

# Translations and Open science

---

## Report of the Study

# Operating model for a technology-aided collaborative translation service dedicated to open scholarly communication

**Authors: Charles Talbot, Rafael Torres**

**Contributors: Jérôme Giusti, Metalaw & Joaquin Maguiña Contreras**

Laboratoire  
d'Intelligence  
Collective &  
Artificielle



## DISCLAIMER

The ideas and views expressed in the exploratory reports only reflect those of the experts involved in the studies and may not be representative of the opinions or policies promoted by any specific organisation, institution, or government entity. The present report is therefore only intended for informational purposes.

## AVERTISSEMENT

Les idées et les perspectives exprimées dans les rapports exploratoires reflètent uniquement celles des spécialistes ayant contribué aux études et ne sont pas nécessairement représentatives des opinions ou des politiques promues par une organisation, une institution ou une entité gouvernementale spécifique. Le présent rapport est donc uniquement diffusé à des fins d'information.

## English summary

This document presents an overview of the study *Operating model for a technology aided, collaborative translation service dedicated to open scholarly communication*. The aim of the study was to identify technical and organisational requirements as well as suggest economic and ethical models for the deployment of the service.

The purpose of this document is to formulate :

- (1) the technical scope and possible uses of a technology-aided translation ecosystem intended for a variety of stakeholders, from researchers to decision-makers;
- (2) suggestions to ensure the economic, ethical and legal sustainability of this system;
- (3) a roadmap for the implementation of the ecosystem.

## Objet de ce rapport

L'objectif général de ce document est d'étudier le périmètre économique, juridique, technique et éthique<sup>1</sup> pour la mise en place d'un service de traduction outillée destiné à la communication scientifique.

---

<sup>1</sup>Source: rapport Traductions et science ouverte (2020).

<https://hal-lara.archives-ouvertes.fr/OUVRIR-LA-SCIENCE/hal-03640511>

« un nouveau modèle d'accès, universel et multilingue, à l'information scientifique, verra le jour. Mais une condition est indispensable : l'humain doit rester au cœur du processus, les technologies devant optimiser le travail sans devenir une contrainte ou une source de frustration pour les utilisateurs, que ce soit les intervenants dans le processus de traduction ou les lecteurs finaux. »

## Table des matières

<b>DISCLAIMER</b>	<b>2</b>
<b>AVERTISSEMENT</b>	<b>2</b>
<b>English summary</b>	<b>3</b>
<b>Objet de ce rapport</b>	<b>4</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. Recommandations d'usage</b>	<b>7</b>
2.1. Consentement et transparence	8
2.2. Rétribution	8
2.3. Qualité de la base de données : biais et discriminations	8
2.4. Déconstruction du mythe "IA de bout-en-bout"	9
<b>3. Fonctionnalités par usage</b>	<b>9</b>
3.1. Usages gratuits	9
3.2. Usages payants	9
3.3. Usages basés sur des contributions	10
<b>4. Modèles économiques et traduction juridique</b>	<b>17</b>
4.1. Les services gratuits	18
4.2. Les services payants	19
4.3. Les services "semi-fermés" ou troqués	19
<b>5. Autres recommandations</b>	<b>22</b>
5.1. Choix des langues	22
5.2. Approfondissement de modèles économiques	24
5.3. Intégration du service	25
<b>6. Feuille de route</b>	<b>26</b>
<b>7. Conclusions</b>	<b>29</b>

# 1. Introduction

La recherche devenant de plus en plus internationalisée et son ouverture un enjeu crucial pour la société, la quête d'un multilinguisme durable et équilibré dans la communication scientifique est plus que jamais un sujet d'intérêt pour de nombreuses parties prenantes : des universités et institutions de recherche aux éditeurs académiques et aux décideurs politiques.

Les études réalisées dans le cadre du projet Traductions et science ouverte<sup>2</sup> ont montré que le paysage des pratiques et des besoins en matière de traduction scientifique est trop complexe pour mettre en œuvre une approche unique, basée sur la technologie seule, et qui prenne en compte toutes les exigences qualitatives, organisationnelles, techniques et financières.

À partir des sujets préalablement pointés et à l'origine de cette étude, on identifie les enjeux suivants :

- produire des traductions qualitatives de textes hautement spécialisés avec des ressources humaines et financières souvent limitées ;
- améliorer l'attractivité vis-à-vis d'experts capables de traduire ces textes hautement spécialisés ;
- optimiser les processus de traduction et mettre à disposition des ressources linguistiques spécialisées et fiables ;
- mieux définir les cadres juridiques et techniques de la traduction scientifique et des activités connexes.

Dans ce contexte, notre étude propose des hypothèses techniques, organisationnelles et financières pour développer et déployer un service de traduction outillée et collaborative destiné à la communication scientifique .

---

<sup>2</sup> [Projet Traductions et science ouverte](#), carnet hébergé sur Hypotheses.org [consulté le 01.04.2024]

## 2. Recommandations d'usage

De nombreuses pistes d'usages du futur service ont été explorées dans les études précédentes. Repartant des travaux effectués dans l'étude 1<sup>3</sup>, des usages ont été circonscrits pour concilier réponse à des besoins, exigences éthiques et pérennité économique. Par ailleurs, par rapport à l'étude 1, nous nous sommes davantage concentrés sur les fonctionnalités et les composantes technologiques, notamment le moteur de traduction automatique, ainsi que les attentes et les contraintes associées.

Ainsi, trois typologies d'utilisateurs du service ont été proposées :

- les chercheurs (nous intégrons les étudiants dans cette catégorie),
- les traducteurs de contenus scientifiques,
- les éditeurs (nous intégrons également les diffuseurs de contenus scientifiques dans cette dernière catégorie).

