Tutoriel – WebStamp2Excel

Cet outil est conçu pour créer une collection sous Excel à partir de sites Web qui ne proposent pas cette option, et accessoirement de récupérer les images de timbres pour construire vos albums.

L'intérêt d'une collection maintenue dans un fichier Excel hormis son évaluation est son indépendance par rapport à des bases de données propriétaires et sa portabilité ; elle peut être visualisée sur tous les tableurs.

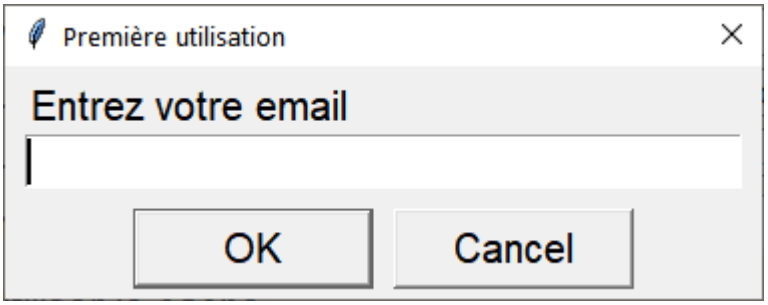
Le logiciel intègre des aides sous forme d'un champ texte central et de bulles d'informations en survol sur les champs.

Les fichiers Excel générés sont modifiables, si des colonnes sont supprimées tant que les titres des colonnes sont préservés ils resteront compatible avec ce logiciel.

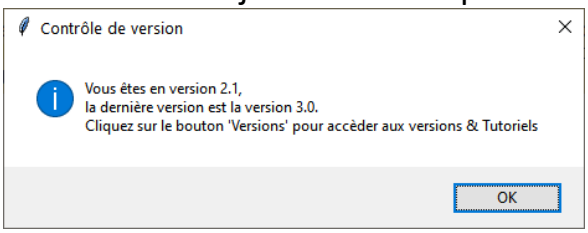
Ce logiciel est un programme Python dont l'exécutable est fourni pour PC, il sera possible de le produire pour MAC.

L'exécutable fourni est « safe », il a été contrôlé par des anti-antivirus, il n'écrit sur le disque sous son arborescence d'installation et il nécessite bien sûr l'accès à Internet.
Il est signé mais avec un certificat auto-signé, cependant il peut générer un avertissement pour des antivirus restrictifs ne laissant passer que des certificats publiés, il suffit dans ce cas de passer l'avertissement pour pouvoir l'utiliser.

Ecran de mise en route
Un email vous sera demandé à la première mise en route, il ne servira qu'à vous enregistrer pour vous tenir informé des évolutions du produit, il ne vous sera plus demandé par la suite.



Si vous aviez déjà une version précédente vous aurez un avertissement au démarrage :



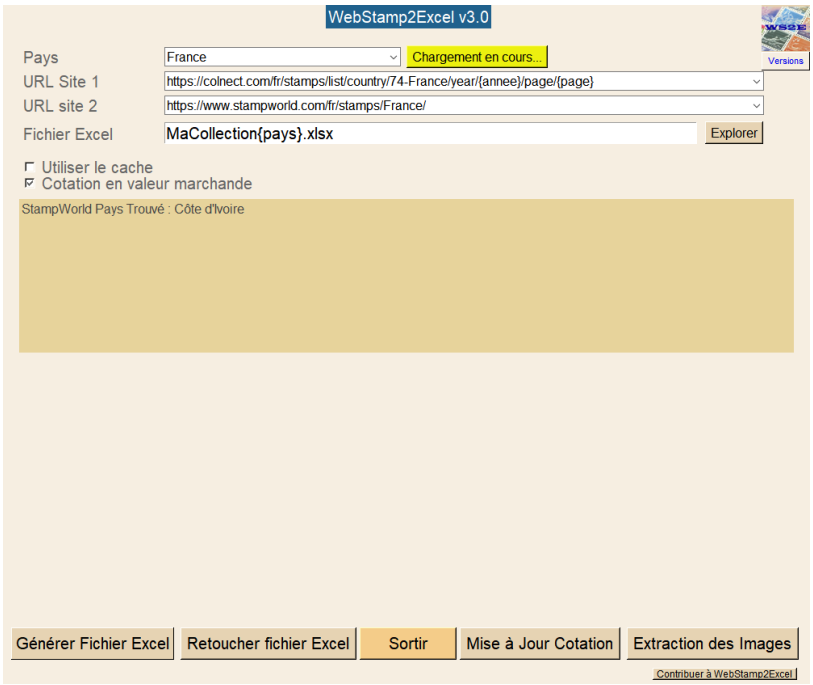
Ecran principal

Pour charger une nouvelle version cliquez sur le lien en bleu « Versions » pour aller sur la page de téléchargement où vous trouverez l'historique des versions et un lien de contact.

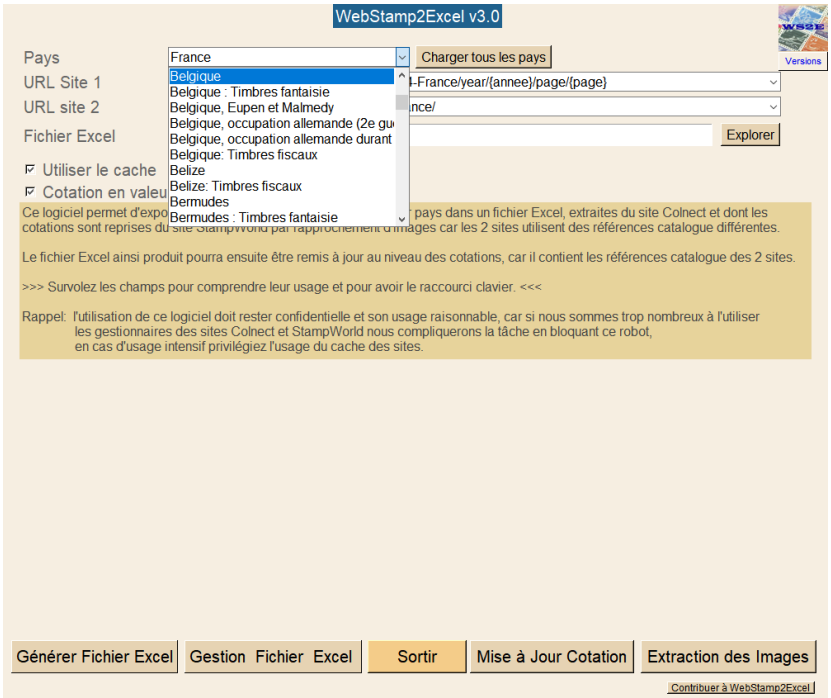
Si vous connaissez des sites ou si vous avez des idées qui apporteraient un plus à ce logiciel n'hésitez pas à les communiquer par le lien de contact.

