented later in this document.

## Service Layer

This layer provides the core functionalities of the NEC TM platform. These services will be used by both Suppliers and Consumers.

Component	Description
<b>User Management</b>	This is a functionality intended to create, modify and delete users. We can also assign scopes to restrict the access to the Languages Resources repository.
<b>Domain Management</b>	This is a functionality intended to create and delete domains. These domains can later be assigned to uploaded TMXs, or added to translations (through the CAT Tools plugins) , and several or no domains can be selected. These domains are intended for categorization purposes. There are domains which will be used for the public to share.
<b>Scope Management</b>	This is a functionality intended to create, modify and delete permissions assigned to users. Permissions can be defined by using the following parameters: <ul style="list-style-type: none"> <li>● Domains</li> <li>● Language combination</li> <li>● Start date/ end date</li> <li>● Upload/ download data</li> <li>● Add new translations through the plugins in the CAT Tools</li> <li>● Amount of times that the scope can be used</li> </ul>
<b>Batch Process Management</b>	This is a functionality intended to modify and delete the Batch Processes which allows the automatic sharing of content between the different instances of NEC TM. This functionality uses API Calls from NEC TM, and a valid User -with the necessary permissions- is required.
<b>Data Stats</b>	This is a functionality intended to view certain stats (like the amount of data per domain, per language combination, the amount of queries by user, amount of MT calls by user...) from the Users and Language Resources in the NEC TM instance.

**NOTE:** SysAdmin is the main responsible to administer the Batch Processes.

---

## Data Layer

This layer provides and stores the data used by the NEC TM system. This data will be stored in a database and file repository.

Component	Description
<b>Language Resources Data</b>	This is a non-relational database that stores all the necessary information about Translations Memories and other data related necessary to run the NEC TM system.
<b>Users Data</b>	This is a relational database that stores all the necessary information about the Users and other data related necessary to run the NEC TM system.

## MT Suppliers

NEC platform has the ability to call external services to provide Machine Translation as a fallback to the typical query/search functionality.

These external calls are triggered after the query/search produces results with a fuzzy-match lower than a pre-defined threshold.

The NEC platforms will be configured by default with the following external access:

Suppliers	Description
iADAATPA	Multi vendor MT platform for PA access
eTranslation	UE DGT Service

The NEC platform provides the required storage functionality to keep the authentication data necessary to access the external services.

---

## Users

These are Public Administrative bodies that require Translation Memory services from **NEC TM**. They access the services of the NEC TM platform via a RESTful interface which is authenticated by a user and password to get a valid **API token**. A customized Add-in/Connector/Plug-in provides a direct connection between CAT Tools and the NEC TM platform.

The purposes of the NEC TM System are defined as :-

Public Administration	Description
SEAD	Public Administration, Spain

## Anonymization Toolkit

In many cases the TMX owner will want his/her data anonymized before uploading it to the NEC platform.

In order to ensure that non-anonymized data never gets uploaded or processed without the owner's control, NEC team will provide an Anonymization Toolkit that will function locally and independently from both the central and national NEC systems.

The Anonymization Toolkit will provide the user/owner with the ability to anonymize a TMX by generating a version where the sensible data is tagged and substituted with placeholders.

The Anonymization procedure will have 3 degrees of aggressivity:

- Weak Anonymization: the user provides a glossary file with terms (personal names, organization names,...) that are to be tagged and substituted.
- Normal Anonymization: the user provides an additional choice of the entity-types that have to be anonymized (names, locations, dates, numbers,...)
- Strong Anonymization: Is the same as normal anonymization but if sensitive information is found, the complete translation-unit is deleted from the outputted TMX.

The output of the Anonymization Toolkit is a modified version of the input TMX. This 'anonymized' TMX should be revised before uploading it to NEC.

## Business View / Usage Case

The Business View / Usage Case describes the interaction between the Users, MT Suppliers and Admins of the NEC TM system.