Les échanges avec des représentants de ces typologies d'utilisateurs ont fait émerger la vision commune d'un service de traduction (sa composante automatisée ainsi que les fonctionnalités connexes) avec les caractéristiques suivantes :

- **Qualité** - les fonctionnalités proposées doivent être pertinentes pour un usage professionnel ;
- **Traçabilité** - le service doit proposer des résultats fiables et traçables ;
- **Conception centrée sur l'humain**<sup>4</sup> - le service doit répondre aux besoins de ses utilisateurs et valoriser ses contributeurs ;
- **Fonctionnement non intrusif** - le service doit prévoir des consentements libres, spécifiques, éclairés et univoques.

Des freins importants au déploiement du service, parfois considérés comme étant rédhibitoires, ont par ailleurs été identifiés et nommés par les différentes parties prenantes. Nous proposons une hypothèse de service qui prend en considération ces problématiques.

---

<sup>3</sup> X. Auffret, C. Lapassat, P. Lacour, Use case study for a technology-aided, collaborative translation service in scholarly communication (2023). <https://doi.org/10.5281/zenodo.10972812>

<sup>4</sup> Ce point fait référence au concept de *Human centric* par opposition à celui de *Human in the loop*, ce dernier consistant à intégrer une supervision humaine pendant l'entraînement d'un modèle de machine learning. Ici, il s'agit d'inverser ce paradigme : ce n'est pas l'humain qui enrichit l'IA, mais l'humain qui s'enrichit d'une intervention de l'IA.

Ces enjeux, dont la prise en compte est fondamentale pour la création d'une "communauté" d'utilisateurs et de contributeurs, sont énumérés ci-après.

## 2.1. Consentement et transparence

Par nature inhérente à la technologie d'intelligence artificielle (IA), les utilisateurs nourrissent le service avec leur propriété intellectuelle (PI). Le traitement qui est fait de leurs données doit être partagé de façon transparente et consentie.

S'inscrivant dans la logique d'un consentement éclairé, le devoir de transparence est à comprendre, non pas au sens réduit comme la simple mise à disposition de cette information, mais comme un devoir d'intelligibilité et d'assimilation de cette information.

## 2.2. Rétribution

Les contributeurs apportent de la valeur au service et l'améliorent. Leurs compétences et leur travail sont à valoriser. Le modèle économique du service doit prendre en considération une logique de contribution-rétribution qui puisse être suffisamment engageante et respectueuse du travail des personnes.

Cette approche peut être considérée comme un atout pour gagner l'adhésion de professionnels de la traduction.

## 2.3. Qualité de la base de données : biais et discriminations

La constitution de la base de données peut générer des biais et engendrer des discriminations dans des inférences proposées par le modèle d'IA, mais également de mauvaises interprétations.

La pertinence et la fiabilité des traductions automatiques doivent permettre une appropriation en confiance.

## 2.4. Déconstruction du mythe “IA de bout-en-bout”

Les contributions humaines ne doivent pas être invisibilisées par la façon dont le service fonctionne et est présenté. Le mythe d’une IA autonome et omnipotente porte atteinte à tous les contributeurs du service.

## 3. Fonctionnalités par usage

L’hypothèse formulée suite aux ateliers réalisés dans le cadre de l’étude est celle du développement d’un service **ouvert et collaboratif d’aide à la traduction scientifique** s’appuyant sur trois types d’usages : gratuits, payants et basés sur des contributions (utilisateurs-contributeurs).

### 3.1. Usages gratuits

L’usage gratuit pourrait donner accès à deux fonctionnalités :

- **Outil de traduction automatique** : possibilité de traduire en différentes langues et dans différents domaines scientifiques, grâce à des modèles entraînés sur des corpus spécialisés – par exemple, dans les trois macro-domaines distingués par le Conseil Européen de la Recherche : sciences humaines et sociales, sciences médicales, sciences physiques et de l’ingénieur. Le nombre de langues supportées pourrait être incrémenté au fil du temps (cf. [5.1 Choix des langues](#) dans un souci d’optimisation des ressources financières).
- **Annuaire** : possibilité de consulter l’annuaire des contributeurs du service. Cet annuaire permettrait de prendre connaissance des personnes et de leurs compétences ainsi que de rentrer en contact avec elles.

### 3.2. Usages payants

L’utilisateur payant pourrait avoir accès aux fonctionnalités incluses dans l’usage gratuit, auxquelles s’ajouteraient deux fonctionnalités soumises à abonnement :

- **Accès à des ressources linguistiques tracées et validées par des experts** : possibilité de consulter des mémoires de traduction et des glossaires scientifiques de qualité.

- **Accès au forum** : possibilité d'échanger avec la communauté de contributeurs et d'utilisateurs payants. Cela permettrait notamment d'initier des dynamiques de coopération sur des travaux de traduction.

### 3.3. Usages basés sur des contributions

Enfin, l'utilisateur contributeur pourrait avoir accès aux fonctionnalités des usages gratuits et payants ainsi que deux fonctionnalités supplémentaires. Notamment, il y accéderait gratuitement en échange de 2 types de contributions :

- **Contribution en ressources linguistiques** : l'utilisateur pourrait céder un volume régulier de mémoires de traduction et/ou glossaires, en contribuant ainsi à enrichir le moteur d'IA et les ressources linguistiques du service.
- **Contribution d'examen par les pairs** : l'utilisateur pourrait effectuer la révision d'un volume régulier de contributions en ressources linguistiques cédées par d'autres utilisateurs contributeurs, afin de garantir la qualité des contributions au service.

Le volume et la régularité des contributions en ressources linguistiques pourraient être encadrés par un nombre de caractères minimum et maximum par mois, afin d'éviter l'influence prépondérante d'acteurs majeurs.

De même, le volume et la régularité des contributions en examen par les pairs pourraient être encadrés avec un nombre de caractères à examiner minimum et maximum par mois, afin d'assurer une diversité de points de vue équilibrée.

