European Language Resource Coordination
Connecting Europe Facility

Atelier ELRC France

The European Language Resource Coordination

La Coordination Européenne des Ressources Linguistiques

ELRC

European Language Resource Coordination
Connecting Europe Facility

Atelier ELRC France

Préparer et partager les données avec le service de dépôt ELRC

Khalid Choukri, Valérie Mapelli, Hélène Mazo
ELRA/ELDA

Retour sur la notion de ressources linguistiques

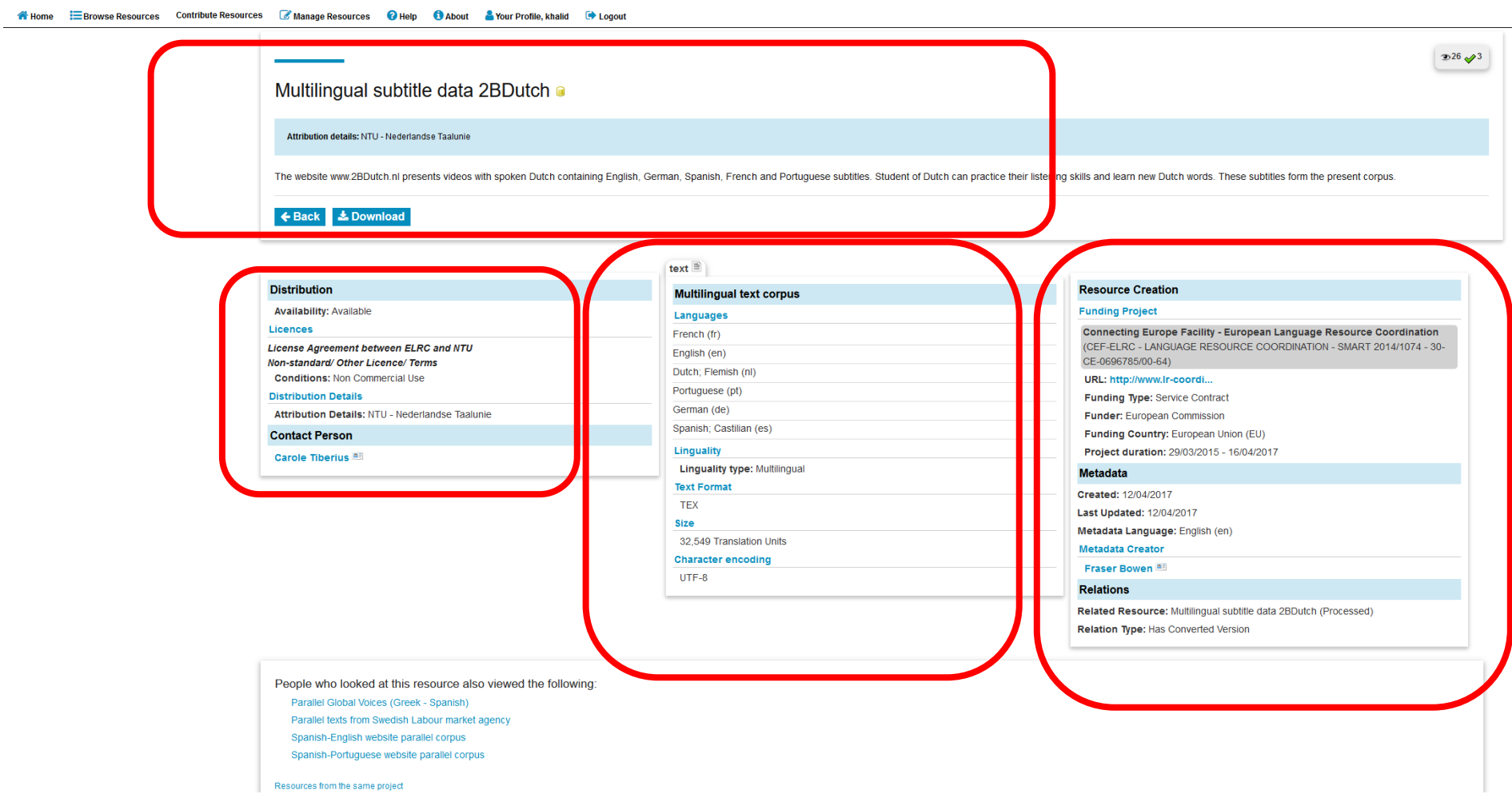
Données

- Tout contenu numérique

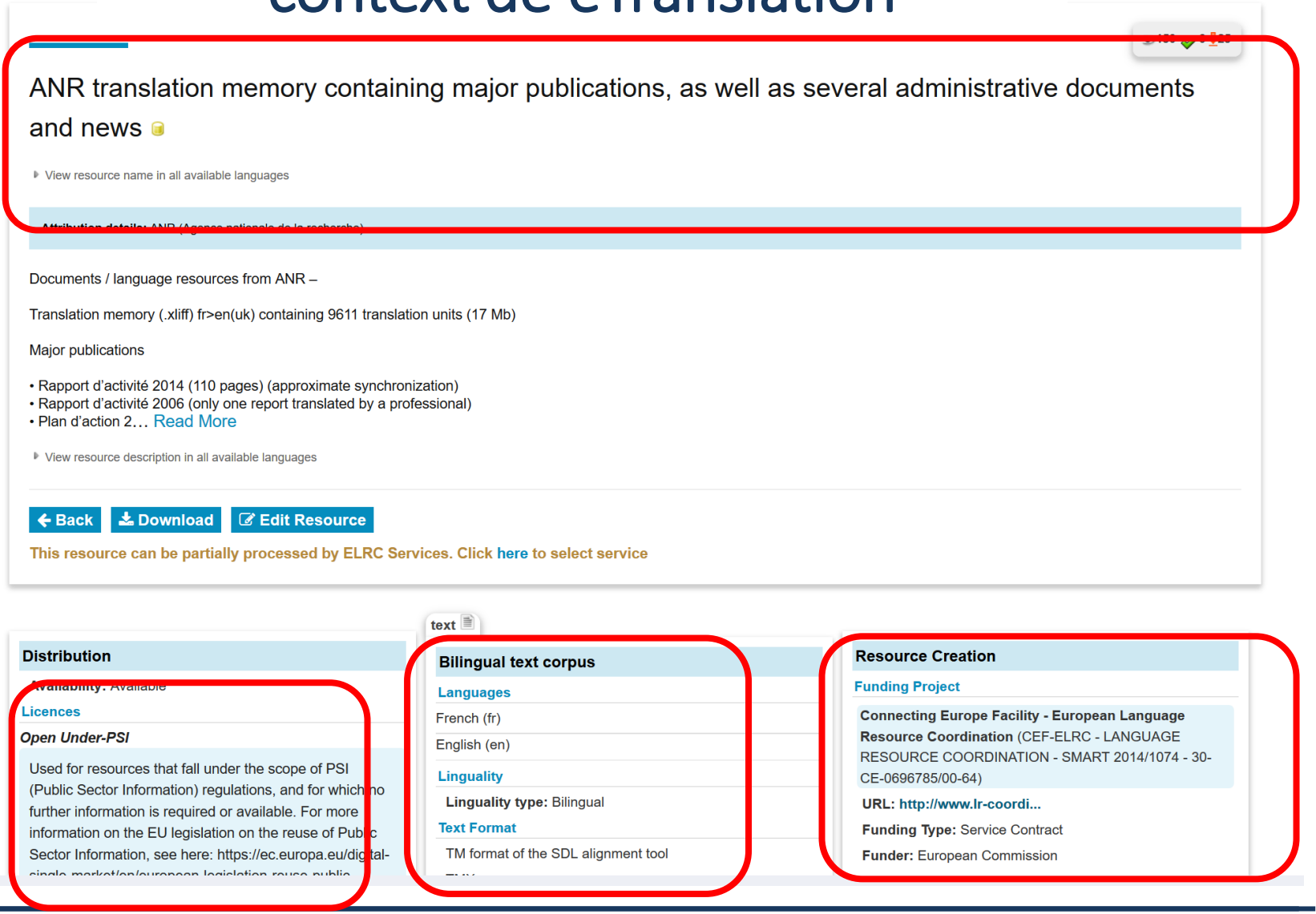
Données Linguistiques (Textuelles)

- Tout contenu textuel au format numérique

La notion de données dans le context de eTranslation



La notion de données dans le context de eTranslation



La notion de données dans le context de eTranslation

```
File01_it.txt
File01_en.txt
File02_it.txt
File02_en.txt
File03_it.txt
…
```

Trans. Data

```
tuv xml:lang="nl"
changedate="20151214T133604Z">
    <seg>Deze gecoördineerde wet
stelt een regeling voor
verplichte verzekering voor
geneeskundige verzorging en
uitkeringen in; ze organiseert
die in twee onderscheiden takken
die betrekking hebben, de ene op
de geneeskundige verstrekkingen,
de andere op de uitkeringen
wegens arbeidsongeschiktheid […]
en op de
moederschapsverzekering.</seg>
```

```
 </tuv>
   <tuv xml:lang="fr"
changedate="20151214T133604Z">
    <seg>La présente loi
coordonnée institue un régime
d'assurance obligatoire soins
de santé et indemnités; elle
l'organise en deux secteurs
distincts relatifs, l'un aux
prestations de santé, l'autre
aux indemnités d'incapacité de
travail, […] et à l'assurance
maternité.</seg>
   </tuv>
```

The notion of data in the context of eTranslation

File01_nl.txt
File01_en.txt
File02_nl.txt
File02_en.txt
File03_nl.txt
…

Trans. Da

Amsterdam is de hoofdstad van ons land.