Page web de téléchargement

Pour commencer si vous collectionnez d'autres pays que la France ,
chargez la liste des pays.



Liste déroulante des pays : les noms de pays sont aussi accessibles par la saisie des 3 premières lettres du pays quand on se place dans le champ [Pays].



Une fois chargée, elle restera en mémoire dans le fichier « config.ini »
avec la plupart des options courantes.

On va pouvoir passer maintenant à la génération de fichiers Excel.

Dans tous les cas il est recommandé de constituer un cache des sites sur votre machine pour ne pas saturer les sites utilisés de requêtes répétées qui risquent de vous rendre ces sites inaccessibles.

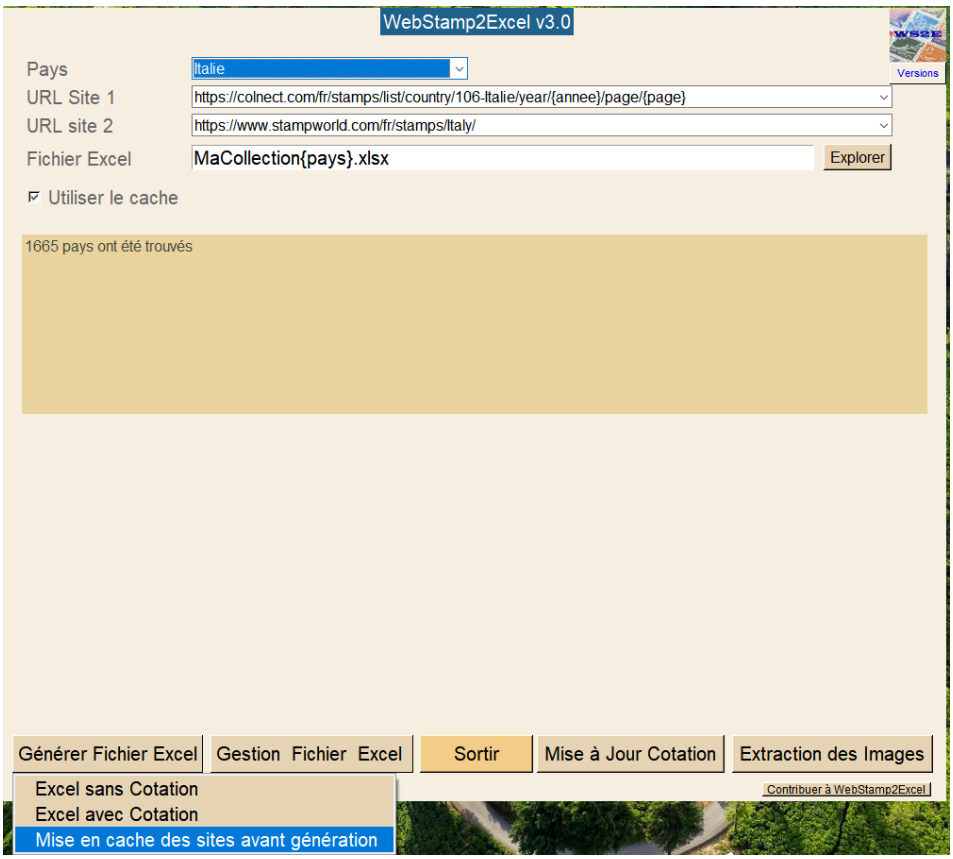
Les caches sont organisés par site et pays, si vous voulez constituer la collection d'un pays, constituez son cache en précisant un intervalle d'année qui recouvre complètement votre collection.

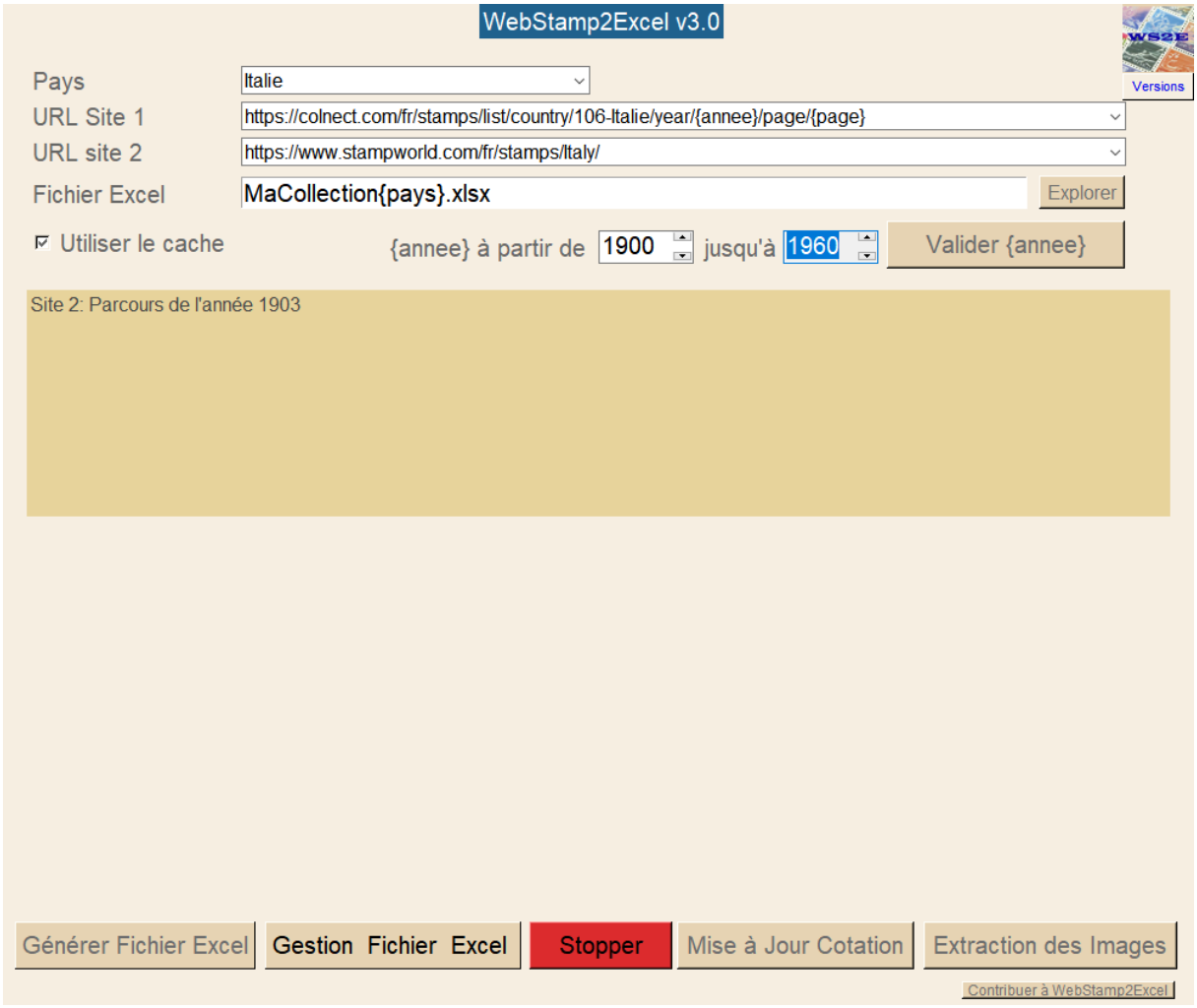
Vous pourrez ensuite générer vos fichiers Excel à loisir sans avoir besoin de vous connecter et elles pourront donc être générées très rapidement, ce qui vous permettra d'ajuster plus facilement les paramètres qui conviennent en fonction du rendu final.