En retour, les utilisateurs contributeurs pourraient être rétribués grâce aux deux fonctionnalités suivantes :

- **Traçabilité des contributions en ressources linguistiques** : chacune des contributions de l'utilisateur contributeur pourrait être référencée à son nom et horodatée. Son activité pourrait être suivie en consultant l'historique de ses contributions. Ces dispositifs permettraient de donner de la visibilité à l'activité du contributeur et d'offrir aux autres utilisateurs des ressources linguistiques plus transparentes et fiables.
- **Rétribution des contributions en ressources linguistiques** : les contributions en ressources linguistiques de l'utilisateur contributeur alimenteraient une base de données

utilisée pour l'entraînement du moteur de traduction automatique. L'utilisateur pourrait être rétribué à hauteur de sa contribution à cette base de données (cf. [2.2. Rétribution](#)).

Ces fonctionnalités s'appuieraient sur deux technologies à développer :

- **Le moteur de traduction automatique** fondé sur une technologie d'IA;
- **la plateforme web collaborative** qui permet à plusieurs utilisateurs d'interagir.

**Nota Bene** : Le scénario décrit dans ce rapport se concentre sur le développement et le déploiement de la technologie de traduction automatique. L'étude du développement et du déploiement d'une plateforme web collaborative, prenant en considération d'autres usages et enjeux, n'est pas considérée ici.

A titre d'exemple les principales fonctionnalités sont illustrées dans les [figures 1, 2, 3](#) ci-dessous.

**OPERAS** open scholarly communication in the european research area for social sciences and humanities

**Mes objectifs** Juin 2024 ▾

Ressources linguistiques 4

Examens par les pairs 2

**Contribuer** **Examiner**

**Mon profil** ⋮

Jeanne MARTIN  
jeanne.martin@operas.org

Suédois > Français

Coréen > Français

---

**Mes contributions** Tout afficher Juin 2024 ▾

- 📅 Aujourd'hui - 12:44 Suédois > Français  
Mémoire [Analyse chronologique des faits marquand en Normandie durant l'année 1978](#)
- ☰ 12.12.24 - 12:44 Suédois > Français  
Glossaire [Paradoxe](#)
- ☰ 12.12.24 - 12:44 Coréen > Français  
Glossaire [Paradoxe](#)
- 📅 Aujourd'hui - 12:44 Suédois > Français  
Mémoire [Randonnée en Jordanie](#)
- 📅 Aujourd'hui - 12:44 Suédois > Français  
Mémoire [Rencontres dans la Vallée verdoyante du Tigre](#)
- ☰ 12.12.24 - 12:44 Coréen > Français  
Glossaire [Ambiance nocture](#)
- ☰ 12.12.24 - 12:44 Coréen > Français  
Glossaire [Jours nouveaux](#)
- ☰ 12.12.24 - 12:44 Coréen > Français  
Glossaire [Quarré de l'hypothénuse](#)
- ☰ 12.12.24 - 12:44 Suédois > Français  
Glossaire [Trottoire d'en face](#)

**Fig1** : illustration de l'interface utilisateur - **Page profil** : un utilisateur contributeur peut suivre sur son tableau de bord les contributions qu'il doit effectuer en échange d'un usage gratuit du service. Ses contributions sont tracées.

Mémoire - Dernière contribution : aujourd'hui - 12:44

Mémoire pépîte ☆

Signaler un problème 🚩

## Randonnée en Jordanie

Français	Coréen <span>Valider</span>	Suédois
<p>La Jordanie est un pays du Proche-Orient, entouré par la Syrie au nord, l'Irak au nord-est, l'Arabie saoudite à l'est et au sud, et enfin Israël et la Cisjordanie à l'ouest. Toutes ces frontières représentent 1 619 km. La Jordanie possède également 26 km de littoral le long du golfe d'Aqaba au nord-est de la mer Rouge. Les principales villes du pays sont Amman (3 216 369 habitants en 2010), Irbid et Zarka, toutes trois situées dans le Nord-Ouest du pays.</p> <p>La Jordanie est composée principalement d'un plateau désertique aride à l'Est, et d'une région montagneuse à l'Ouest. La vallée du Grand Rift et le Jourdain séparent la Jordanie d'Israël. Le point culminant du pays est le jabal Umm ad Dami (1 854 mètres), tandis que la mer Morte en est le point le plus bas.</p>	<p>요르단은 중동의 국가로 북쪽으로는 시리아, 북동쪽으로는 이라크, 동쪽과 남쪽으로는 사우디아라비아, 서쪽으로는 이스라엘과 서안 지구와 국경을 접하고 있습니다. 이 국경의 총 길이는 1,619km입니다. 요르단은 또한 홍해 북동쪽의 아카바 만을 따라 26킬로미터의 해안선을 가지고 있습니다. 요르단의 주요 도시는 암만 (2010년 인구 3,216,369명), 이르비드, 자르카이머, 이 세 도시는 모두 북서쪽에 위치하고 있습니다.</p> <p>요르단은 주로 동쪽의 건조한 사막 고원과 서쪽의 산악 지역으로 구성되어 있습니다. 그레이트 리프트 계곡과 요르단 강은 요르단과 이스라엘을 분리합니다. 요르단에서 가장 높은 지점은 자발 움 아드 다미(1,854미터)이고 가장 낮은 지점은 사해입니다</p>	<p>Jordanien är ett land i Mellanöstern som gränsar till Syrien i norr, Irak i nordost, Saudiarabien i öster och söder samt Israel och Västbanken i väster. Dessa gränser är totalt 1 619 km långa. Jordanien har också en 26 km lång kustlinje längs Akababukten nordost om Röda havet. Landets största städer är Amman (befolkning 3 216 369 år 2010), Irbid och Zarka, alla tre belägna i nordvästra delen av landet.</p> <p>Jordanien består huvudsakligen av en torr ökenplatå i öster och en bergsregion i väster. Great Rift Valley och Jordanfloden skiljer Jordanien från Israel. Den högsta punkten i landet är jabeln Umm ad Dami (1 854 meter), medan Döda havet är den lägsta punkten.</p>

**Fig2** : illustration de l'interface utilisateur - **Page examen de mémoire** : un utilisateur contributeur peut examiner une mémoire de traduction en l'éditant, en signalant un problème technique ou éthique à l'aide du bouton "signaler un problème", ou en soulignant la qualité du travail de traduction à l'aide du bouton "mémoire pépîte".

