

TACT

Plateforme de
transcription et
d'annotation
de corpus textuels

Plateforme TACT

MANUEL DU GESTIONNAIRE DE PROJET

Mise à jour en mars 2021



Ce guide a pour vocation d'accompagner toute personne qui dispose déjà d'un projet sur TACT ou qui cherche à en créer un.

Le sommaire est organisé en fonction des différents menus de la plateforme.

TACT [Les projets](#) [Actualités](#) [À propos](#) [Se connecter](#) [Créer un compte](#)

PLATEFORME DE TRANSCRIPTION ET D'ANNOTATION DE CORPUS TEXTUELS

Bienvenue sur TACT !

TACT est une plateforme de transcription et d'annotation collaborative qui accueille des projets de recherche et permet à chacun d'apporter sa contribution ou de solliciter la communauté de la plateforme.

Vous voulez participer à un projet ? Vous pouvez transcrire et annoter des textes originaux, relire et vérifier des transcriptions faites par d'autres internautes, consultez ce [manuel du contributeur](#) et allez [explorer les projets](#) !

Vous souhaitez déposer un projet sur TACT ? Vous deviendrez gestionnaire de votre projet TACT, vous pourrez en définir les modalités et gérer les contributions. Nous vous invitons à lire ce [manuel du gestionnaire](#) qui vous donnera toutes les informations nécessaires ainsi que la marche à suivre. Bonne lecture !

[Les projets](#) [Manuel du contributeur](#)

UGA CIP 3iM Hum-Num ELAN

[Haut de page](#)
[Conditions Générales d'Utilisation](#)

La plateforme TACT est le résultat d'un travail d'équipe (Brigitte Combe, Cécile Meynard, Thomas Lebarbé, Elisabeth Greslou, Arnaud Bey, Anne Garcia-Fernandez, Patrick Guillou, Sylvain Hatier) et nous tenons à remercier toutes celles et tous ceux qui y ont participé et en particulier Célia Marion qui est l'autrice de la première version de ce document.

Table des matières

1. Principe de la plateforme.....	4
2. Création d'un projet TACT.....	4
2.1. Créer un compte en tant que gestionnaire.....	4
2.2. Votre rôle.....	4
2.3. Éditer.....	5
2.3.1. Édition des propriétés de base :.....	5
2.3.2. Édition des propriétés avancées :.....	6
2.3.3. Édition des serveurs d'images.....	7
2.3.4. Édition des métadonnées.....	7
2.3.5. Autres opérations.....	8
2.4. Gestion des médias.....	8
2.4.1 Hébergement interne des médias.....	8
2.4.2 Hébergement externe des médias.....	8
Le principe.....	8
En pratique.....	9
2.5. Gestion des utilisateurs.....	9
3. Que la transcription commence !.....	10
3.1. Aider les contributeurs.....	10
3.1.1. Manuel du contributeur.....	10
3.1.2. Tutoriel.....	10
3.1.3. Instructions de transcription.....	11
3.1.4. Océrisation.....	11
3.2. Relecture.....	11
4. Et maintenant ?.....	12
4.1. Exports.....	12
4.2. Clôture du projet.....	13

1.Principe de la plateforme

TACT est une plateforme collaborative de Transcription et d'Annotation de Corpus Textuels.

La transcription collaborative se fait de deux manières :

- Par la transcription de pages numérisées, appelées médias ;
- Par la relecture des transcriptions préalablement produites.

Les utilisateurs s'inscrivent à un projet et transcrivent les pages disponibles selon vos instructions ; ils peuvent aussi relire des transcriptions déjà faites.

Les projets sur la plateforme accueillent deux types de Contributeurs :

- Transcripteurs : ils transcrivent des médias ;
- Relecteurs : ce sont des Transcripteurs qui peuvent également relire et valider des transcriptions.

2.Création d'un projet TACT

Ce chapitre vise à présenter les différentes étapes à suivre pour créer un projet sur la plateforme TACT.