De plus si vous voulez vous servir des images pour constituer vos propres albums vous n'aurez pas besoin de passer par l'extraction d'image, car vous les retrouverez toutes dans le répertoire « **cache** »

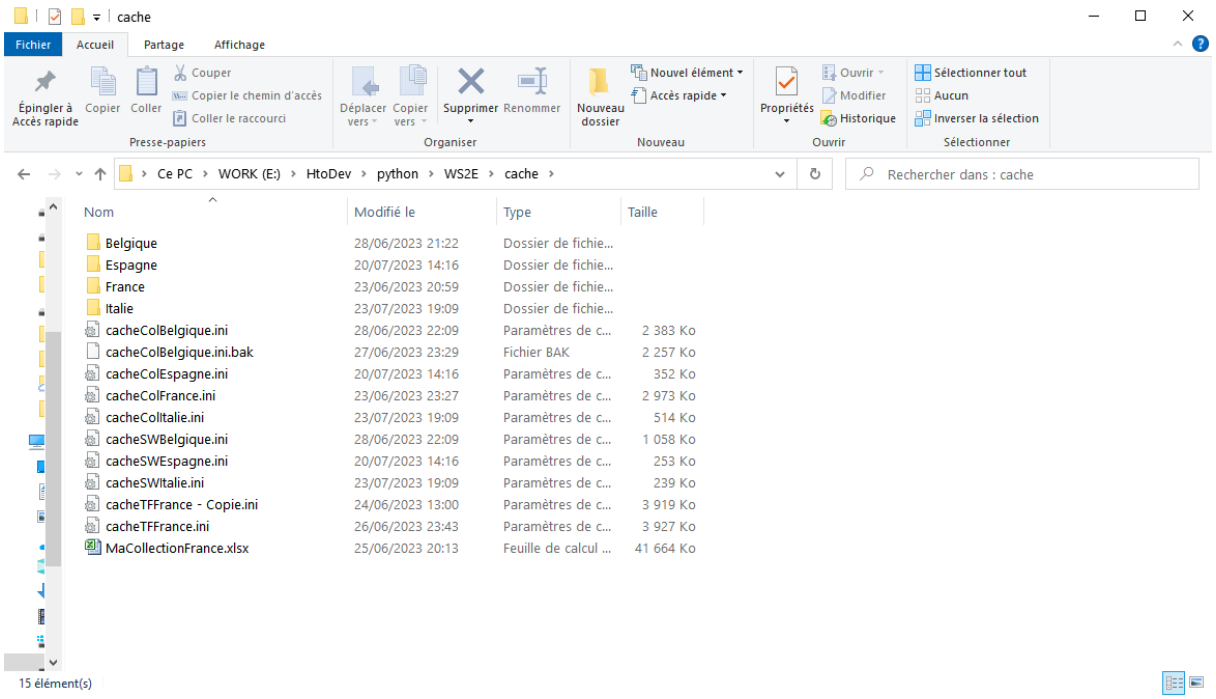
Mise en cache d'un pays

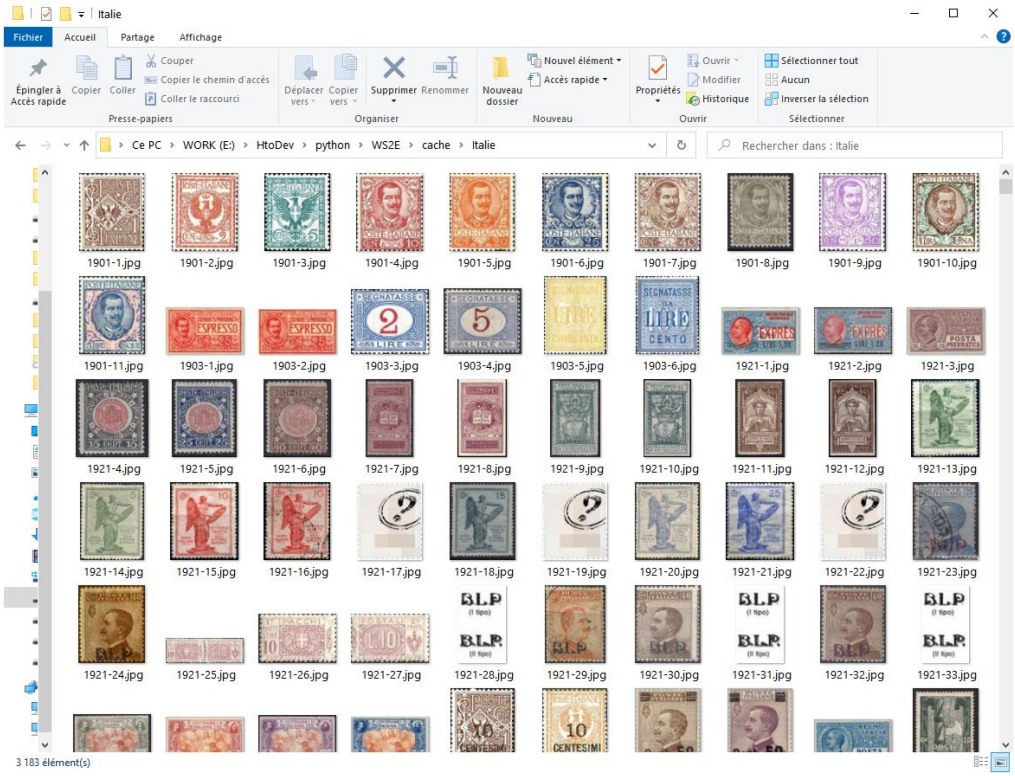
Sélectionnez [Générer Fichier Excel] puis [Mise en cache des sites avant génération], sélectionner votre intervalle d'année puis [Valider {année}]