## Traduction automatique

Domaine ▾

Français	> Coréen ▾	> Suédois ▾	+
Écrivez ou collez votre texte ici			

**Fig3** : illustration de l'interface utilisateur - **Page traducteur**: un utilisateur peut effectuer une traduction multilingue en comparant la traduction proposée dans chaque colonne. Il a la possibilité d'ajouter des colonnes pour augmenter le nombre de langues de traduction simultanées. La traduction automatique est spécialisée dans un domaine que l'utilisateur choisit dans un menu déroulant : sciences humaines et sociales / sciences médicales / sciences physiques et de l'ingénieur.

Les fonctionnalités peuvent être résumé dans le [tableau #1](#) suivant :

	<b>Fonctionnalité</b>	<b>Type d'usage</b>	<b>Technologie</b>
#1	Outil de traduction automatique	Gratuit Payant Contributeur	Traduction automatique
#2	Annuaire	Gratuit Payant Contributeur	Plateforme collaborative
#3	Accès aux ressources linguistiques tracées et validées par des experts	Payant Contributeur	Plateforme collaborative
#4	Accès au forum	Payant Contributeur	Plateforme collaborative
#5	Traçabilité des contributions en ressources linguistiques	Contributeur	Plateforme collaborative
#6	Rétribution des contributions en ressources linguistiques	Contributeur	Plateforme collaborative

Tableau 1: classification des fonctionnalités par type de technologie et d'usage

En repartant des résultats de l'étude 3<sup>5</sup>, il est important de noter que réaliser un service qui propose ces fonctionnalités avec un niveau de performance satisfaisant par rapport à l'état de l'art demande de se fonder sur des solutions techniques existantes et de constituer une base de données massive. Cette contrainte court terme ne doit pas occulter l'objectif long terme de s'émanciper de solutions commerciales pour des raisons de souveraineté économique et politique.

<sup>5</sup> T. Vanallemeersch, S. Szoc, K. Migdisi, L. Meeus, L. Macken, A. Tezcan, 2023, Machine translation evaluation in the context of scholarly communication (2023). <https://doi.org/10.5281/zenodo.10972872>

Pour concilier ces différents enjeux, le scénario présente une feuille de route en 4 étapes clés, étalée sur 3 ans (cf [6. Feuille de route](#)), dont l'objectif serait de pouvoir démarrer rapidement en s'appuyant, en partie, sur une solution commerciale externe puis progressivement migrer vers une solution ouverte propre au service (nommée par la suite **Moteur Open Science**).

## 4. Modèles économiques et traduction juridique

L'objet de cette section consiste à proposer un modèle économique, conforté par un modèle juridique, qui permette la pérennité d'un service de traduction outillé ouvert, c'est à dire :

1. de lui assurer un revenu minimum ;
2. de consolider la contribution d'une communauté afin d'améliorer les services de manière continue et pérenne ;
3. d'appréhender une approche éthique où l'on peut considérer la place de l'humain à côté de l'IA.

Dans ce contexte, nous avons proposé des services gratuits ou libres, des services qui pourraient être échangés ou "troqués" entre les parties prenantes, et enfin des services payants notamment à destination d'institutions et organisations tierces.

Cette démarche est consolidée par une approche juridique fondée sur les communs, laquelle fait une distinction entre :

- (i) des contenus "ouverts", généralement sous licences libres, comme les licences Créative Commons,
- (ii) des contenus "semi-fermés" sous licence à réciprocité qui permettent un échange réciproque de ressources, biens ou services, entre les membres d'une même communauté de personnes, cet échange pouvant être ou non monétaire, et
- (iii) des contenus "fermés", faisant l'objet d'une exclusivité, opposable à tous, justifiant une autorisation préalable par les auteurs des contenus au bénéfice de tiers, en contrepartie d'une utilisation, gratuite ou payante, plus ou moins limitée dans l'objet, la durée, le territoire et les finalités de ladite utilisation.

Des autorisations croisées, sous la forme de licences mutuelles<sup>6</sup>, pourraient être envisagées entre les parties prenantes, détentrices de droits d'auteur sur les contenus en question.

---

<sup>6</sup> Licence mutuelle : à l'instar des dons et des contre dons facilitant l'engagement de la communauté.

## 4.1. Les services gratuits

Les services gratuits ou de libre accès sont ceux qui répondent à une grande majorité des usages de la traduction automatique de nos jours.

Effectivement, un très grand nombre d'utilisateurs ont recours à des interfaces de programmation (API) et des sites en ligne pour traduire des phrases dans une ou plusieurs langues, tant pour des fins professionnelles que personnelles. La version gratuite de ces sites, faciles d'accès et d'utilisation, ne demande pas de contrepartie aux utilisateurs.

**Nota Bene:** Les traductions générées en ligne par le biais d'un service gratuit ne sont pas explicitement protégées sur le plan de la propriété intellectuelle. Il en découle une suspicion de *scrapping*. Les CGU de ces sites gratuits ou API de traduction ne mentionnent pas ce type de pratique. C'est le cas du site DeepL qui ne présente les CGU que pour la version payante : <https://www.deepl.com/fr/pro-license>

Les sites de traduction en ligne gratuits sont des démonstrateurs de capacité de traduction automatique et produits d'appel pour les versions payantes.