1. **NEC Team** installs the NEC-Central instance in a single central European (DGT Owned?) location. A **CentralAdmin** is appointed, who is given the credentials and an admin guide. A **SysAdmin** is also chosen, who will be in charge of system administration and typical duties including; backup, monitoring, system restart... and also will be responsible to define the operational parameters of the platform (including the credentials to access external systems). The **SysAdmin** also needs a specific admin guide.
2. **NEC Team** can also install a NEC instance in the national PA and can then appoint a **NationalAdmin**. A **NationalSysAdmin** is also chosen, who will be in charge of system administration and typical duties such as backup, monitoring, restart..., he/ she will also need a specific admin guide.
3. **CentralAdmin** and **NationalAdmin** uses the NEC UI Website to create domains and to load TMX files. Every TMX should be categorized by a domain which will be utilised to restrict access. The domains can be of type "public", "private" or unspecified.
4. **CentralAdmin** can also create **CentralUsers**. Each user is associated to a number of Domains at creation time, this is the Scope. The **Scope** defines what the subset is for the total language database that a user can access. Content that has been characterized with a "private" domain can only be accessed by users with this domain specified in their Scope. Content that is flagged with a "public" domain can be accessed by any users. Two special domain names are created by default, both public: the "Public" and "Share" domains. Any content flagged with the Public domain will be accessed by anyone, no matter the other domain specification. The domains for "unspecified" type (example: legal, automotion, health...) can be selected by all **CentralUsers**, but the content will only be seen if it has a "public" type domain or "private" type domain if the **CentralUser** has access in the **Scopes**. The **Scopes** can be configured to enable a **User** to update, upload and/or download TMX files in Domains that they have access to. The minimal permissions by default enabled for the **CentralUsers** are to submit QUERY operations using the NEC API (CAT tools) or a web Access.
5. **NationalAdmin** can also create **NationalUsers**. Each user is associated to a number of Domains at creation time, this is the Scope. The **NationalScope** defines what the subset is for the total language database that a user can access. Content that has been characterized with a "private" domain can only be accessed by users with this domain specified in their Scope. Content that is flagged with a "public" domain can be accessed by any users. Two special domain names are created by default, both public: the "Public" and "Share" domains. Any content flagged with the Public domain will be accessed by anyone, no matter the other domain specification. Any content with the "Share" domain will have the same access

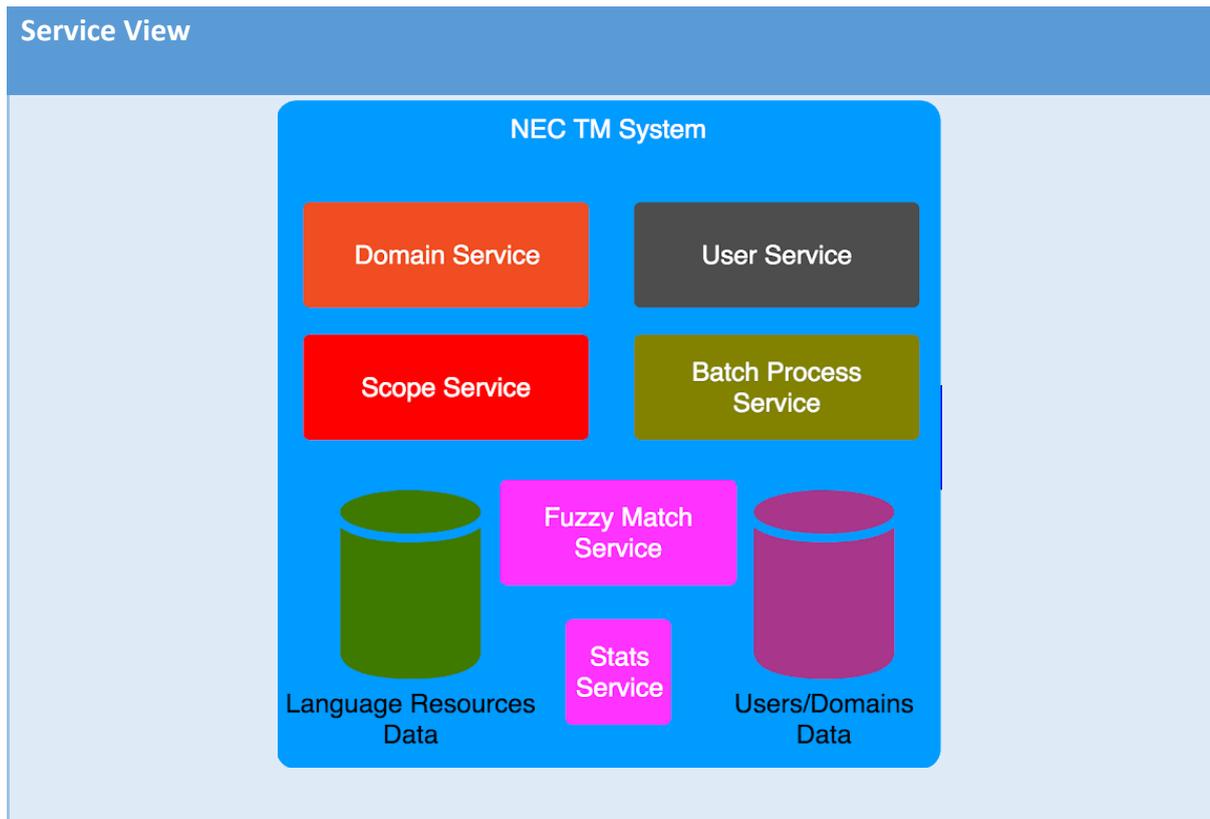
definition as if it were tagged with “Public” but also it will be marked for asynchronous upload to the Central NEC. The domains for “unspecified” type (example: legal, automotion, health...) can be selected by all **NationalUsers**, but the content will only be seen if it has a “public” type domain or “private” type domain if the **NationalUser** has access in the **NationalScopes**. The **NationalScopes** can be configured to enable a **NationalUser** to update, upload and/or download TMX files in Domains that they have access to. The minimal permissions by default enabled for the **NationalUsers** are to submit QUERY operations using the NEC API (CAT tools) or a web Access.