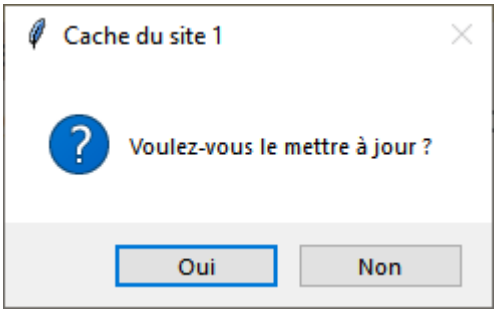


Aperçu du répertoire cache





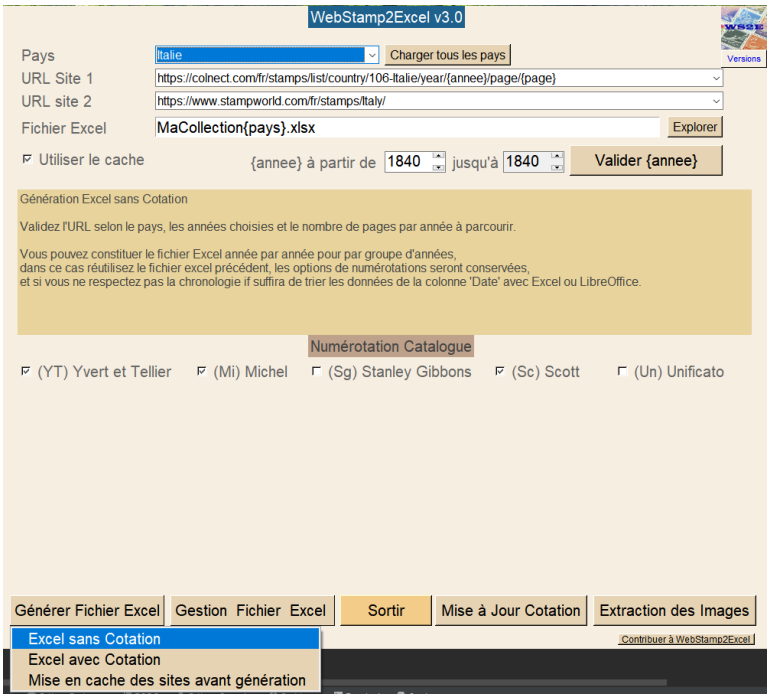
A noter <u>le cache est évolutif</u>, si vous le régénérer une boite de dialogue vous proposera de le faire évoluer au lieu de l'écraser, au cas où vous voudriez compléter des années.



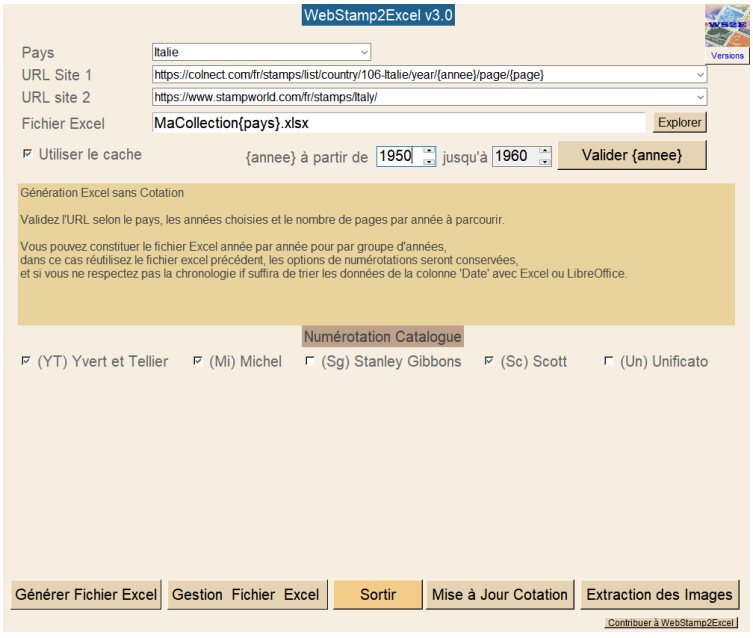
Génération d'un fichier Excel non côté à partir du cache
La génération se fait de la même façon sans utiliser le cache, la seule différence sera la vitesse de la génération qui sera plus lente !

La génération se fait par intervalle d'année, mais vous pouvez revenir sur le même fichier Excel dans ce cas une fenêtre vous proposera d'écraser l'ancien fichier ou de l'enrichir d'un nouvel intervalle, dans ce cas le fichier Excel restera triable sous Excel au cas où la chronologie des années n'était pas respectée.

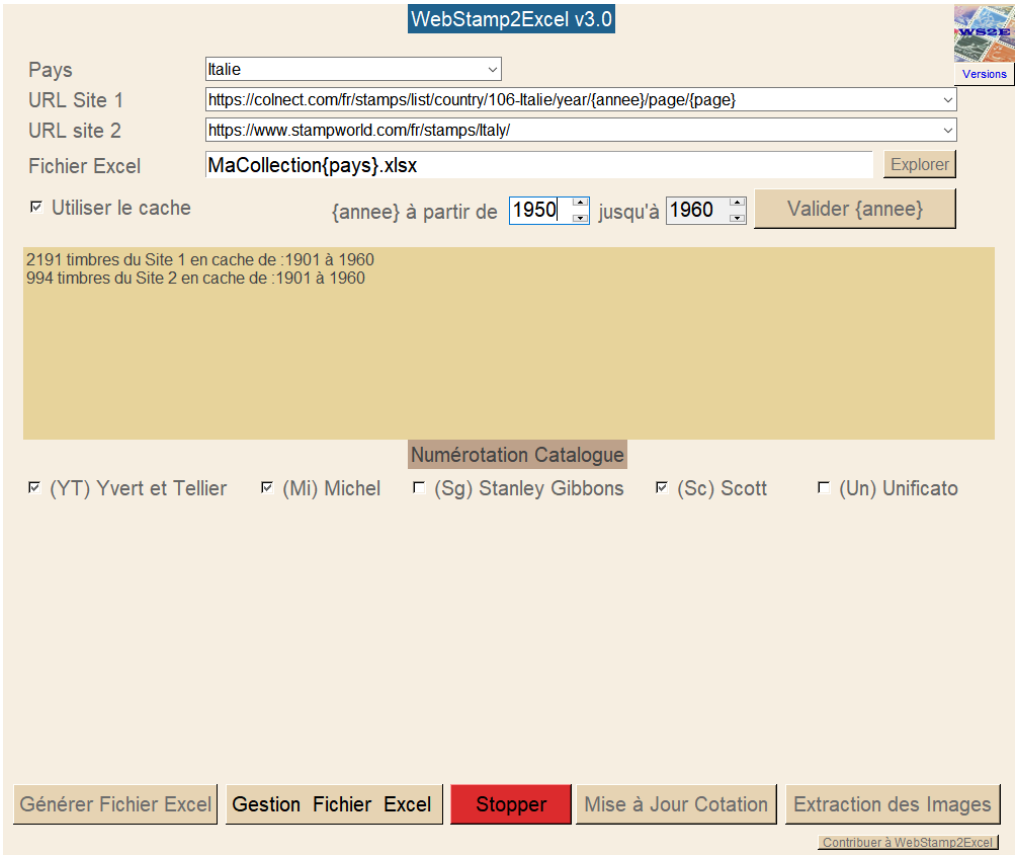
Sélection menu [Excel sans Cotation]