Nu is het een grote en drukke stad.

En dit is 'De Dam',

het beroemde plein in Amsterdam.

Amsterdam is the capital of our country

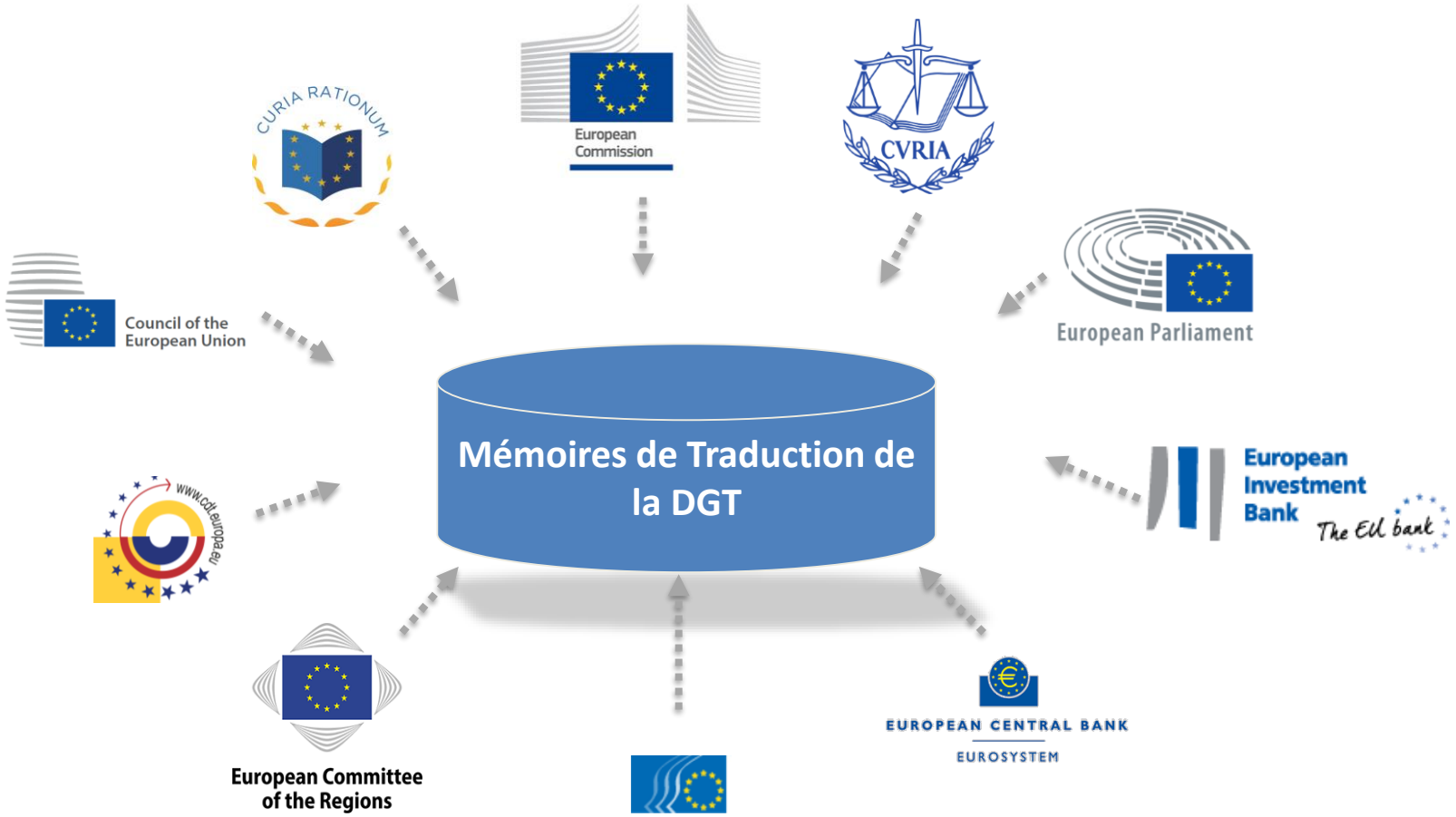
It is a big and busy city

This is Dam square

It is the famous square of Amsterdam

European Language Resource Coordination
Connecting Europe Facility

Données utilisées par eTranslation



Ces données sont déjà utilisées par la DGT mais il en faut beaucoup plus

De quoi a besoin eTranslation?

- Toutes données détenues par des organismes publics,
- Produites en interne ou en sous-traitance
  - Rapports
  - Documents de Communication
  - Documents d'information (Nouvelles)
  - Contenu Web produit en plusieurs langues
  - Politiques
  - Terminologies
  - Archives
  - Formulaires
  - Foires aux Questions (FAQs)
  - ….

Quelles données sont utiles pour eTranslation selon le type | 1

- Tout texte numérique dans une des langues de l'UE + Norvège/Icelande
- **Des textes et leurs traductions (par exemple, bilingues ou multilingues parallèles)**

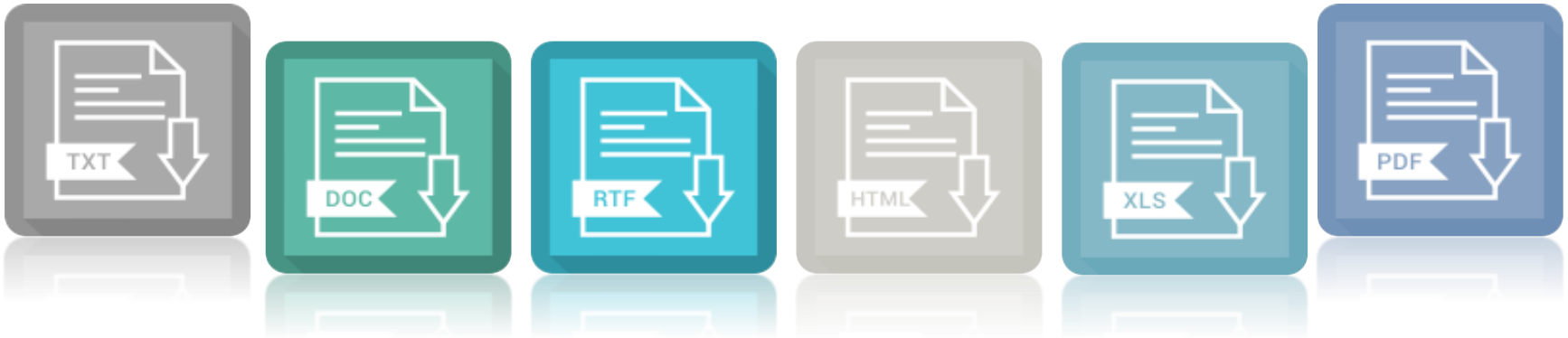
Texte en NL

Ter toepassing van de bepalingen van deze gecoördineerde wet worden de landsbonden gemachtigd die het waren ter toepassing van het organiek koninklijk besluit van 22 september 1955 van de ziekte- en invaliditeitsverzekering

Traduction en Français

Sont agréées pour l'application des dispositions de la présente loi coordonnée les unions nationales qui l'étaient pour l'application de l'arrêté royal du 22 septembre 1955 organique de l'assurance maladie-invalidité.

Quelles données sont utiles pour eTranslation selon le format | 1



- En principe, tout texte au format numérique (format que peut lire des ordinateurs),
- Mais certains formats sont plus "prêts" pour le traitement que d'autres, c'est-à-dire qu'ils nécessitent moins de traitement (manuel ou automatique).
- Plus de traitement introduit plus d'erreurs dans le résultat final, le rendant moins utile pour eTranslation

Quelles données sont utiles pour eTranslation selon le format |2

Les formats suivants sont particulièrement utiles:

- Pour les textes parallèles bilingues / multilingues
  - Mémoires de traduction (.tmx)
  - Fichiers de traduction XML (.xliff)
  - Texte brut (.txt, .csv)
  - Feuilles de calcul (par exemple xlsx)
- Pour les terminologies
  - TermBase eXchange (.tbx)
  - Texte brut (.txt, .csv)
  - Feuilles de calcul (par exemple xlsx)
- Pour les textes monolingues
  - Texte brut (.txt, .csv)

European Language Resource Coordination
Connecting Europe Facility

Formats de fichiers de textes parallèles et leur manipulation

Preparation de données |1

A ne pas faire!!

Ter toepassing van de bepalingen van deze gecoördineerde wet worden de landsbonden gemachtigd die het waren ter toepassing van het organiek koninklijk besluit van 22 september 1955 van de ziekte- en invaliditeitsverzekering.

Sont agréées pour l'application des dispositions de la présente loi coordonnée les unions nationales qui l'étaient pour l'application de l'arrêté royal du 22 septembre 1955 organique de l'assurance maladie-invalidité.

De landsbonden waarborgen in hun statuten de bij deze wet bedoelde prestaties.

Les unions nationales garantissent, dans leurs statuts, les prestations prévues par la présente loi.

De machtiging van landsbonden die deze gecoördineerde wet of haar uitvoeringsbesluiten en –verordeningen niet naleven kan door de Koning, op advies of voorstel van het Algemeen comité van het Instituut, worden ingetrokken.

L'agréation peut être retirée par le Roi, sur avis ou sur proposition du Comité général de l'Institut, aux unions nationales qui n'observent pas la présente loi coordonnée ou ses arrêtés et règlements d'exécution.