Les points faibles de ce type d'usage sont :

- des traductions imparfaites
- le non-traitement des thématiques de niches, telles que celles que l'on retrouve dans les sciences.

Dans le cadre du futur service, des CGU seraient à rédiger afin d'encadrer l'accès à la fonctionnalité de traduction automatique pour tout type d'utilisateur gratuit, c'est-à-dire pour toute personne se connectant, sans recours à une identification quelconque.

Les CGU pourraient couvrir :

- les modalités d'ouverture,
- les limites découlant de l'utilisation à titre gratuit,
- les responsabilités pour l'utilisateur, etc.

**Concernant le règlement général sur la protection des données (RGPD)**, celui-ci implique de l'appliquer à tout traitement des données personnelles effectué dans le cadre du service en ligne. En matière de documentation juridique, il conviendra notamment de s'assurer de publier à destination des utilisateurs du dispositif une charte de confidentialité à laquelle sera associée des mentions **d'opt-in**<sup>7</sup> pour s'assurer du consentement des utilisateurs du dispositif. À titre d'exemple, les emails ou pseudos des utilisateurs ne doivent pas être collectés ou bien conservés sans le consentement de ces derniers.

Un service de **newsletter** pourrait être utilisé afin de consolider la communauté et de l'informer de nouvelles fonctionnalités et versions du service, ainsi que des spécificités de fonctionnement des communications scientifiques en science ouverte et d'opérations de communication destinées aux potentiels financeurs.

En termes de fonctionnalités, dans le cadre d'un service gratuit, nous suggérons une limitation technique du nombre de traductions possibles (quantité de texte limité) ainsi qu'un accès réduit à l'ensemble des fonctionnalités recensées (cf [3.2. Usages payants](#)).

## 4.2. Les services payants

Les services payants doivent permettre, a *minima*, la soutenabilité des coûts récurrents du service en production. Ils s'inscrivent dans une logique d'amélioration permanente du dispositif de traduction outillée.

Ces prestations payantes, dès lors que les parties prenantes y accèdent via une plateforme numérique, pourraient être encadrées par un instrument juridique constitué de CGV, supplétives à un autre instrument juridique qui seraient des CGU, elles-mêmes gratuites, pour appréhender l'utilisation des services de libre accès (cf [3.1. Usages gratuits](#)). Ainsi, les parties prenantes pourraient, en s'inscrivant sur la plateforme, adhérer, par défaut, aux CGU qui fourniraient un service minimal et standard et, par choix, aux CGV, au cas où une utilisation réservée et payante les intéresseraient, selon la méthode classique dite du "premium".

## 4.3. Les services "semi-fermés" ou troqués

Dans une communauté restreinte ou "resserrée", le troc peut s'organiser.

---

<sup>7</sup> L'opt-in, c'est obtenir le consentement du destinataire de la publicité : s'il n'a pas dit "oui", c'est "non".

Pour mettre en œuvre ces échanges, les parties prenantes (éditeurs, chercheurs et traducteurs) pourraient souscrire à des licences à réciprocité, dont l'adhésion pourrait également être prévue via la plateforme, lesquelles permettraient un échange de biens et/ou services entre personnes intéressées.

Cette souscription serait également volontaire et facultative, comme les conditions particulières de vente, laissant ainsi le choix aux parties prenantes de moduler leur engagement contractuel.

Les réciprocités pourraient être monétaires ou non. Elles pourraient être complétées par un système de valorisation des contributions et rétributions, soit en euros soit dans une monnaie n'ayant aucun cours légal (système de points).

Entre professionnels, cet échange monétaire ou non devrait faire l'objet d'une valorisation fiscale et nécessiterait l'édition de factures croisées qui se compenseraient ou pas, selon la méthode dite du "bartering".<sup>8</sup> En 2013, le ministère de l'Economie et des Finances a publié un guide des plateformes d'échanges de biens et de services entre entreprises qui peut inspirer de telles pratiques.

**L'une des hypothèses de contribution évoquée lors des ateliers est celle où les traducteurs et les chercheurs apportent des mémoires et glossaires spécifiques pour permettre d'affiner les outils, c'est-à-dire les entraîner sur des données spécialisées afin d'améliorer la traduction.**

Cela implique une tarification/valorisation de la contribution au cas par cas. Nous suggérons aujourd'hui une estimation quantitative des contributions dans la base de ressources linguistiques afin de déterminer l'entité de la rétribution.

L'un des autres intérêts explicités par les parties prenantes en qualité de potentiels contributeurs consiste à **être identifiés** pour leur contribution et **bénéficier d'un accès premium**. Cette démarche nécessite un suivi des identités des contributeurs. (cf [3.3. Usages contributeurs](#)).

Au nom d'un intérêt commun pour gagner du temps et afin de baisser les coûts, les différentes parties prenantes seraient prêtes à échanger différents services et données.

---

<sup>8</sup> Bartering : Désigne l'échange de marchandises entre entreprises. Le barter joue un rôle intermédiaire entre les entreprises qui peuvent ainsi échanger un produit contre un autre produit et un service contre un autre service.