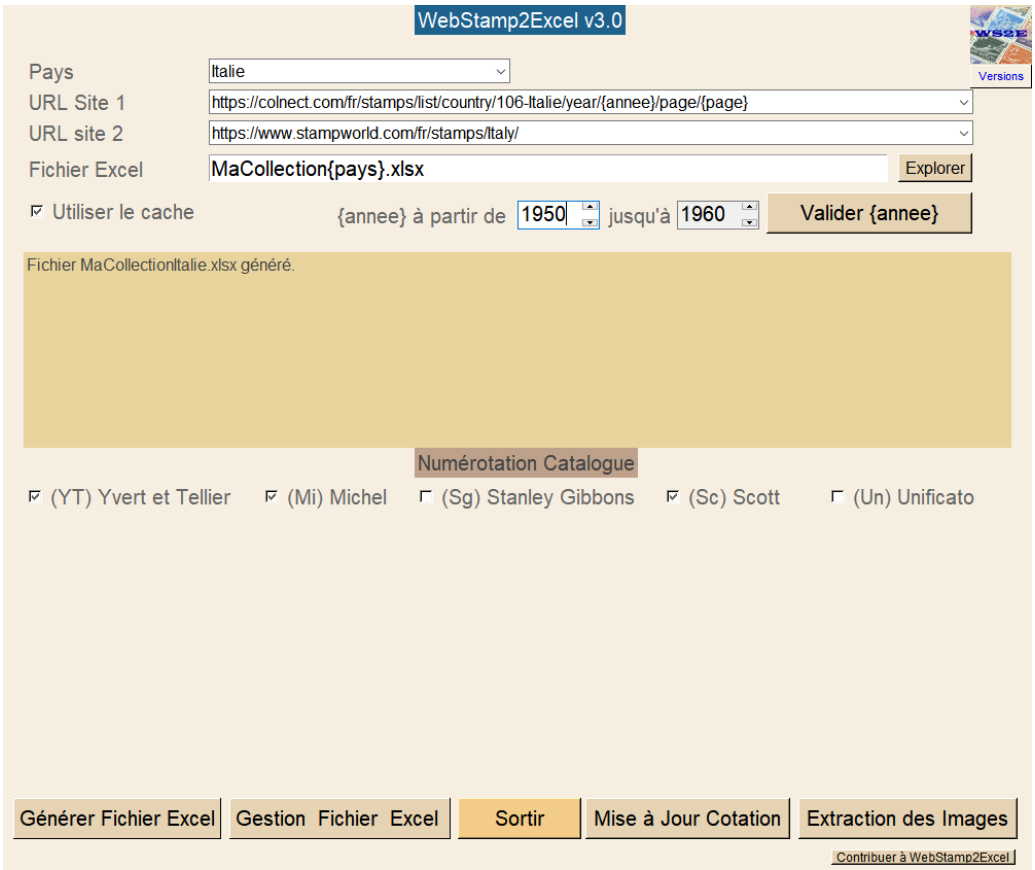
Validation des années à produire puis [Valider {année}]



Le processus reste interruptible durant la génération par [Stopper]



A la fin de la génération, vous êtes informé dans le champ central.