2.1. Créer un compte en tant que gestionnaire

Pour créer un projet sur TACT, nous vous conseillons d'abord de lire l'intégralité de ce manuel. Ensuite, il vous faudra **créer votre compte utilisateur** sur la plateforme et remplir le formulaire de demande de création de projet disponible à l'adresse suivante :

<http://enquete.elan-numerique.fr/index.php/854998?lang=fr>

Vous pouvez bien sûr à tout moment **contacter les administrateurs** à l'adresse tact@univ-grenoble-alpes.fr pour poser vos questions, obtenir plus d'informations, etc. Ce sont les administrateurs qui créeront un squelette de projet dont vous serez nommé gestionnaire.

Une fois gestionnaire de projet, vous avez la main sur tous les aspects de votre projet TACT via plusieurs onglets à votre disposition.

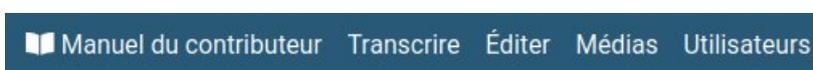


Figure 1 : Onglets disponibles pour les gestionnaires de projet

2.2. Votre rôle

En tant que gestionnaire, votre rôle est de :

- définir les tâches que vous confiez à vos contributeurs, ceux et celles qui transcriront et/ou valideront les transcriptions, notamment en mettant à leur disposition un manuel du contributeur spécifique à votre projet ;
- gérer les utilisateurs et animer votre projet sur TACT : par exemple, inviter des personnes à participer à votre projet et suivre régulièrement leur avancement ;

- surveiller l'activité de votre projet sur TACT : les inscriptions, les demandes de relectures, etc. Si vous avez peur de perdre le fil, vous pouvez consulter les événements du projet mais aussi vous abonner aux notifications d'activité du projet pour recevoir toutes ces informations par mail (depuis la page d'accueil de votre projet aussi appelée page



« Détails »).

- Figure 2 : Cadre visible par les gestionnaires de projet en bas à gauche de la page « Détails » du projet. À gauche, avant inscription aux notifications, à droite, après inscription.

2.3. Éditer

Diverses informations sont nécessaires pour renseigner votre projet et le documenter. Dans l'onglet « Éditer » du bandeau bleu de la plateforme, vous pouvez ajouter et modifier de nombreux éléments indispensables. Les paramètres modifiables sont répartis en différentes catégories.

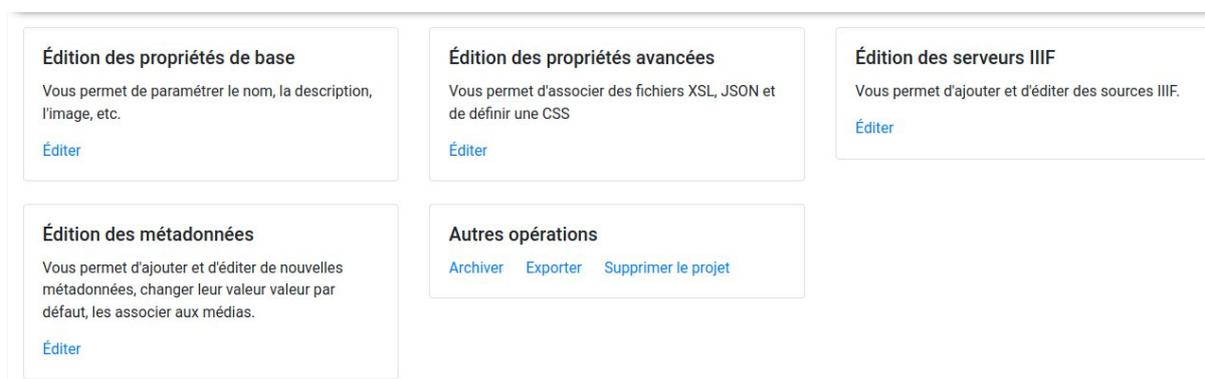


Figure3 : Catégories d'édition d'un projet TACT.

2.3.1. Édition des propriétés de base :

- **Nom du projet** : choisissez un titre accrocheur qui donne envie aux utilisateurs de participer à votre projet !

- **Description du projet** : assez brève, pour que les utilisateurs comprennent ce à quoi ils vont participer et en quoi leurs contributions seront utiles à votre projet de recherche. C'est ici que vous expliquez à vos contributeurs ce que vous attendez d'eux.