Ne pas fusionner le texte source et le texte traduit (cible) en un seul document

Preparation de données |2

A ne pas faire!

Ter toepassing van de bepalingen van deze gecoördineerde wet worden de landsbonden gemachtigd die het waren ter toepassing van het organiek koninklijk besluit van 22 september 1955 van de ziekte- en invaliditeitsverzekering.

Sont agréées pour l'application des dispositions de la présente loi coordonnée les unions nationales qui l'étaient pour l'application de l'arrêté royal du 22 septembre 1955 organique de l'assurance maladie-invalidité.

De landsbonden waarborgen in hun statuten de bij deze wet bedoelde prestaties.

Les unions nationales garantissent, dans leurs statuts, les prestations prévues par la présente loi.

De machtiging van landsbonden die deze gecoördineerde wet of haar uitvoeringsbesluiten en –verordeningen niet naleven kan door de Koning, op advies of voorstel van het Algemeen comité van het Instituut, worden ingetrokken.

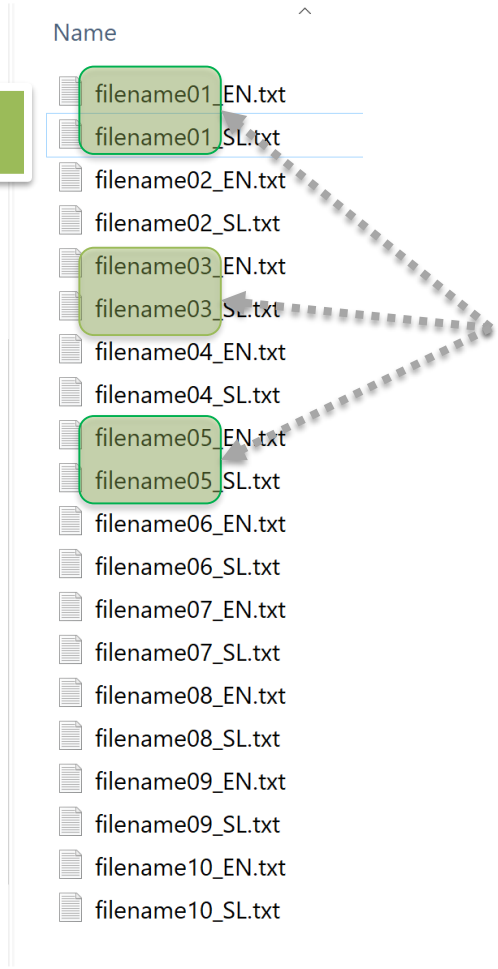
L'agréation peut être retirée par le Roi, sur avis ou sur proposition du Comité général de l'Institut, aux unions nationales qui n'observent pas la présente loi coordonnée ou ses arrêtés et règlements d'exécution.

Preparation de données |3

A ne pas faire!

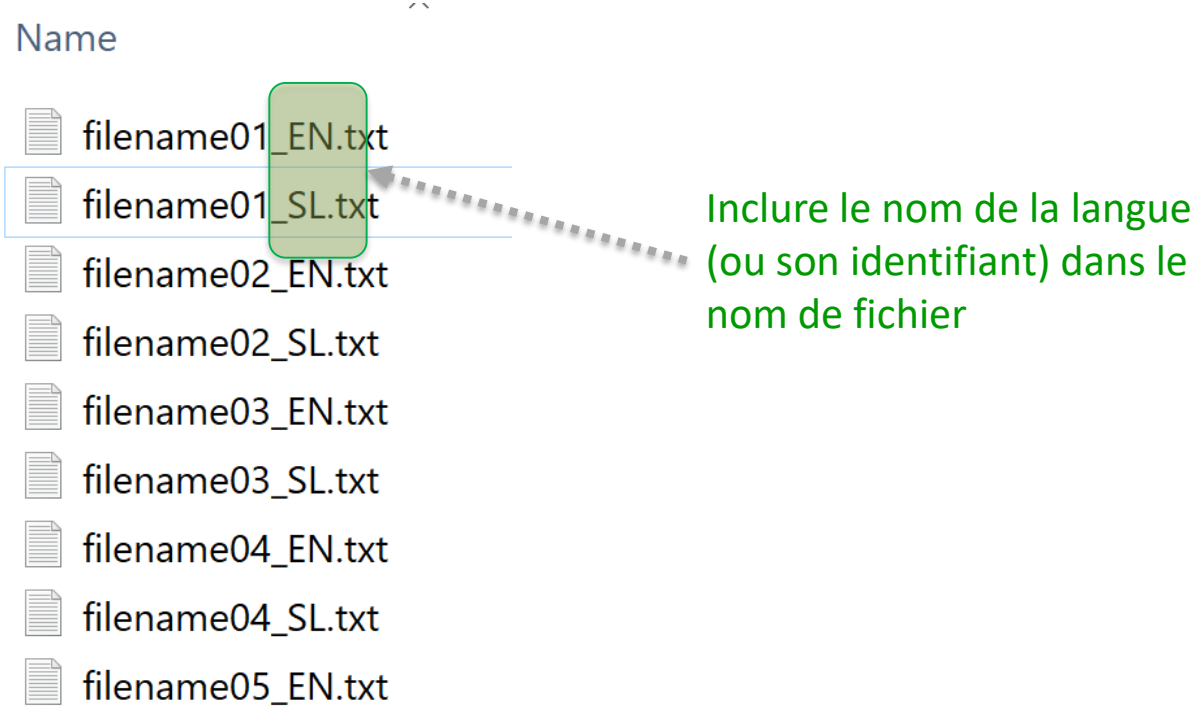
Ter toepassing van de bepalingen van deze gecoördineerde wet worden de landsbonden gemachtigd die het waren ter toepassing van het organiek koninklijk besluit van 22 september 1955 van de ziekte- en invaliditeitsverzekering.	Sont agréées pour l'application des dispositions de la présente loi coordonnée les unions nationales qui l'étaient pour l'application de l'arrêté royal du 22 septembre 1955 organique de l'assurance maladie-invalidité.
De landsbonden waarborgen in hun statuten de bij deze wet bedoelde prestaties.	Les unions nationales garantissent, dans leurs statuts, les prestations prévues par la présente loi.

Preparation de données |4



Utilisation d'un fichier par langue et la même façon de nommer les paires de fichiers (source – traduction)

Preparation de données |5



Quelques critères de regroupement de vos données

- Un ensemble de données est un ensemble de données "regroupées selon certains critères"
- Afin de renforcer et d'adapter le système eTranslation, deux critères sont essentiels:
  - **Langue** (s): chaque collection est définie par la ou les paires de langues de ses données, par ex:
    - Recueil de textes en anglais - allemand
    - Documents en anglais - norvégien - finnois
  - **Domaine**: chaque collection appartient idéalement à un seul domaine, par exemple:
    - Recueil de textes en anglais - allemand dans le domaine de la culture
    - Documents de sécurité sociale en anglais - norvégien - finnois

Domaines importants

- Domaines des écrits Administratifs / règlementations et tout ce qui concerne les CEF DSI (Infrastructures de Services Numériques)

CEF DSI	Domaines
Résolution des litiges en ligne (ODR)	Droits des consommateurs, litiges
Échange électronique d'informations sur la sécurité sociale,	Sécurité Sociale, Assurance
eProcurement (passation électronique des marchés)	Commande Publique,
Portail e-Justice	Justice, Loi
Portail eHealth	Santé, Médecine,
Business Registers Interconnection System (Registres de Commerce Interconnectés)	Affaires, Marché
Internet Sans Crainte (Safer Internet)	
Cybersécurité	
Données Publics Ouvertes	
Europeana	Culture

European Language Resource Coordination
Connecting Europe Facility

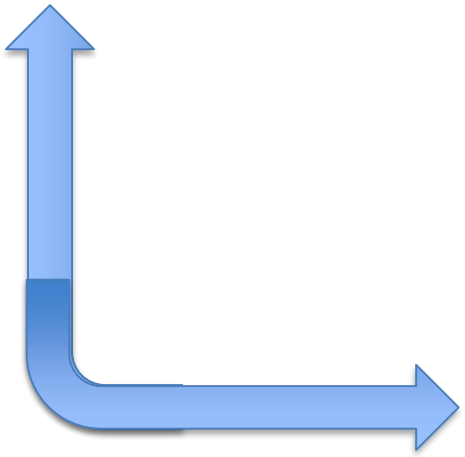
Comment contribuer vos données à eTranslation
Un guide en quelques étapes