Fichier Excel généré

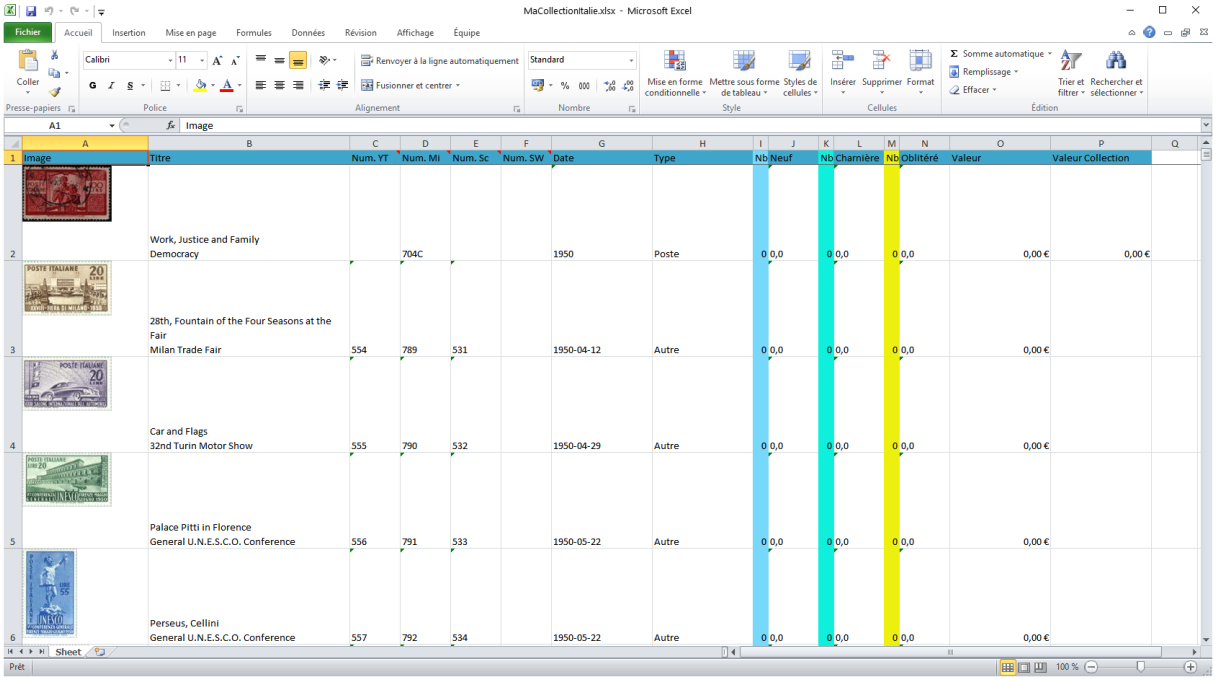
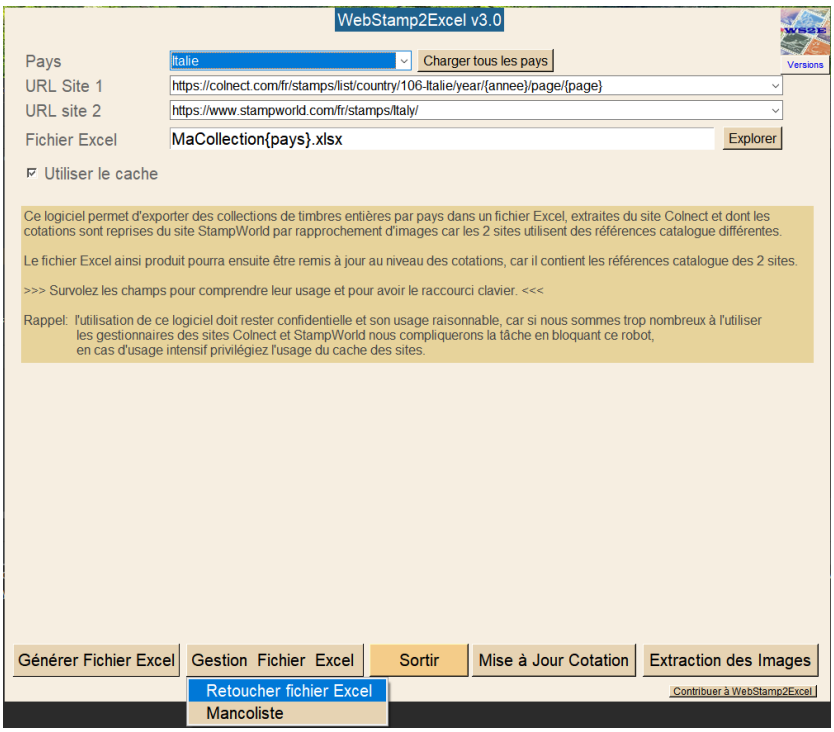


Image	Titre	Num. YT	Num. MI	Num. Sc	Num. SW	Date	Type	Nb	Neuf	Nb	Charnière	Nb	Oblitéré	Valeur	Valeur Collection
	Work, Justice and Family Democracy		704C			1950	Poste	0	0,0	0	0,0	0	0,0	0,00 €	0,00 €
	28th, Fountain of the Four Seasons at the Fair Milan Trade Fair	554	789	531		1950-04-12	Autre	0	0,0	0	0,0	0	0,0	0,00 €	
	Car and Flags 32nd Turin Motor Show	555	790	532		1950-04-29	Autre	0	0,0	0	0,0	0	0,0	0,00 €	
	Palace Pitti in Florence General U.N.E.S.C.O. Conference	556	791	533		1950-05-22	Autre	0	0,0	0	0,0	0	0,0	0,00 €	
	Perseus, Cellini General U.N.E.S.C.O. Conference	557	792	534		1950-05-22	Autre	0	0,0	0	0,0	0	0,0	0,00 €	

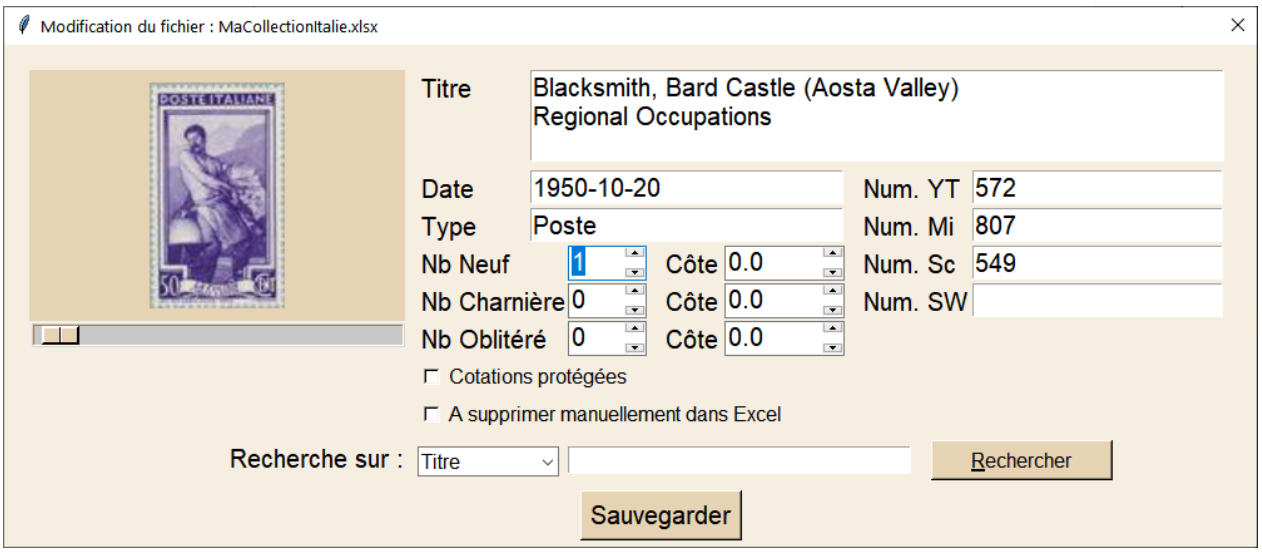
Les formules de calcul de la valeur de la collection sont intégrées au fichier Excel, il suffit de saisir le nombre (Nb) possédé pour chaque état (Neuf/Charnière/Oblitéré)

Modification du fichier Excel

Sélectionnez [Gestion Fichier Excel] puis [Retoucher fichier Excel]



Ecran de modification du fichier Excel après chargement



Vous pouvez naviguer dans la collection via la barre de défilement, la molette de la souris, par les flèches page haut/bas, ou par recherche sur un champ.

Tous les champs sont modifiables durant la navigation en préservant le focus sur le champ courant, la sauvegarde se fait en fin de saisie.