6. Share options in the NEC-Central case:
  - a. The **CentralAdmin** can create “public” type domains that all **CentralUsers** can access.
  - b. The **CentralAdmin** can create “private” type domains that **CentralUsers** cannot access or that only a selection of **CentralUsers** can access.
  - c. **CentralUsers** cannot create new domains.
  - d. The **CentralAdmin** can access all “private” type domains.
  - e. **CentralUsers** can add upload/download TMXs to “public” type domains or to “private” domains that they have access to.
  - f. **CentralUsers** can select several domains (from permitted domains) when uploading TMXs.
  - g. **CentralUsers** can also select “unspecified” type domains.
  - h. **CentralUsers** can add a new “public” type domain to content that already has a “private” type domain that they have access to.
  - i. The **CentralAdmin** can add new “public” type domain to content that already has a “private” type domain.
  - j. If **CentralUsers** want to share anonymised data from a “private” domain, he/ she should download the data, use the “anonymisation workflow” and upload it in a “public” domain.
  - k. The **NationalUser** can upload and/or download data using a valid **CentralUser** account that has upload and/or download permissions.
  
7. Sharing options in a national instance:
  - a. The **NationalAdmin** can create “public” type domains that all **CentralUsers** can access.
  - b. The **NationalAdmin** can create “private” type domains that **NationalUsers** cannot access or that only a selection of **NationalUsers** can access.
  - c. **NationalUsers** cannot create new domains.
  - d. The **NationalAdmin** can access all “private” type domains.
  - e. **NationalUsers** can add upload/download TMXs to “public” type domains or to “private” domains that they have access to.
  - f. **NationalUsers** can select several domains (from permitted domains) when uploading TMXs.
  - g. **NationalUsers** can also select domains of an “unspecified” type.

- h. **NationalUsers** can add a new “public” domain to content with an already “private” domain that they have access to.
  - i. The **NationalAdmin** can add a new “public” type domain to content that already has a “private” domain.
  - j. If **NationalUsers** want to share anonymised data from a “private” domain, he/ she should download the data, use the “anonymisation workflow” and upload it in a “public” domain.
8. Automatic sharing:
- a. An automatic **BatchProcess** will take place every certain time (this can be adjusted by the **NationalSysAdmin**), where data with the “Share” domain is exported and uploaded to the central instance (using a valid **CentralUser** domain with import permissions).