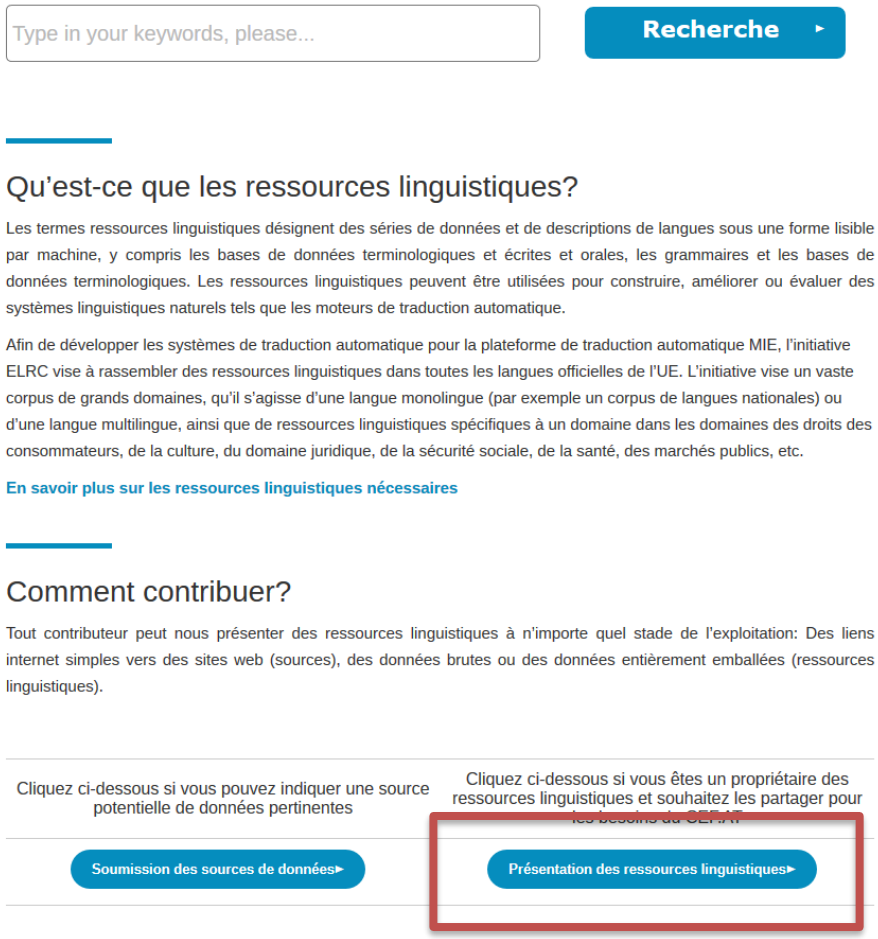
www.lr-coordination.eu vers www.elrc-share.eu

www.lr-coordination.eu vers www.elrc-share.eu

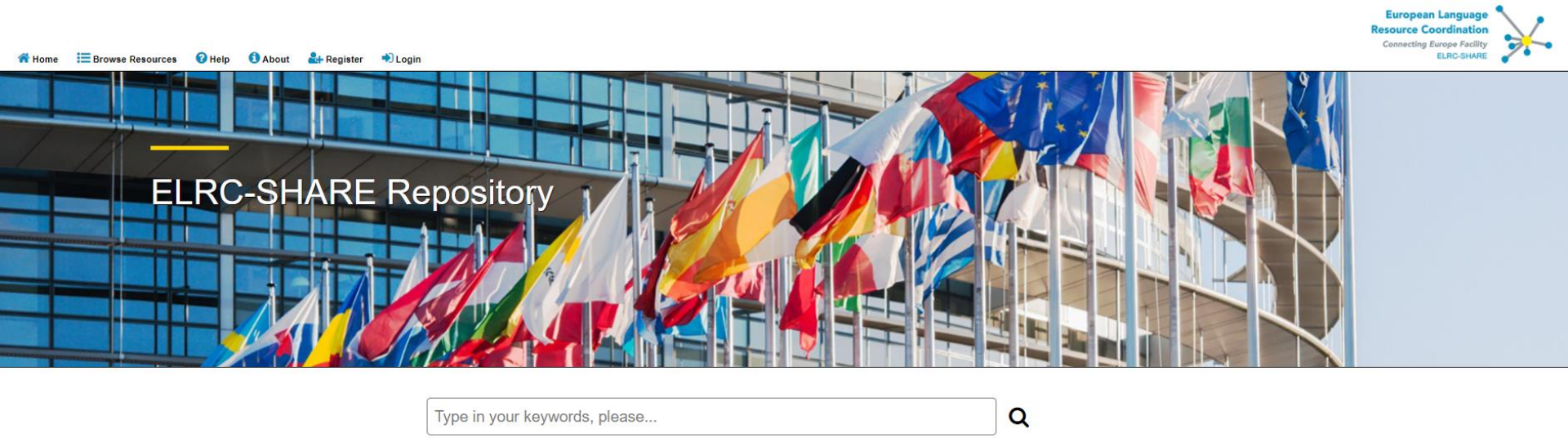
- Sur le portail ELRC, cliquez sur le bouton «Soumission de ressources linguistiques».
  Ou
  Tapez l'adresse URL:

elrc-share.eu

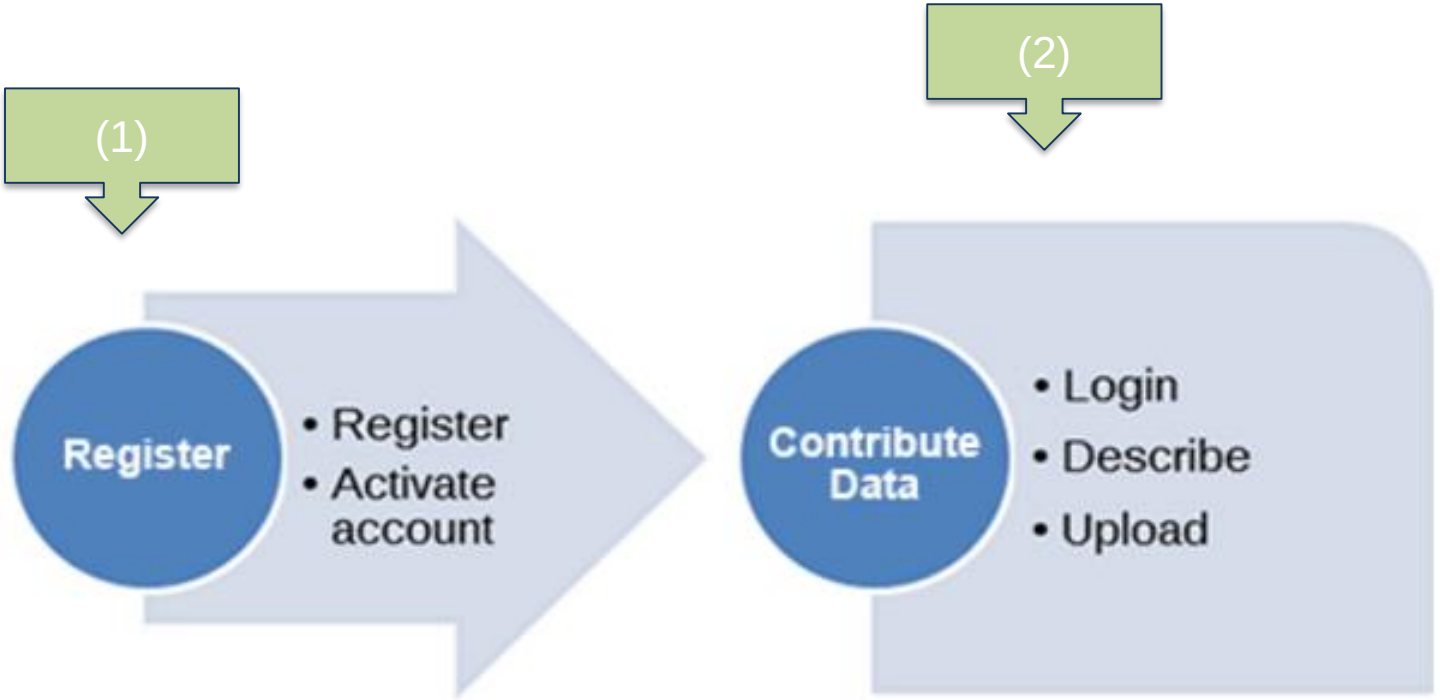




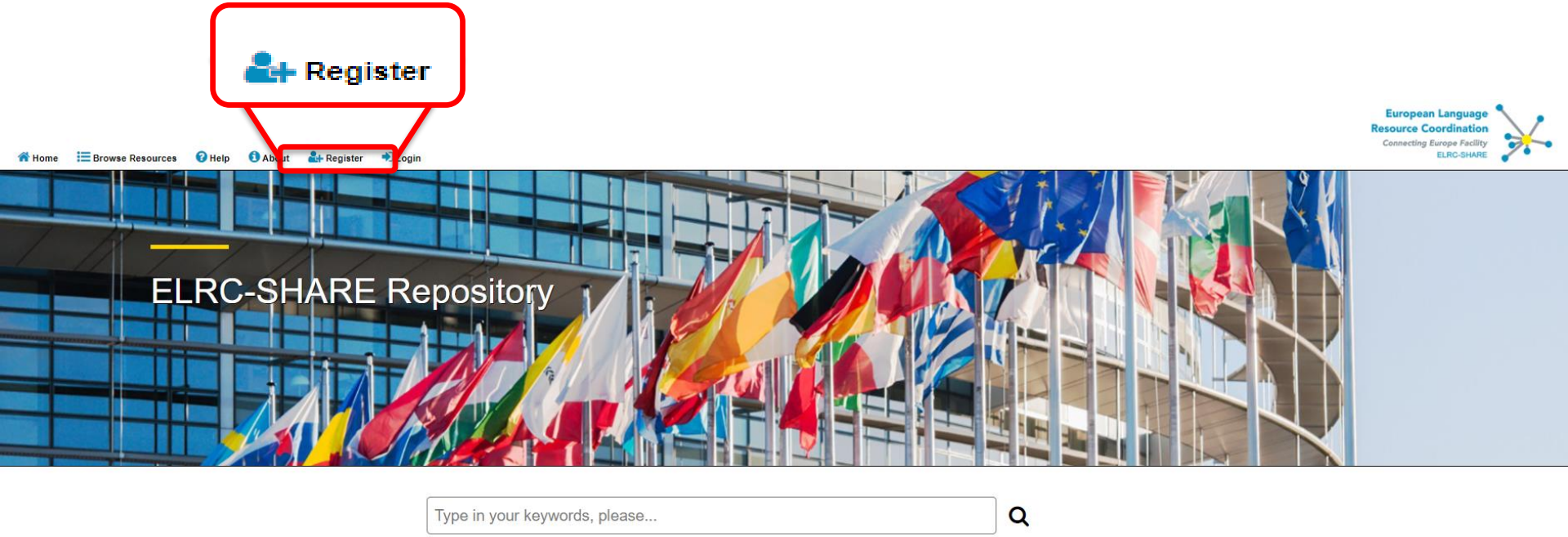
Catalogue / Inventaire ELRC-SHARE



Comment contribuer des données

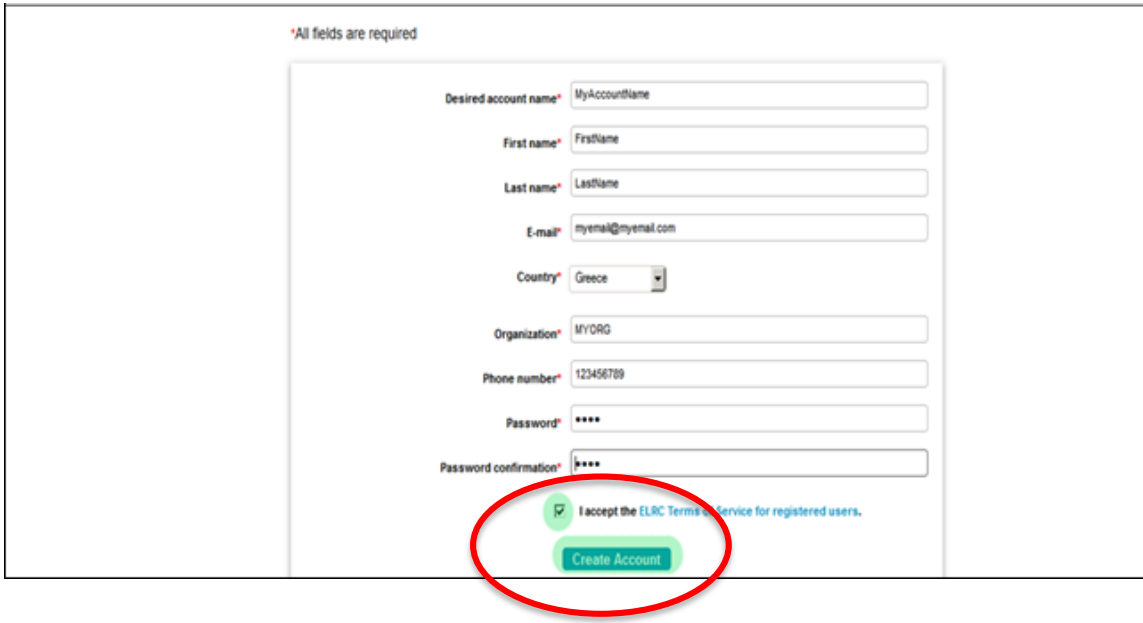


Comment s'enregistrer (1/2)

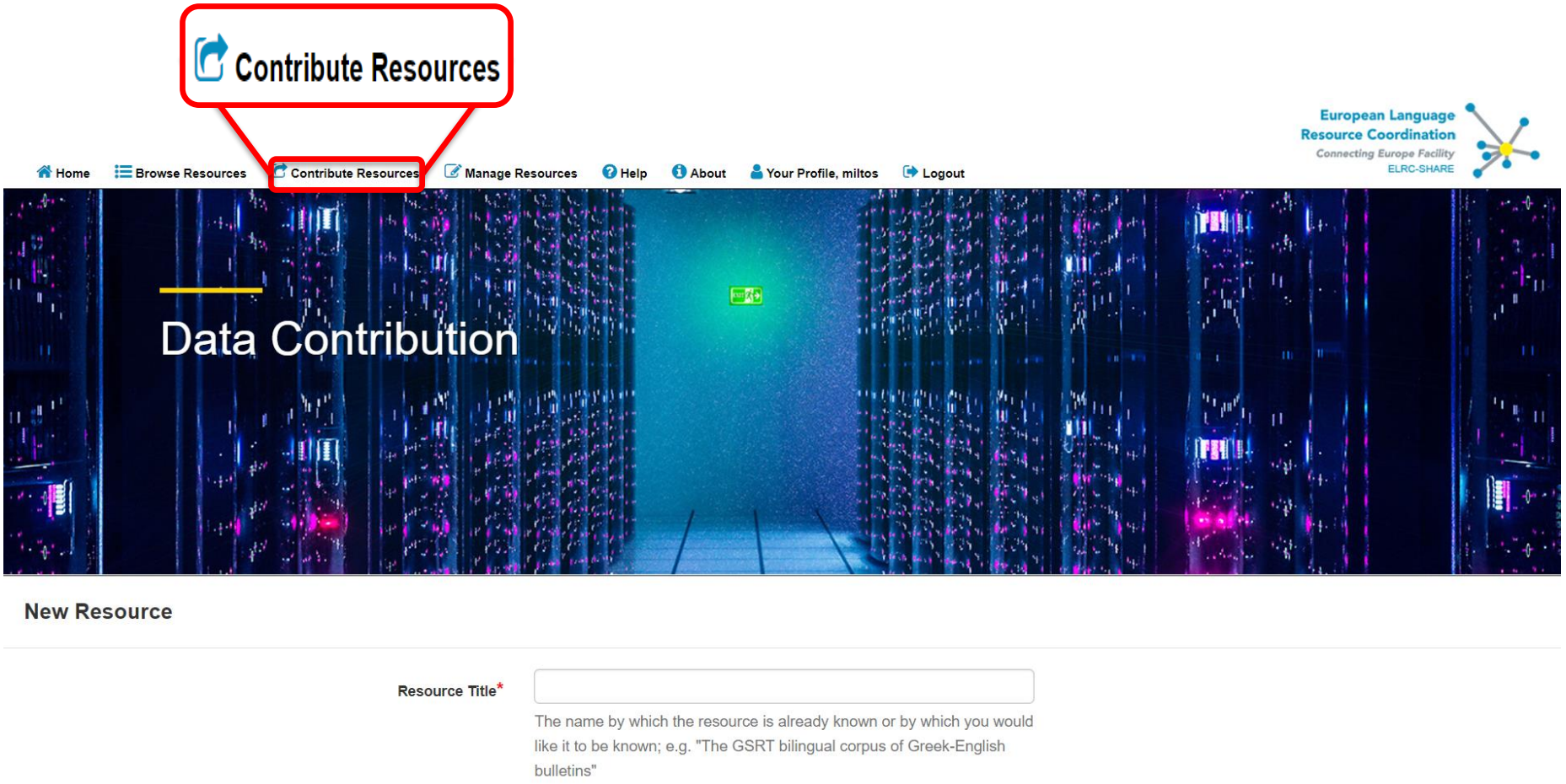


Comment s'enregistrer (2/2)

- Remplissez les informations requises
- Lisez les conditions d'utilisation et cliquez sur "Accepter", si vous êtes d'accord.
- Cliquez sur le bouton "Créer un compte".
- Activez votre compte conformément aux directives qui vous sont envoyées par courriel.

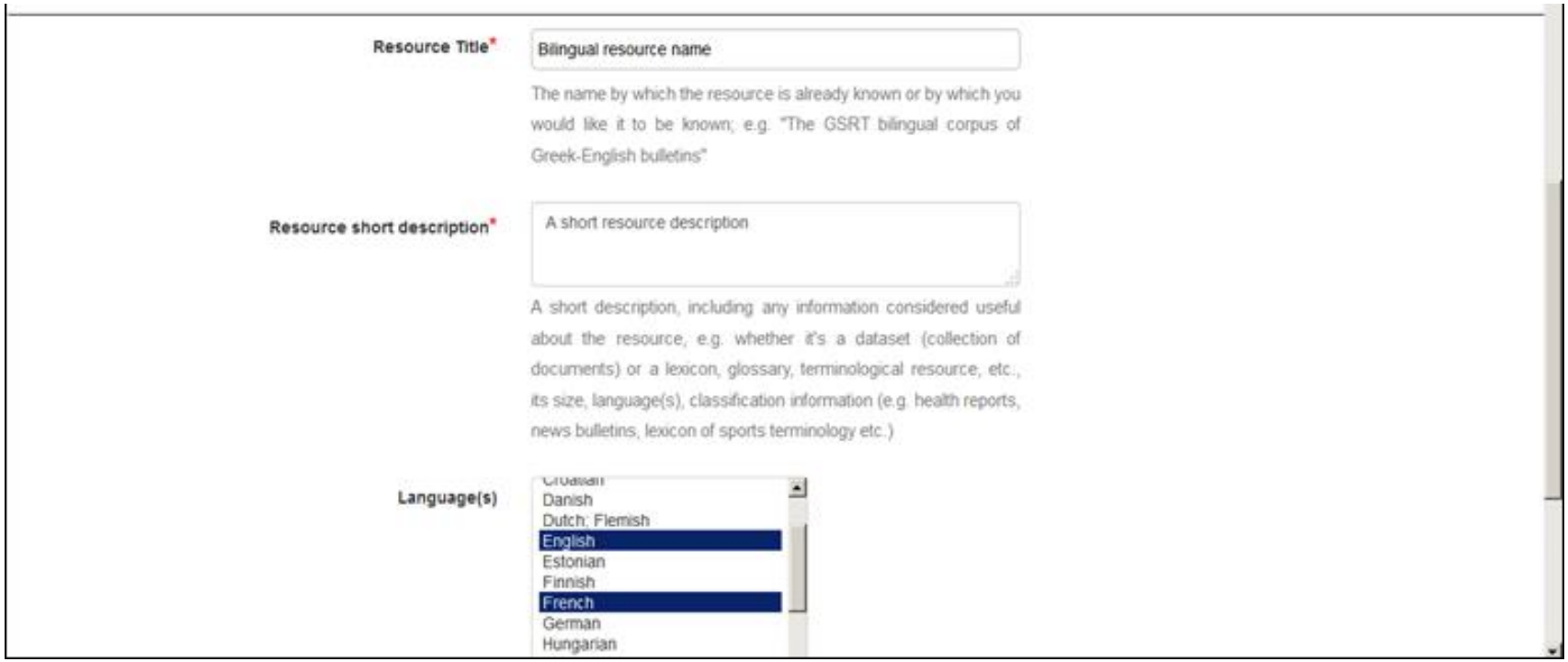


Comment contribuer des données (1/6)