Après chaque modification du fichier une copie datée du fichier précédente est générée sous la forme suivante : MaCollectionItalie_copieDu20230727_19h28m29.xlsx

case à cocher « Cotations protégées » :

Si cette case est cochée, les cotations de ce timbre ne pourront plus être mise à jour par le Bouton [Mise à jour Cotation], mais elles resteront modifiables manuellement. Dans le fichier Excel le survol de la cellule affichera : Cotations protégées.

case à cocher « A supprimer manuellement dans Excel » :

si cette case est cochée, le timbre sera à supprimer en marquant les lignes Excel correspondantes d'un fond rouge ; On ne fait pas la suppression dans le logiciel car la librairie Python utilisée ne permet pas la mise à jour automatique des formules associées aux cellules, il faudra donc supprimer ces lignes dans votre Tableur.

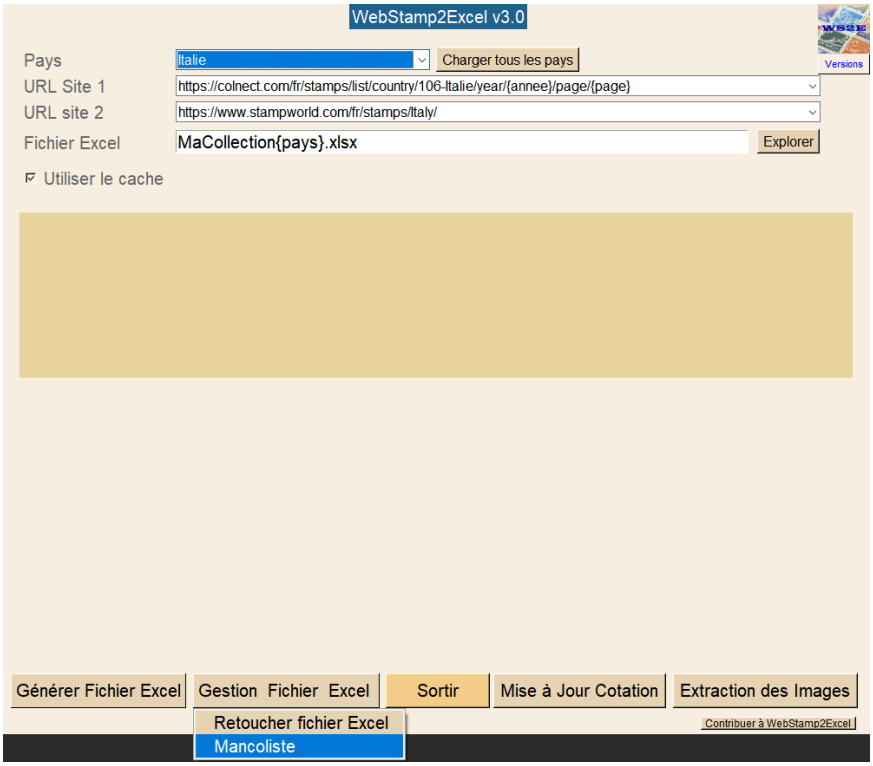
Génération de la Mancoliste

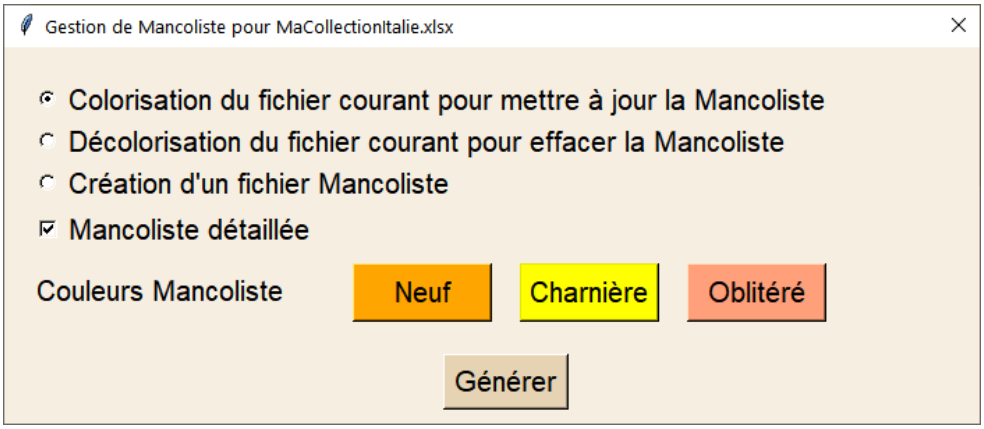
La mancoliste permet d'établir une liste des timbres manquants dans une collection, 2 méthodes sont proposées :

- ⇨ Marquer d'une couleur spécifique les timbres manquants dans le fichier Excel courant selon l'état (Neuf/Charnière/Oblitéré) si [Mancoliste détaillée] est cochée, ou simplement marquer d'une couleur les manquants quelque soit leur état si décochée. Dans ce cas il sera possible de décoloriser le Fichier Excel si vous souhaitez retirez le marquage de la mancoliste.

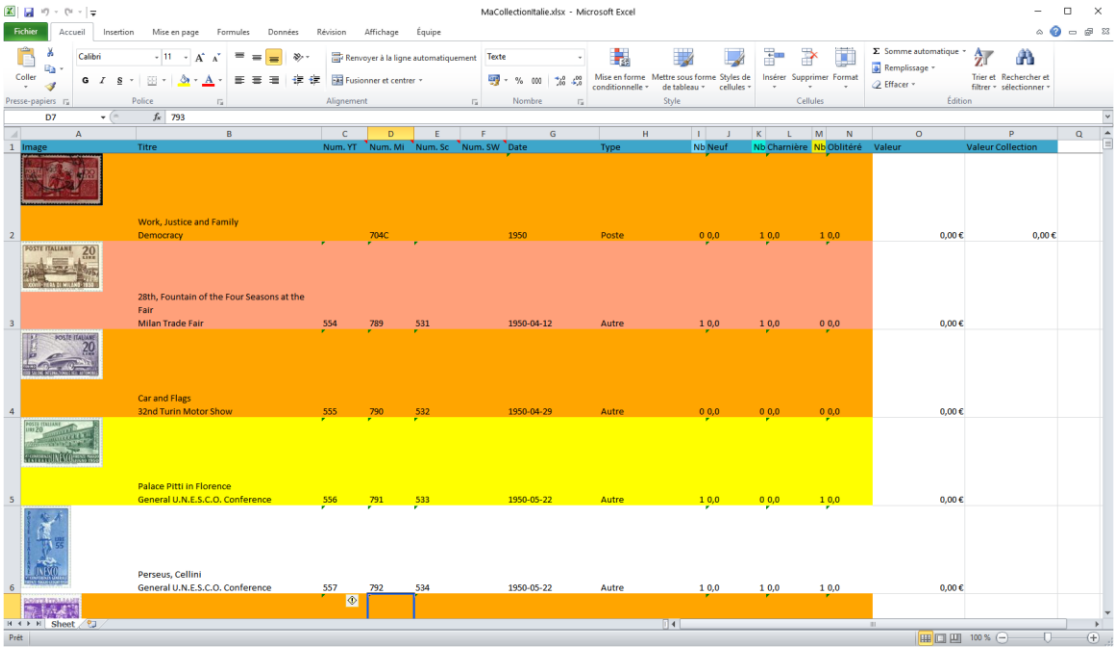
- ⇨ Créer un fichier Excel Mancoliste avec uniquement les timbres manquants marqués d'une couleur spécifique selon l'état (Neuf/Charnière/Oblitéré) si [Mancoliste détaillée] est cochée. Si [Mancoliste détaillée] est décochée, le fichier Excel Mancoliste comportera les timbres manquants sans colorisation. A noter la création du fichier Mancoliste prend plus de temps que la colorisation du fichier.