Lors des ateliers réalisés dans le cadre de l'étude, les intérêts mentionnés par les chercheurs sont les suivants :

- **L'accès à un glossaire spécialisé** (par champ) tracé et de bonne qualité, nourri par d'autres chercheurs et par des traducteurs, et susceptible d'être enrichi par les utilisateurs, par conséquent non biaisé (cf. [tableau #1](#), fonctionnalité 4)
- **Annuaire** (cf. [tableau #1](#), fonctionnalité 2)

Lors des ateliers réalisés dans le cadre de l'étude, les intérêts mentionnés par les traducteurs sont les suivants :

- L'accès premium à un glossaire spécialisé et tracé (**nourri par les chercheurs et traducteurs de la communauté des contributeurs**), non biaisé (cf. [tableau #1](#), fonctionnalité 4)
- L'accès à une base de mémoires de traduction fournie par les utilisateurs (la communauté des **traducteurs**) (cf. [tableau #1](#), fonctionnalité 4)

**Les éditeurs** ont en leur possession des textes bilingues qui pourraient servir à créer des mémoires de traduction. Selon le type de contrat mis en place avec les chercheurs et les traducteurs en matière de droits d'auteur, ils pourraient céder ces textes contre rétribution dans le système de comptabilité de la plateforme.

La motivation principale dans le partage de ces données serait l'accès aux glossaires spécifiques créés par la communauté restreinte des contributeurs (ressources linguistiques de niche).

La contrepartie pour les éditeurs serait donc de payer de manière identique par rapport aux produits du marché tout en obtenant des fonctionnalités nouvelles (collaboratives, annuaire, etc.)

## 5. Autres recommandations

### 5.1 Choix des langues

Le précédent rapport<sup>9</sup> préconisait de travailler vers une dualité d'objectifs, notamment :

- Promouvoir la dissémination de production scientifique initialement en français vers d'autres langues sur tous les continents
- Casser la barrière de la langue pour les citoyens, organisations et entreprises francophones qui souhaitent avoir accès à des résultats de recherche internationale.

Ce choix des langues pour le dispositif de traduction est stratégique. D'une part en termes d'impact et de pertinence du dispositif, d'autre part parce que le nombre de langues à implémenter pour la traduction a un impact non négligeable sur le coût du développement du service et sur les partenariats qu'il implique pour sa mise en œuvre.

Afin de travailler avec un objectif plus précis du nombre de langues potentielles à traduire, nous avons essayé un système de score. Pour ne pas cloisonner le périmètre de l'outil uniquement au regard du nombre de publications scientifiques par pays et de langues d'origine des publications scientifiques, nous avons cherché à intégrer des critères géopolitiques et macroéconomiques pour la pertinence du choix de ces langues.

a) Langues les plus parlées dans le monde<sup>10</sup>: 5 points

Anglais, mandarin, hindi, espagnol, français, arabe, bengali, russe, portugais, ourdou

b) Pays les plus riches et les plus influents économiquement<sup>11</sup> dans l'ordre mondial actuel: 5 points pour le G7, 4 points pour la BRICS, 3 points pour les autres langues de l'UE :

- Le G7 s'exprime dans 5 langues différentes: anglais, français, allemand (5 points), italien (5 points) et le japonais (5 points)
- Les autres langues parlées par les pays de la BRICS (4 points): portugais (4 points), russe (4), hindi (4), bengali(4), ourdou (4), le mandarin (4) et l'indonésien (4)
- Autres langues de pays économiquement influents dans l'union européenne (3 points) : néerlandais, polonais, suédois.

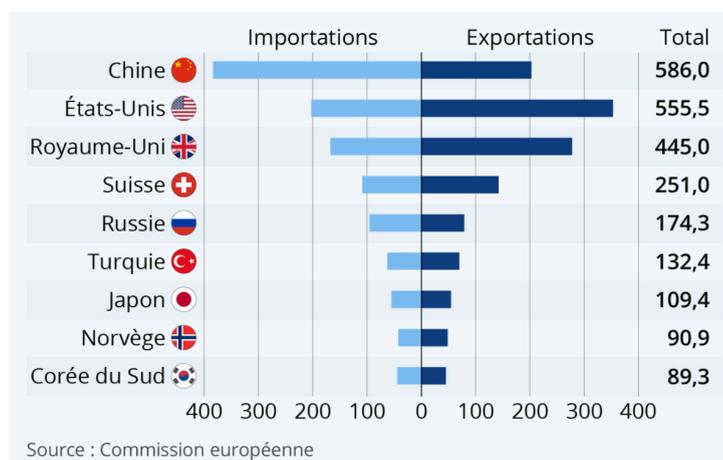
---

<sup>9</sup> <https://www.ouvrirlascience.fr/report-by-the-translations-and-open-science-working-group/>

<sup>10</sup> Source: [https://fr.wikipedia.org/wiki/Liste\\_de\\_langues\\_par\\_nombre\\_total\\_de\\_locuteurs](https://fr.wikipedia.org/wiki/Liste_de_langues_par_nombre_total_de_locuteurs)

<sup>11</sup> Source : G20 [https://fr.wikipedia.org/wiki/Groupe\\_des\\_vingt#Pays\\_et\\_organisations\\_membres](https://fr.wikipedia.org/wiki/Groupe_des_vingt#Pays_et_organisations_membres)

- c) Langues parlées par les principaux partenaires économiques de l'Union européenne en 2023 <sup>12</sup>: de 5 à 1 point: mandarin , anglais, français (5 points), russe (4 points), turque (3 points), japonais (2 points), norvégien (1 points), coréen (1 point).