# Service View

This section describes the Service View of the NEC TM System.



## Fuzzy Match Services

There are 2 fuzzy match services that need to be provisioned for the NEC TM system. These are:-

1. Search translation memory segments
2. Batch search translation memory segments

### Search translation memory segments

The service used by User and Admin to search translation segments. These fields can be specified/modified during the use of this service:-

## Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Text	Text to be searched
Source language	Source language of the text searched
Target language	Target language of the translation units found where the source language is similar to the text searched
Limit	Limit the number of segments in the output
Minimum match	Return only matches above or equal to a threshold
Domain	Filter segments by domain(s)

### Output

Translation	Units matching the query
-------------	--------------------------

### Batch search translation memory segments

Service used by User and Admin to batch search translation segments (it is possible to specify different texts to be searched). These fields can be specified/modified during the use of this service:-

### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Text	Text to be searched ( <b>multiple values allowed</b> )
------	--

Source language	Source language of the text searched
Target language	Target language of the translation units found where the source language is similar to the text searched
Limit	Limit the number of segments in the output
Minimum match	Return only matches above or equal to a threshold
Domain	Filter segments by domain(s)

## Output

Translation	Units matching the query
-------------	--------------------------

## User Management Services

There are 6 user management services that need to be provisioned for the NEC TM system. These are:-

1. Add/update user details
2. Delete user
3. Get user details
4. Add/update user scope
5. Delete user scope
6. Authorization

### Add/update user details

Service used by Admin to add and update Users. These fields can be specified/modified during the use of this service:-

## Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Username	Name of User
Password	Password for User Account
User role	Either "user" or "admin".

### Error 4xx

403	Insufficient permissions
-----	--------------------------

### Delete user

Service used by Admin to delete Users. These fields can be specified/modified during the use of this service:-

### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Username	Name of User
----------	--------------

### Error 4xx

403	Insufficient permissions
404	User doesn't exist

### Get user details

Service used by Admin to get User details (included his/ her scopes). These fields can be specified/modified during the use of this service:-

#### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

#### Parameters

Username (optional)	Name of User, if not user specified, allows users' details to be returned
---------------------	---

#### Error 4xx

403	Insufficient permissions
404	User doesn't exist

#### Output

Users	List of user details matching the query
-------	---

### Add/update user scope

Service used by Admin to add and update User scopes. These fields can be specified/modified during the use of this service:-

#### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

## Parameters

Username	Name of User
Language pairs (optional)	List (separated with comma) of allowed language pairs to query, ex.: en_es,es_en,es_fr. By default, allows all pairs
Domains (optional)	List (separated with a comma) of allowed domains to access.
Can update (optional)	Allow/forbid update TM in the given scope. Default is false.
Can import (optional)	Allow/forbid import new TMX files in the TM. Default is false.
Can export (optional)	Allow/forbid export content from the TM. Default is false.
usage_limit (optional)	Limit allowed usage (queries number). Default: no limit
Start date (optional)	Scope starts at that date. Default - not limited
End date (optional)	Scope ends at that date. Default - not limited

## Error 4xx

403	Insufficient permissions
404	User doesn't exist

### Delete user scope

Service used by Admin to delete User scopes. These fields can be specified/modified during the use of this service:-

## Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

Parameters	
Username	Name of User
ID	Scope ID

Error 4xx	
403	Insufficient permissions
404	User doesn't exist

### Authorization

Service to get Authorization Token. These fields can be specified/modified during the use of this service:-

Parameters	
Username	Name of User
Password	Password for account

Error 4xx	
401	Invalid credentials

Output	
Token	Token needed for security reasons in other services

## Domain Management Services

There are 2 domain management services that need to be provisioned for the NEC TM system. These are:-

1. Add/update domain
2. Delete domain
3. Get domain details

### Add/update domain

Service used by Admin to add and update Domains. These fields can be specified/modified during the use of this service:-

Authorization	
Token	Token returned by the Authentication service and is needed here for security reasons

Parameters	
Domain name	The name of the domain to be added
Type domain	Type of domain: public, private or unspecified. Default: unspecified

Error 4xx	
403	Insufficient permissions

### Delete domain

Service used by Admin to delete Domains. These fields can be specified/modified during the use of this service:-

### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Domain name	Name of the domain to be added
-------------	--------------------------------

### Error 4xx

403	Insufficient permissions
404	Domain doesn't exist

### Get domain details

Service used by Admin and Users to get Domain details (Users will only see the domains they have access). These fields can be specified/modified during the use of this service:-

### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

### Parameters

Domain name (optional)	Name of Domain, if not domain specified, alls domain' details are returned.
------------------------	---

### Error 4xx

403	Insufficient permissions
404	Domain doesn't exist

## Output

Domains	List of domain details matching the query
---------	---

## Import Services

There are 2 import management services that need to be provisioned for the NEC TM system. These are:-

1. Import translation memory segments from TMX file
2. Add new translation memory unit

### Import translation memory segments from TMX file

Service used by Admin and Users to import translation memory segments from a TMX file. These fields can be specified/modified during the use of this service:-

#### Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

#### Parameters

File	TMX file to import
Domain	Domain(s) of the imported file

Language pair (optional)	Language pair to import (for multilingual TMX files)
--------------------------	--

### Add new translation memory unit

Service used by Admin and Users to add a translation memory segment. These fields can be specified/modified during the use of this service:-

Authorization	
Token	Token returned by the Authentication service and is needed here for security reasons

Parameters	
Source text	Source segment text
Target text	Target segment text
Source language	Source language
Target language	Target language
Domain	Translation unit domain(s)
Filename (optional)	Filename (or source name)

## Export Services

There is one export management service that needs to be provisioned for the NEC TM system. This is:-

1. Export translation memory segments to TMX file

### Export translation memory segments to TMX file

A service used by User and Admin to export translation segments to TMX file. These fields can be specified/modified during the use of this service:-

## Authorization

Token	Token returned by the Authentication service and is needed here for security reasons
-------	--

## Parameters

Source language	Source language of the translation units
Target language	Target language of the translation units
Domain (optional)	Filter segments by domain(s)
Filename	Filter segments by filename

## Output

binary	TMX containing the translation units matching the query
--------	---

## Data Services

There are three data services that need to be provisioned for the NEC TM system. These are :

1. Language Resource Management
2. User Registration & Management
3. MT Supplier Registration & Management

**NOTE:** Data Services 1 and 2 are Database oriented while Data Service 3 is a file store.

### Language Resource Management

A Translation Memory uploaded through a TMX (using the NEC TM UI or API Call) or added through a the CAT Tool plugin, it can sumistrate certain data that can be used later in the interaction with

other services (like Fuzzy Match or Export. These data details will be stored in a Database which is part of the Data Layer of the NEC TM system.

The Language Resource meta data describes each Translation Memory stored. This will include, at a minimum:-

Language Resource Meta Data	
Source Text	Text of the Source Text used later by the Fuzzy Search Service
Target Text	Text of the Target Text that returns the Fuzzy Search Service
Source Language	Source Language code (example: ES, FR, DE...)
Target Language	Target Language code (example: ES, FR, DE...)
Filename	List field of the different filenames indicating where the translation unit comes from.
Domain	List field of the different domains assigned to the translation unit.

### User Registration & Management

Users will require secure access to the NEC TM system. Access to this account will be secured by the use of an Account Name and Secure Password which will be stored.

The User Registration & Management meta data describes the User. This will include, at a minimum:-

User Meta Data	
Username	Name of User
User role	Either "user" or "admin". Admin has access to all NEC TM interfaces, User is limited to where the scope allows.
Password	Password for User Account
is_active	Whether user is active. Inactive user's access will be denied
Scopes	Limits applied to the User by Admin so can access or make changes where is defined.

## External MT Supplier Management

A Supplier needs the facility to define their MT Services so that Consumers can interact with them. This registration details will be stored in a file which is part of the Data Service Layer of the NEC TM system.

The Supplier Registration meta data describes the information required to access the MT service. This will include, at a minimum:-

MT Supplier Meta Data	
APIKey / Application Token	Authentication information
URL of Endpoint	External services API Endpoint

---

**NOTE #1:** Provisions will be done to at least access iADAATPA and eTranslation.

---