Attention, ceci n'est pas une description scientifique de votre projet !

- **Inscription libre** : si vous cochez la case, tout utilisateur pourra s'inscrire à votre projet et commencer à transcrire ; sinon, l'inscription sera modérée et vous devrez valider l'inscription d'un utilisateur à votre projet pour qu'il puisse accéder aux médias et participer à la transcription.
- **Phrase d'accroche** : le slogan de votre projet TACT
- **Instructions de transcription** : déposez ici un document rédigé par vos soins spécifiant concrètement ce que vous attendez de vos contributeurs (quoi transcrire, quels phénomènes annoter et comment). Pour plus de détails, voir la section 3.1.3.
- **Nombre de relectures nécessaires par fiche** : au bout d'un certain nombre de relectures que vous définissez ici (obligatoirement supérieur ou égal à 1), une transcription est considérée comme "terminée", elle n'est plus modifiable sur la plateforme et le gestionnaire peut la récupérer.

À noter que *plus les transcriptions sont ouvertes à un public large , plus le nombre de relectures nécessaires est élevé. En effet, plus les transcriptions sont revues par des personnes différentes et moins elles contiendront d'erreurs ou d'imprécisions. Cela signifie également que votre projet aura besoin d'un grand nombre de contributeurs pour avancer rapidement !*

- **Image du projet** : image sympathique en lien avec votre projet pour donner envie aux utilisateurs de participer. Elle est affichée en bandeau dans l'espace dédié à votre projet ainsi qu'en miniature sur la page d'accueil de TACT.

Quelques recommandations :

- veillez à utiliser une image dont vous disposez des droits ou libre de droits ;
- préférez un ratio de 4:1, sans quoi l'image sera rognée automatiquement à l'affichage.

2.3.2. Édition des propriétés avancées :

- **Définition de la CSS inline pour le projet** : vous pouvez ici définir des règles CSS afin d'améliorer la mise en forme de la transcription en mode « Prévisualisation ».
- **Fichier XSL appliqué à l'export des transcriptions** : vous pouvez déposer ici un fichier XSL utilisant la syntaxe XSLT 1.0 ; La transformation sera alors appliquée à l'export des fichiers de transcription. Un exemple de fichier .xsl est disponible sur la plateforme TACT : <https://tact.demarre-shs.fr/data/xsl-exemple.xsl>.
- **Définition de la langue pour la fonctionnalité OCR Tesseract** : vous pouvez ici définir la langue des médias pour permettre la génération d'une transcription automatique (via Tesseract) dans la vue *Transcrire*
- **Schéma de transcription au format JSON** : TACT vous permet d'inclure un schéma de transcription (c'est-à-dire un ensemble d'éléments et d'attributs prédéfinis) en fonction de vos objectifs. Si vous utilisez un schéma, il doit respecter le format JSON interne à TACT. Si vous n'utilisez pas de schéma de transcription, alors les contributeurs ne pourront que transcrire du texte seul.

Bon à savoir : Dans le cas, où vous disposez d'un schéma TEI (au format RNG, obtenu avec l'outil ROMA- par exemple [celui du projet Réticence](#)), vous pouvez utiliser le script Python mis à votre disposition pour générer un fichier JSON : <https://gitlab.com/litt-arts-num/tei2json>.

Si vous choisissez d'inclure un schéma de transcription, veillez à le restreindre au maximum pour que seuls **les éléments correspondant aux phénomènes que vous souhaitez annoter soient disponibles** pour vos *contributeurs*.

S'il est préférable de remplir tous ces paramètres dès la genèse du projet TACT, ils restent bien évidemment modifiables à tout moment.

2.3.3. Édition des serveurs d'images

Plutôt que de téléverser sur la plateforme TACT les médias à transcrire, il est possible de faire référence à des fichiers hébergés sur un ou plusieurs serveurs d'images externes.

À noter : le recours à un serveur utilisant la technologie IIIF vous permettra de spécifier la qualité d'image pour les images réduites (celles visibles sur les pages listant les médias de votre projet) et pour les images du mode transcription (pour lesquelles une qualité supérieure est souvent nécessaire).