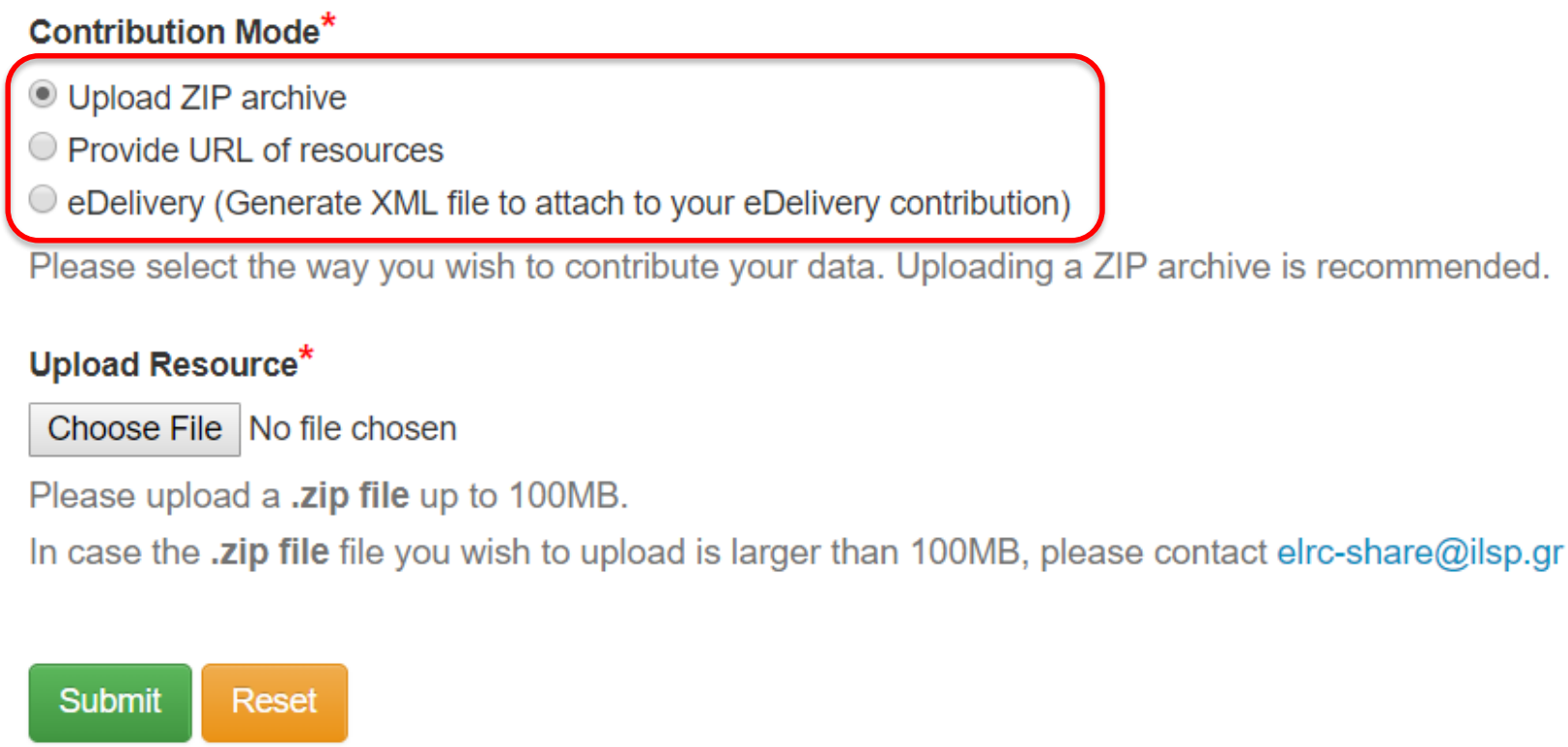
Comment contribuer des données (2/6)

- Remplissez les détails décrivant l'ensemble de données



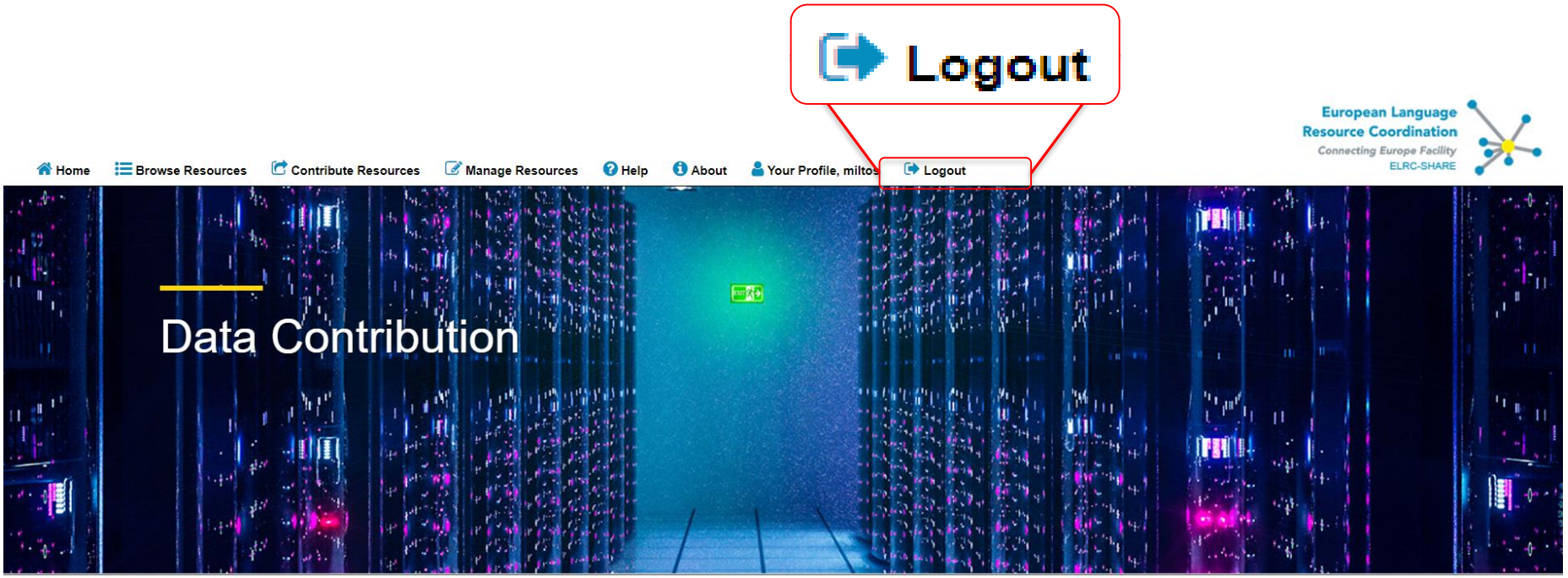
Comment contribuer des données (3/6)