**Fig4** : volume des échanges commerciaux avec les partenaires majeurs de l'Union européenne en 2020, en milliards d'euros

- d) Langues des premiers pays de publications scientifiques<sup>13</sup> :

- 5 points : anglais, mandarin, allemand, hindi, japonais, français, italien,
- 4 points : russe, espagnol, coréen, portugais, néerlandais, iranien,
- 3 points : polonais, turque, suédois,

Notre classement pour le choix des langues à implémenter et résultant de critères centrés sur les partenaires économiques et scientifiques de l'union européenne, donne la liste suivante :

1. Anglais : 20 points; (1,452 milliard de locuteurs 1er et 2e langue)
2. Français : 20 points; (274,1 millions de locuteurs)
3. Mandarin : 19 points; (1,118 milliard de locuteurs)
4. Russe : 17 points;
5. Hindi : 14 points;
6. Allemand : 10 points;
7. Italien : 10 points;

<sup>12</sup> Source:

<https://fr.statista.com/infographie/14742/principaux-partenaires-economiques-ue-union-europeenne/>

<sup>13</sup>Source :

<https://atlasocio.com/classements/education/publications/classement-etats-par-nombre-publications-scientifiques-monde.php>

8. Espagnol : 9 points;
9. Portugais : 9 points;
10. Arabe : 5 points;

Il résulte que si nous implémentons les 10 langues de ce classement, nous maximisons la portée du service en termes de citoyens et scientifiques à travers le monde, de poids de PIB et de potentiels de publications scientifiques couvertes – tout en minimisant le coût liés aux efforts de traduction multilinguistique du dispositif technologique.

L'avantage de ce classement est que le dispositif devrait être plus facilement finançable dans une logique de partenariats publics-privés. Le défaut de ce classement multicritère est qu'il exclut, si on s'arrête à un liste de 10 langues, un certain nombre de pays de la zone euro (notamment les pays du nord de l'Europe) et le Japon (6e au classement en termes de publications scientifiques).

Si l'on devait commencer par les 3 langues préconisées dans le rapport précédent (anglais, français, espagnol), les 2 autres langues européennes suivantes les plus impactantes sur le plan scientifiques et diplomatiques à implémenter seraient l'allemand et l'italien.

A titre indicatif, pour notre feuille de route du déploiement du dispositif, nous sommes donc partis sur un développement du service avec un démonstrateur en 3 langues puis un premier service avec 5 langues puis 10 langues la 3e année.

La réalité diplomatique et les opportunités de financement guideront le choix des langues mais il faut garder en tête que celui-ci a un coût.

## 5.2. Approfondissement de modèles économiques

Un certain nombre d'hypothèses ont été émises par les personnes que nous avons croisées lors de l'étude. Afin de valider ces hypothèses, un questionnaire serait à faire circuler auprès d'un plus grand nombre de personnes dans les différents secteurs de métiers concernés par le dispositif de traduction.

- Question ouverte à approfondir par ailleurs : Est-ce que l'éditeur peut renforcer la visibilité des traducteurs (marché du travail) et chercheurs (impact scientifique) membres de la communauté?

- Échange monétaire et non monétaire vont dépendre d'un ensemble de modalités : contrat de travail, type de contribution, contexte général des droits d'auteur des contributeurs, etc.

### 5.3. Intégration du service

Les usages imaginés pour le service nécessitent des interfaces utilisateur variées et spécifiques. Certains utilisateurs potentiels ont exprimé le besoin d'une intégration du service dans leur outil de traduction et de rédaction, d'autres de pouvoir bénéficier de ces fonctionnalités sur leur propre plateforme web.

Cela représente des coûts dédiés de développement et de maintenance pour chaque intégration envisagée. Le scénario de déploiement présenté dans ce rapport propose de développer une plateforme web comme unique interface et d'exposer une API pour permettre l'adoption dans d'autres applications informatiques.

## 6. Feuille de route

Le déploiement du dispositif est présenté sur une feuille de route en 4 étapes clés, décrites ci-dessous, et étalées sur 3 ans pour une montée en charge progressive.

- **M0 - Démonstrateur** : la première étape correspond au lancement du service sous forme d'un démonstrateur et dont la mise en service nécessite **6 mois de développement**. L'objectif de ce démonstrateur est de montrer les fonctionnalités clés du dispositif envisagé afin d'en évaluer la pertinence. Il se fonde sur l'utilisation d'une solution commerciale externe de traduction automatique. Durant les 6 mois de développement, sont effectués: la structuration de la collecte de données et de sa gouvernance, l'encadrement légal du traitement des données, l'étude d'architecture qui soutient le système global, le développement d'une interface utilisateur dédiée au démonstrateur et d'une API qui servira de base pour le développement du service..
- **M1 - Moteur Open Science V1 : La première version du service nécessite 9 mois de développement**. La version 1 du service de traduction possède les mêmes fonctionnalités que le démonstrateur mais permet de prendre en charge jusqu'à 10000 utilisateurs. Cela permet d'initier une dynamique collaborative ainsi que de recueillir retours et contributions d'utilisateurs. Durant cette phase, la collecte de données, permise par la structuration effectuée lors de la phase M0, débute sous différentes formes (scrapping, appel à contribution, acquisition, contributions d'utilisateurs). Le système d'information est sécurisé et outillé pour le déploiement et la maintenance.
- **M2 - Moteur Open Science V2 : La deuxième version du service nécessite 9 mois de développement**. Dans sa deuxième version, le service anticipe la montée en charge progressive du nombre d'utilisateurs quotidien vers une limite de 20 000 utilisateurs par jour, 10 millions de segments dans les mémoires de traduction et un apport de 500 000 contributions. Ce service fonctionne pour cinq langues. En parallèle, un modèle de traduction automatique Open Science est entraîné sur la base de données collectée grandissante, et évalué en comparaison à la solution commerciale externe.
- **M3 - Moteur Open Science V3 : la troisième version du service nécessite 6 mois supplémentaires de développement**. Dans cette version, la solution de traduction automatique externe est remplacée par le traducteur automatique Open Science, offrant un service autonome. La montée en charge du service supporte jusqu'à 50000 utilisateurs quotidien, 50 millions de mémoires de traduction, 1 millions de contributions,

523 Go de données textuelles. Une traduction en dix langues principales est proposée, permettant de couvrir les 5 continents et les pays les plus influents scientifiquement et économiquement.

Le développement de la plateforme collaborative, non décrit ici, se devra se faire de manière concomitante au développement de la technologie de traduction outillée.