Si vous décidez d'héberger les données sur TACT, vous n'avez pas besoin de renseigner ces options.

Si vos images sont hébergées de façon externe, vous pouvez définir autant de serveur que nécessaire en choisissant d'ajouter un nouveau serveur IIIF. Les options sont les suivantes :

- **Nom du serveur** : Ce nom est indiqué à côté du noms des médias dans la vue « Médias »
- **URL du serveur** : indiquer ici l'URL de base du serveur.
- **Suffixe large** : pour les images à transcrire : à n'utiliser que si vos images sont hébergées sur un serveur IIIF. Spécifiez ici le suffixe de l'URL (en commençant par « / ») à utiliser pour obtenir des images de qualité optimale pour les transcriptions. La syntaxe précise dépend du serveur que vous utilisez mais vous aurez par exemple : `/full/full/0/default.jpg` (on demande, en lisant les options de droite à gauche, une version couleur, sans rotation, de taille maximale, de l'image entière)
- **Suffixe miniature** : à n'utiliser que si vos images sont hébergées sur un serveur IIIF. Spécifiez ici le suffixe de l'URL (en commençant par « / ») à utiliser pour afficher les vignettes des médias. Nous vous conseillons une qualité et une taille restreinte afin de ne pas ralentir l'affichage des pages, surtout si vous manipulez beaucoup de médias. La syntaxe précise dépend du serveur que vous utilisez mais vous aurez par exemple : `/full/250,/0/default.jpg` (on demande, en lisant les options de droite à gauche, une version couleur, sans rotation, de largeur 250px, de l'image entière)

Pour en savoir plus sur les spécifications IIIF, se reporter à la documentation officielle :

- <https://iiif.io/api/image/2.1>
- ou encore à la [page d'aide](#) de Gallica

2.3.4. Édition des métadonnées

Cette section vous permet de créer des métadonnées en définissant leur nom et leur valeur par défaut, et de spécifier les modalités d'application. Il est possible de créer les métadonnées une à une ou bien de téléverser un ensemble de métadonnées décrites dans un fichier .xml. Un exemple de format est disponible sur la plateforme TACT : <https://tact.demarre-shs.fr/data/metadatas-exemple.xml>.

Ces métadonnées seront incluses dans l'export des transcriptions. Elles peuvent aussi être manipulées par votre feuille de transformation XSL (voir section page).

2.3.5. Autres opérations

Cette section permet d'effectuer les actions suivantes :

- Archivage/désarchivage du projet : un projet archivé apparaît toujours sur la plateforme dans la section projets archivés, en bas de la page projets. Les utilisateurs ne peuvent plus s'y inscrire mais les transcriptions sont toujours consultables. Le(s) gestionnaire(s) peuvent exporter le projet mais pas le modifier. Vous pouvez aussi désarchiver à tout moment.
- Exporter : choisissez les éléments du projet que vous souhaitez télécharger (transcriptions, médias, métadonnées, informations utilisateurs ...)du projet ainsi que les informations.
- Supprimer : Il ne sera plus possible d'accéder au projet, à ses médias, transcriptions ou utilisateurs (même pour les administrateurs de la plateforme). Assurez-vous d'exporter les données que vous souhaitez conserver avant de supprimer votre projet. Cette opération est irréversible.

2.4. Gestion des médias

Il existe deux façons de gérer les médias à transcrire : héberger les médias sur un service externe (idéalement un service IIIF) ou bien héberger les images directement sur la plateforme TACT.

2.4.1 Hébergement interne des médias

Dans l'onglet « Médias », vous pouvez mettre en ligne les images (formats png ou jpeg), et, si besoin, les organiser en dossiers et sous-dossiers. Vous pouvez téléverser des images depuis votre ordinateur, ou vous pouvez charger un dossier zippé contenant des images dont l'architecture sera conservée.

2.4.2 Hébergement externe des médias

Le principe

Un hébergement externe consiste à utiliser un service spécialisé dans la gestion d'image plutôt que de déposer vos fichiers sur la plateforme TACT.