- Trois modes pour contribuer vos données

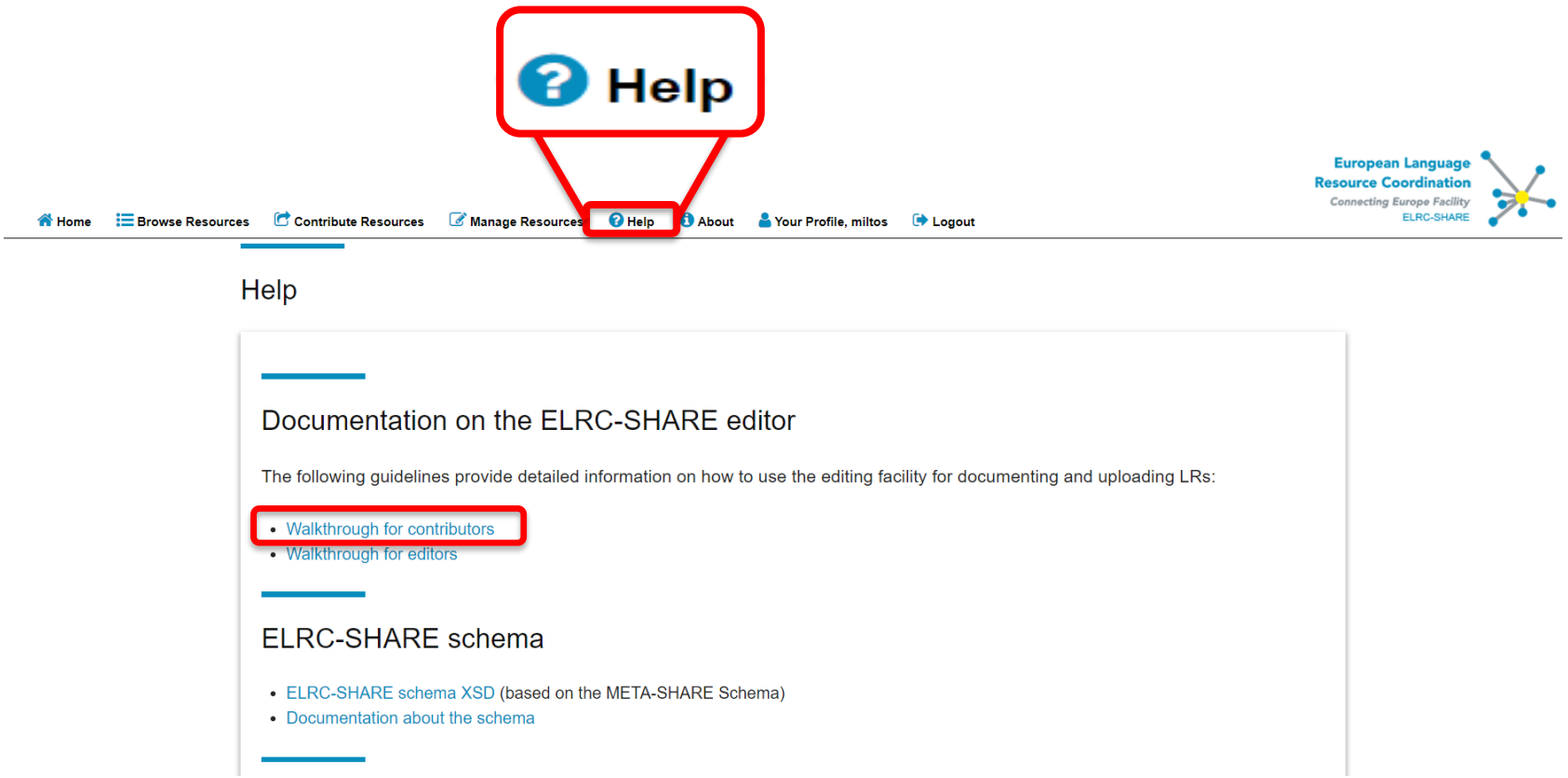


Comment contribuer des données (4/6)

- Repeat the process if you want to contribute another resource, or log out

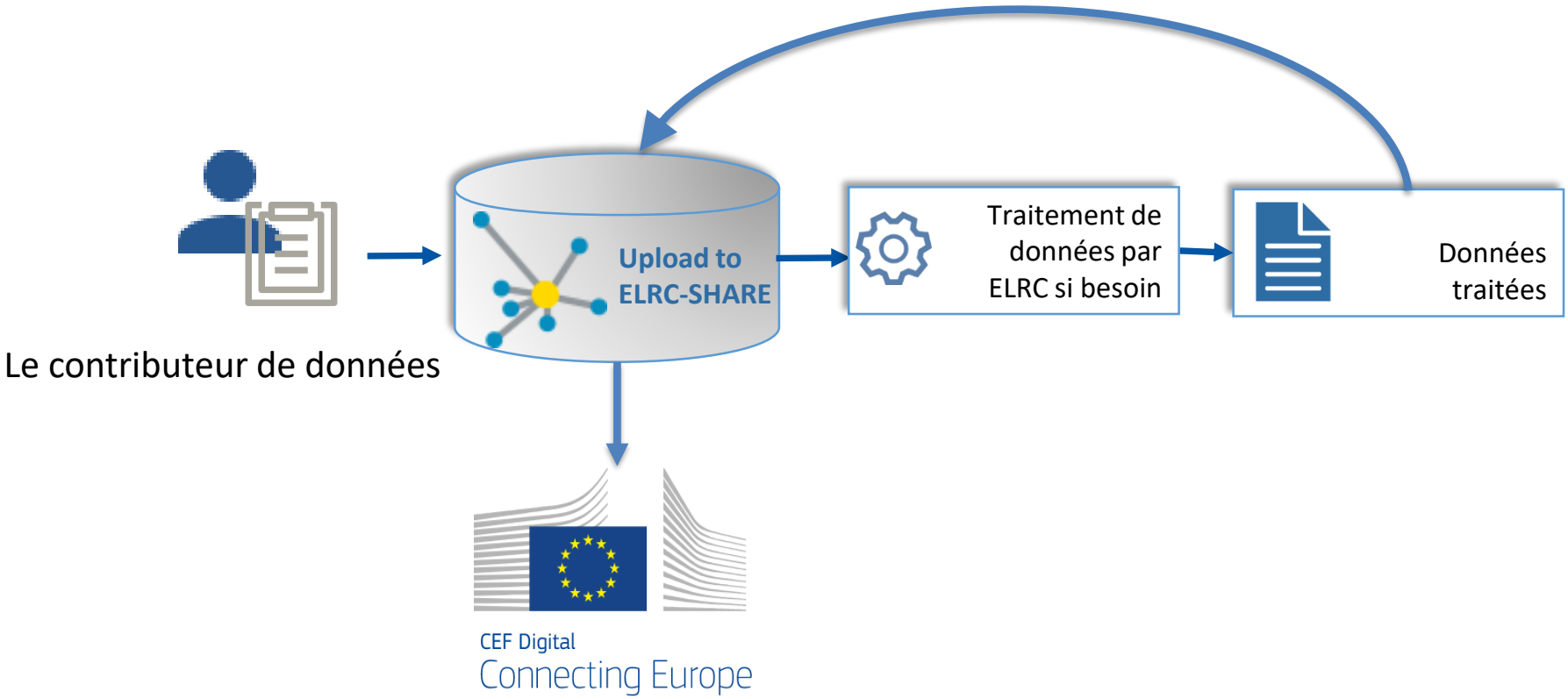


Guide pour les contributeurs

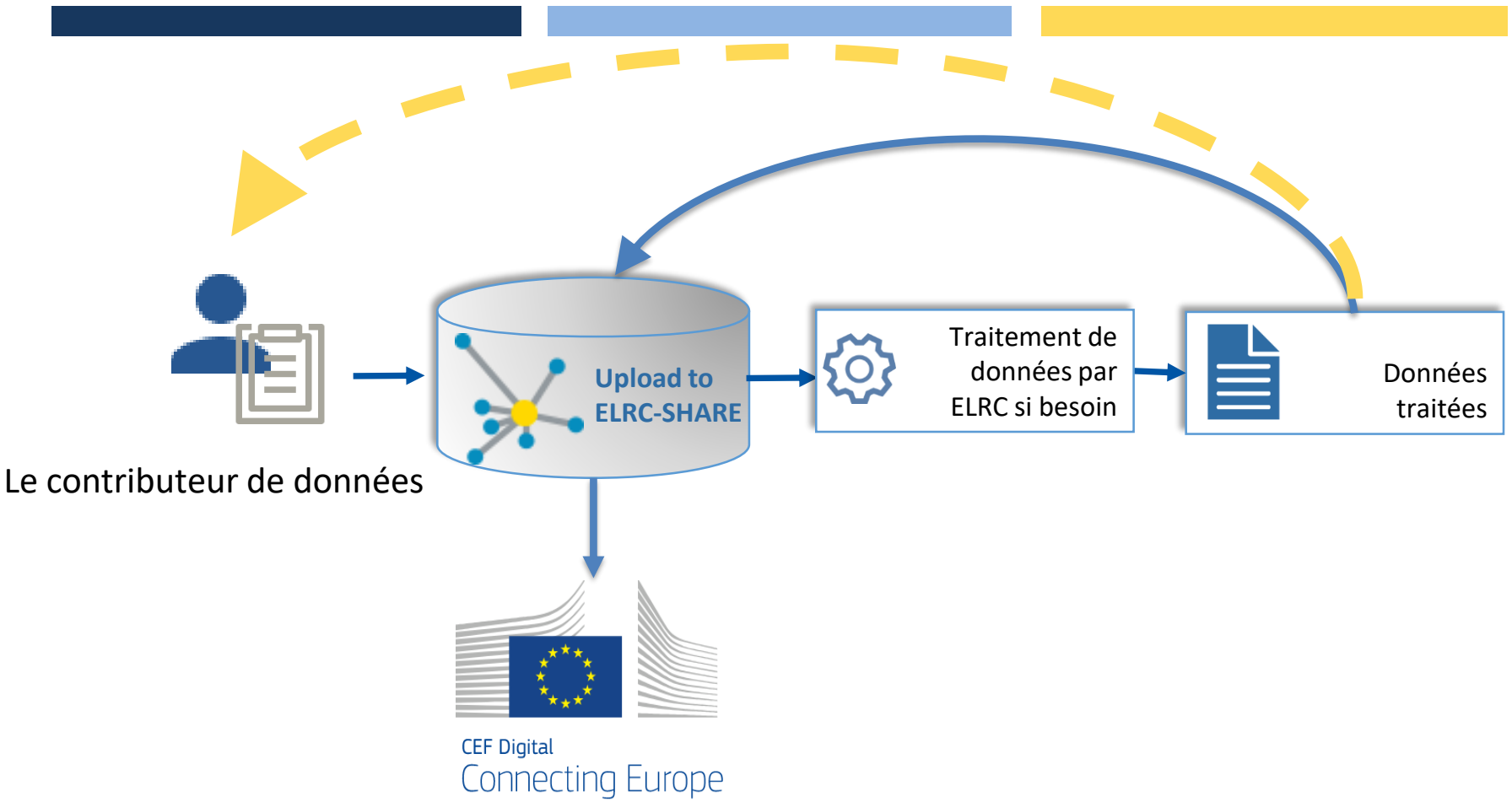


et la suite !!