## Feuille de route

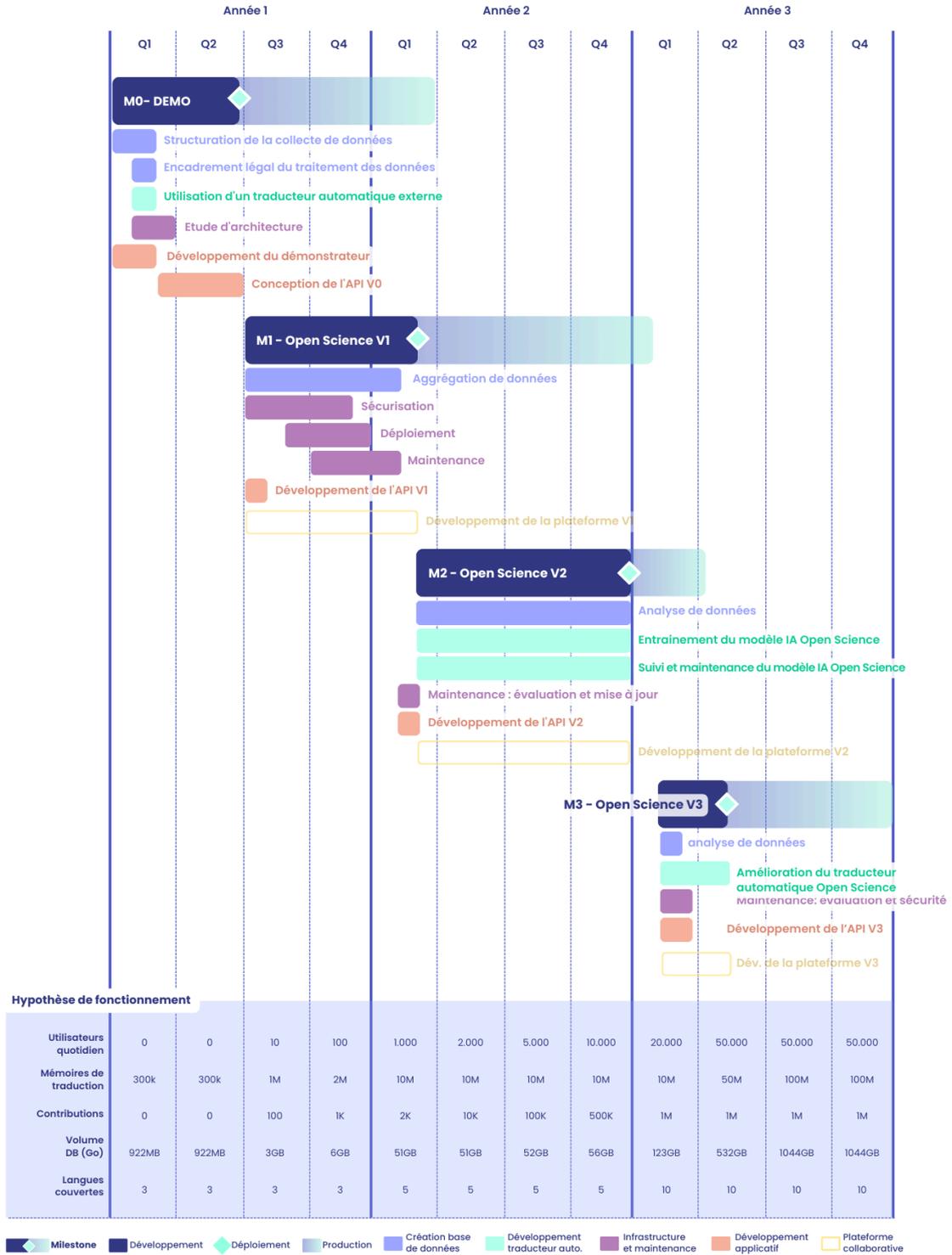


Fig5 : Feuille de route pour le déploiement du service de traduction Open Science

## 7. Conclusions

Le présent rapport a permis de faire émerger des fonctionnalités associées à des usages, leur temps de développement et une première proposition de fonctionnement économique avec sa traduction juridique.

Il est nécessaire de préciser qu'un certain nombre d'hypothèses fortes ont dû être émises afin d'arriver à l'estimation du déploiement d'un tel dispositif de traduction.

La **première hypothèse** forte est que le service de traduction seul ne suffit pas à la pérennisation économique du service du fait de la forte concurrence par la présence de services de traduction en ligne efficaces mais également des fortes évolutions des métiers. En découle la **nécessité d'une plateforme collaborative** qui donne une place importante aux efforts humains consentis pour la création d'un tel service.

La **deuxième hypothèse** forte est que le choix d'une **dizaine de langues** est un choix stratégique qui sécurise la maîtrise des coûts humains dans la construction du dispositif.

La **troisième hypothèse** est que les solutions d'IA fondées sur des grands modèles de langue (LLM) bien que écartées de cette étude, car popularisées alors que le projet touchait à sa fin, sont potentiellement la prochaine étape technologique en termes de traduction automatique. Les IA génératives apportent des capacités rédactionnelles et des niveaux d'interactivité remarquables mais aussi de nouveaux enjeux éthiques, juridiques et économiques.

La **quatrième hypothèse** est qu'un **modèle de type "prestation de service" serait difficilement soutenable d'un point de vue économique** pour le déploiement d'un dispositif de traduction outillé aussi ambitieux. Il est en revanche possible et souhaitable de s'inspirer d'un **système d'économie collaborative pour pérenniser le service**.

D'une manière générale, le contexte de forte progression de l'usage de l'IA dans la traduction des langues implique une évolution aussi forte du métier de traducteur.

En proposant un scénario où les traducteurs peuvent bénéficier directement de la plus-value de leur travail, nous entrevoyons une recommandation d'évolution du métier à la frontière entre la science et la traduction des langues.