Un tel stockage vous permet de séparer physiquement images et transcriptions. Cela peut faire partie de la bonne gestion de vos données en évitant de stocker vos images à plusieurs endroits et en choisissant dès le départ d'utiliser un entrepôt de données dédié, sécurisé et performant. Lors d'autres exploitations de vos images (création d'un site web, inclusion dans un billet de blog, référence au sein d'un article...), vous pourrez systématiquement faire référence au même lieu d'hébergement.

TACT permet l'interopérabilité avec les serveurs IIF, ainsi vous pourrez afficher la qualité optimale d'image pour les contributeurs. Dans ce cas, pensez bien à renseigner lors de la création du serveur d'image (voir section [2.3.3](#)) les suffixes *large* et *miniature*.

En pratique

Deux étapes de configuration sont à prévoir :

1. Paramétrer le serveur d'image : voir la section [2.3.3](#)
2. Déposer une archive descriptive des médias et de leur arborescence

Dans l'onglet « Médias », choisir l'ajout de médias externes, déposer votre archive .zip et choisir le serveur concerné. Vous pouvez répéter cette opération autant de fois que nécessaire, si par exemple vos images sont stockées sur différents serveurs.

+ Ajout de média externe (IIF)

Figure 4 : Bouton pour l'ajout de médias externes.

L'archive .zip doit contenir un fichier .xml par média. Chaque xml doit contenir une balise <identifiant> ou <\$ns:identifiant>. Celle-ci doit intégrer l'identifiant de l'image sur le serveur IIF ou bien, si votre serveur d'image n'utilise pas la technologie IIF, le nom du fichier image extension comprise.

```
<identifiant>11280/2bb7063d</identifie  
r>
```

Figure 5 : Contenu minimal d'un fichier .xml.

Si besoin, l'archive .zip peut être organisée avec des sous-dossiers, TACT respectera cette hiérarchie.

Le nom du fichier .xml servira à nommer le média au sein de TACT. Par exemple, à partir du fichier 0001.xml issu du serveur nommé « Corpus 1530 », on créera le média « 0001 » avec un rappel du nom du serveur.

```
<?xml version="1.0"  
encoding="UTF-8"  
standalone="yes"?>
```

Fichier
0001.xml

+ Ajout de média externe (IIF)

Choisissez l'archive à uploader. Les médias seront associés au répertoire actuellement sélectionné. L'arborescence sera conservée. Chaque média doit disposer de son fichier XML. L'identifiant du média est récupéré dans la balise <identifiant>

Sélectionner un .zip *

Browse

Serveur *

Corpus 1530

L'ajout de nouveau serveur se fait dans "Editer > Edition des serveurs IIF"

Envoyer



Figure 6 : Schéma explicatif du nommage des

2.5. Gestion des utilisateurs

Cet onglet permet de gérer les inscriptions des utilisateurs à votre projet et leurs droits. Dans l'onglet « Utilisateurs », vous pourrez ainsi :

- valider les inscriptions des utilisateurs (si inscription modérée) ;
- invalider ou supprimer des utilisateurs pour le projet en question ;
- éditer le statut des utilisateurs inscrits à votre projet : par défaut, tous les utilisateurs validés inscrits à un projet sont « Transcripteurs ».

3. Que la transcription commence !

Faites une annonce, prévenez vos collègues et lancer votre projet de transcription !

Les transcripteurs doivent eux-mêmes se créer un compte sur la plateforme TACT, puis demander à « s'inscrire au projet ».

Si vous avez choisi une inscription **libre**, l'utilisateur pourra directement commencer à transcrire. Si vous avez plutôt choisi une inscription **modérée**, il vous faudra valider son inscription depuis l'onglet « Utilisateurs ». Nous vous conseillons alors de choisir de recevoir des notifications d'activité par mail (voir [figure2](#)) afin d'être le plus réactif possible !

3.1. Aider les contributeurs

Les utilisateurs qui participent à votre projet sont autonomes sur la plateforme, mais cela ne veut pas dire qu'ils sont seuls !

En effet, plusieurs outils sont disponibles pour venir en aide aux contributeurs : le manuel du contributeur et les aides situées dans l'onglet « Aide » de la page « Transcrire ».