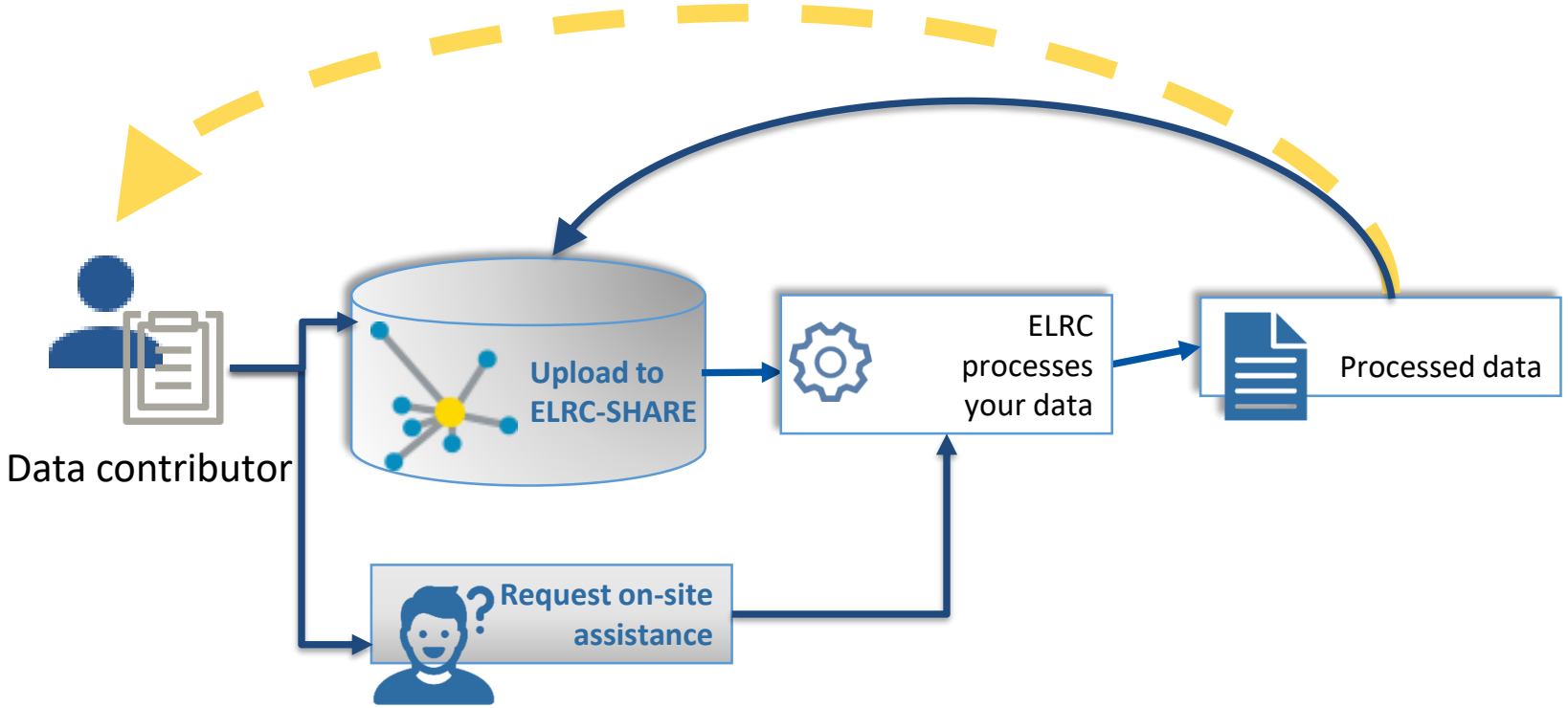
Qu'advient-il de vos données?



Partenariat avec les autres initiatives e.g. ELRI



Partenariat avec les autres initiatives e.g. ELRI



Assistance sur site

Tous les services de préparation de données peuvent également être offerts gratuitement sur votre site
Pour tous les contributeurs de données.

Assistance sur site

L'assistance sera fournie en coopération étroite avec un vaste réseau d'experts

Assistance sur site

Nous pouvons vous assister pour résoudre vos problèmes de données
Une fois "préparées", vous obtenez une copie de cette version.

Nous pouvons également vous aider à améliorer vos processus de gestion des données. Il suffit de demander!

Services de traitement linguistique sur place

Extraction de données

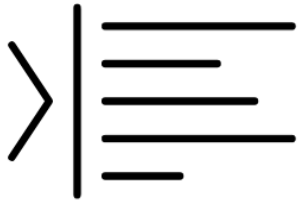
Si vos données sont dans des archives et des bases de données, nous pouvons vous aider à les extraire.

Anonymisation

Vos données contiennent-elles des informations personnelles? Nous pouvons aider à les anonymiser

Nettoyage

Si vos données contiennent beaucoup de "bruit", nous pouvons les nettoyer.



Re-formatage

Besoin de reformater DOCX en XML ou PDF en WORD? On peut vous aider

Comment demander une assistance sur site services et de l'aide

ELRC Assistance sur site

Submit a request for on-site assistance by filling out the form below. See a list of services here.

First name *

Last name *

Institution *

Country *

Email *

Types of assistance required *

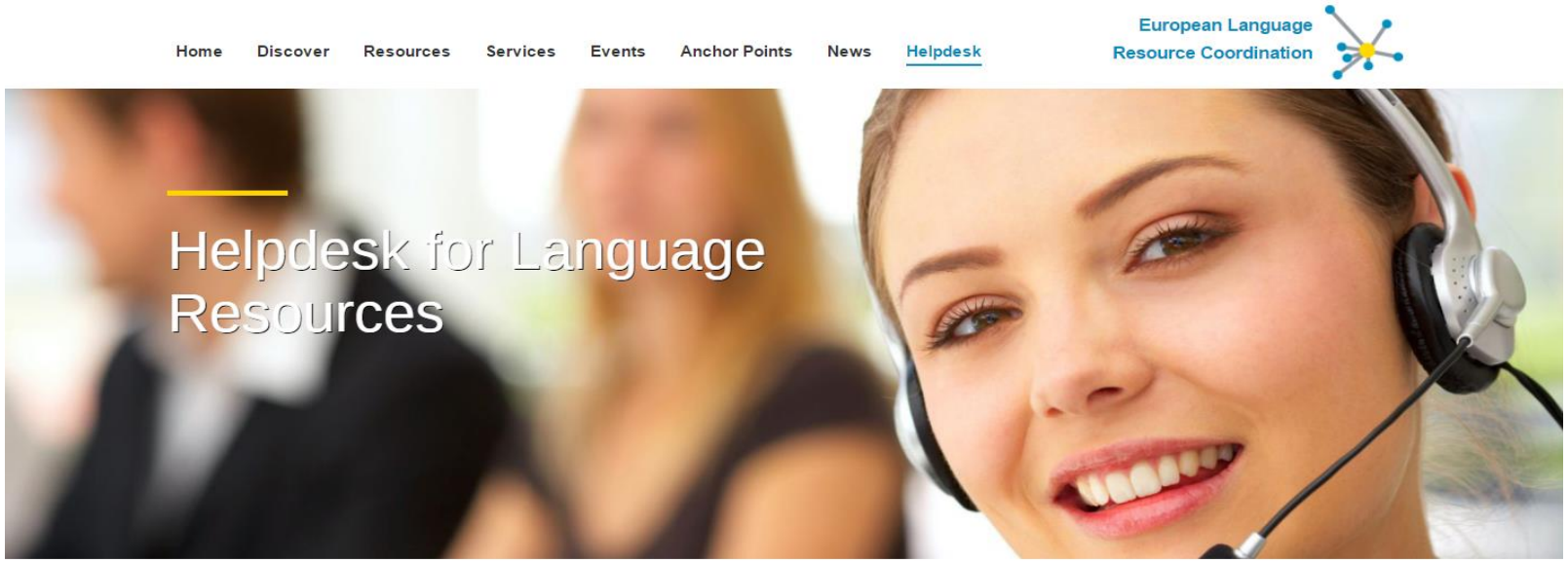
- Legal assistance
- Data processing
- Anonymisation
- Other

Description of assistance required

Submit

lr-coordination.eu/request-onsite-assistance

ELRC Service d'assistance



Helpdesk for Language Resources

We are happy to answer any questions on the technical or legal aspects related to the use, production, collection, processing, and sharing of language resources.

Please feel free to contact us through one of the following channels:

Telephone*	**+33 970 440 522**
Secretariat Support	**+49 681 857 7552 85**
Skype	**ELRC Helpdesk**
E-mail	help@lr-cooridantion.eu

lr-coordination.eu/helpdesk

European Language Resource Coordination
Connecting Europe Facility

Merci de votre attention

Courriel : info@lr-coordination.eu

Site web: www.lr-coordination.eu