Figure 6 : Détail de l'onglet "Aide" de la page "Transcrire".

3.1.1. Manuel du contributeur

Accessible depuis le bandeau bleu de la plateforme, ce menu mène à un *Manuel du Contributeur* qui synthétise les principes généraux de la transcription et du balisage ainsi que le fonctionnement et la prise en main de TACT pour accomplir ces tâches de transcription et d'annotation. Ce manuel est essentiel aux utilisateurs et peut faire écho à de nombreux questionnements, notamment pour les néophytes de la transcription. Identique quel que soit le projet, il est également accessible depuis la page d'accueil de TACT.

Nous vous invitons également à consulter avec attention ce manuel, pour connaître un peu mieux les tâches attendues des Contributeurs.

3.1.2. Tutoriel

Lancé automatiquement à chaque première transcription, le tutoriel passe en revue les différentes fonctionnalités d'une page de transcription. Il reste accessible à tout moment via l'onglet « Aide ».

3.1.3. Instructions de transcription

En tant que gestionnaire de projet, vous êtes vivement encouragé à mettre à disposition de vos contributeurs un document pour leur expliquer précisément ce que vous attendez de leur participation. Soyez clair, concis et précis, indiquez :

- ce qu'ils doivent transcrire : texte, retours à la ligne ou pas, erreurs d'orthographe corrigées ou non... selon votre projet.
- si vous avez un schéma de transcription, les phénomènes à annoter et les balises et attributs correspondants.

Pour rendre accessible ce document, rendez-vous dans Éditer > Édition des propriétés de base > Instructions de transcription (voir section 2.3. de ce guide).

3.1.4. Océrisation

Une fonctionnalité d'océrisation (bouton « Tesseract OCR ») est disponible lorsque aucune transcription n'a encore été enregistrée pour un média. Lorsque le traitement est effectué, le résultat de l'océrisation est automatiquement ajouté à la transcription mais non enregistré/sauvegardé tant que l'utilisateur ne l'a pas décidé.



Figure 7: Utilisation de l'OCR pour la transcription

3.2. Relecture

Les transcribers ne peuvent pas relire les transcriptions. Seuls les Relecteurs et les gestionnaires ont les droits nécessaires. C'est au(x) gestionnaire(s) que revient la décision de faire passer un utilisateur du rôle de transcriber à celui de relecteur. Il peut le faire via l'onglet « Utilisateurs » grâce à l'action « Éditer le statut ».

C'est l'action « Soumettre la relecture » qui permet de valider ou d'invalider une transcription.



Figure 8 : Options disponibles lors de la relecture.

Figure 9 : Boîte de dialogue de la soumission d'une relecture.

Si la qualité de la transcription ne vous satisfait pas, vous pouvez la modifier vous-même puis la valider ou bien invalider la transcription afin qu'elle retourne dans le lot de transcriptions « en cours » (cette action ne supprime pas la transcription déjà produite, mais modifie seulement son statut).

4. Et maintenant ?

Il est possible d'ajouter de nouveaux médias à transcrire, d'en retirer, etc.

4.1. Exports

À tout moment, il est possible d'« Exporter » tout ou partie du projet pour récupérer les transcriptions mais aussi les médias, la liste des utilisateurs ayant participé et toutes les données du projet.

Figure 10 : Options d'export

Il est également possible d'exporter des transcriptions « au fil de l'eau », directement via la page de transcription.

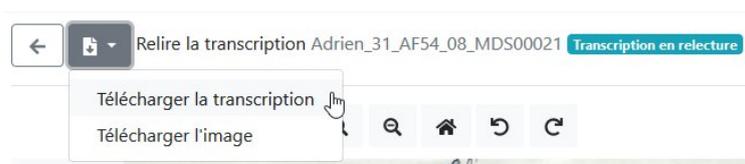


Figure 11 : Export au fil de l'eau

4.2. Clôture du projet

Le gestionnaire peut, lorsque le projet n'est plus actif, l'archiver ou le supprimer (voir section « [Autres opérations](#) »).

N'hésitez pas à nous faire part de vos questions ou suggestions à l'adresse :

tact@univ-grenoble-alpes.fr