Sélectionnez [Gestion Fichier Excel] puis [Mancoliste]

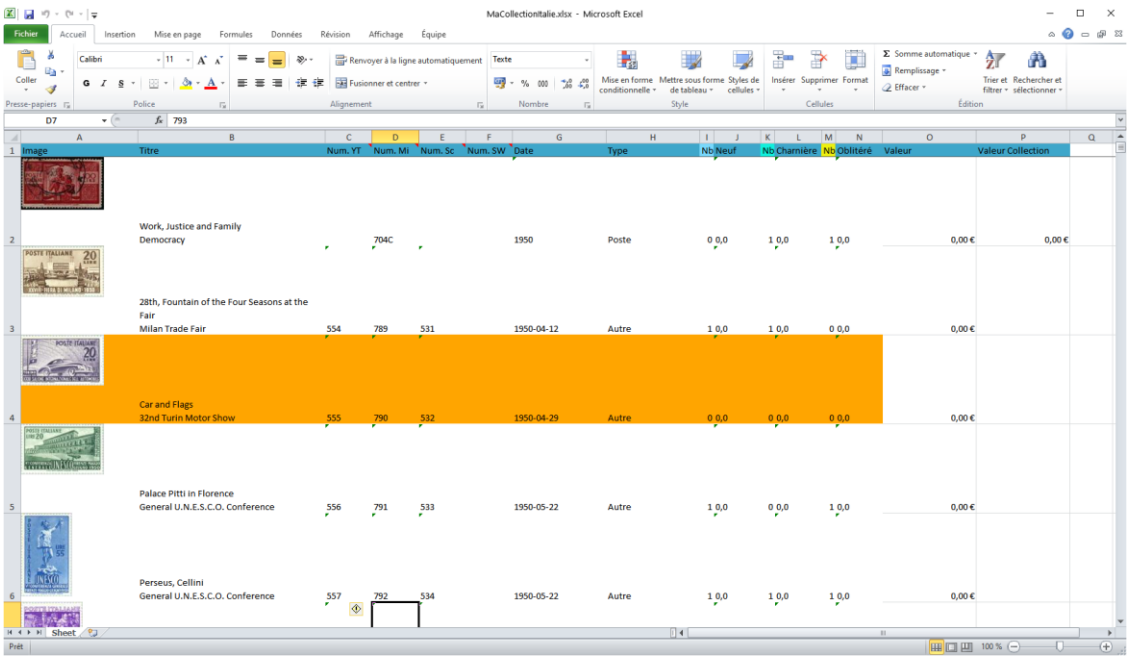




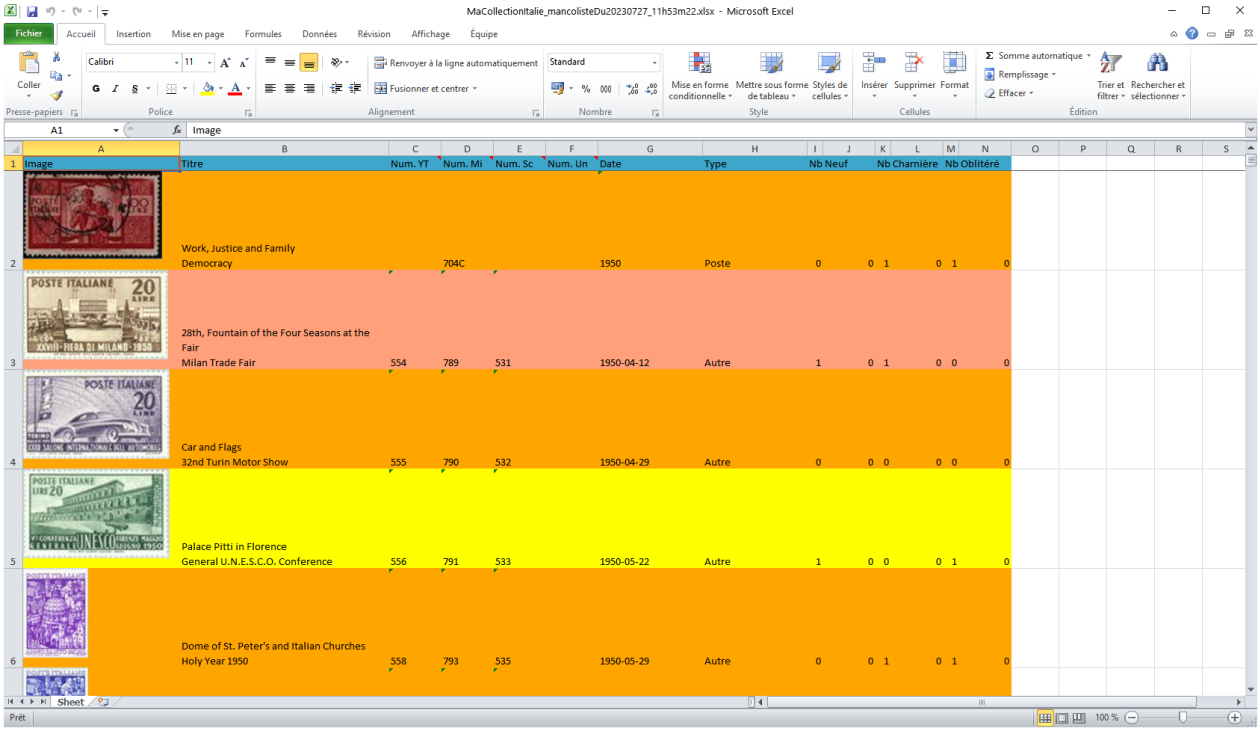
Résultat [Colorisation du fichier courant...] **avec** option Mancoliste détaillée



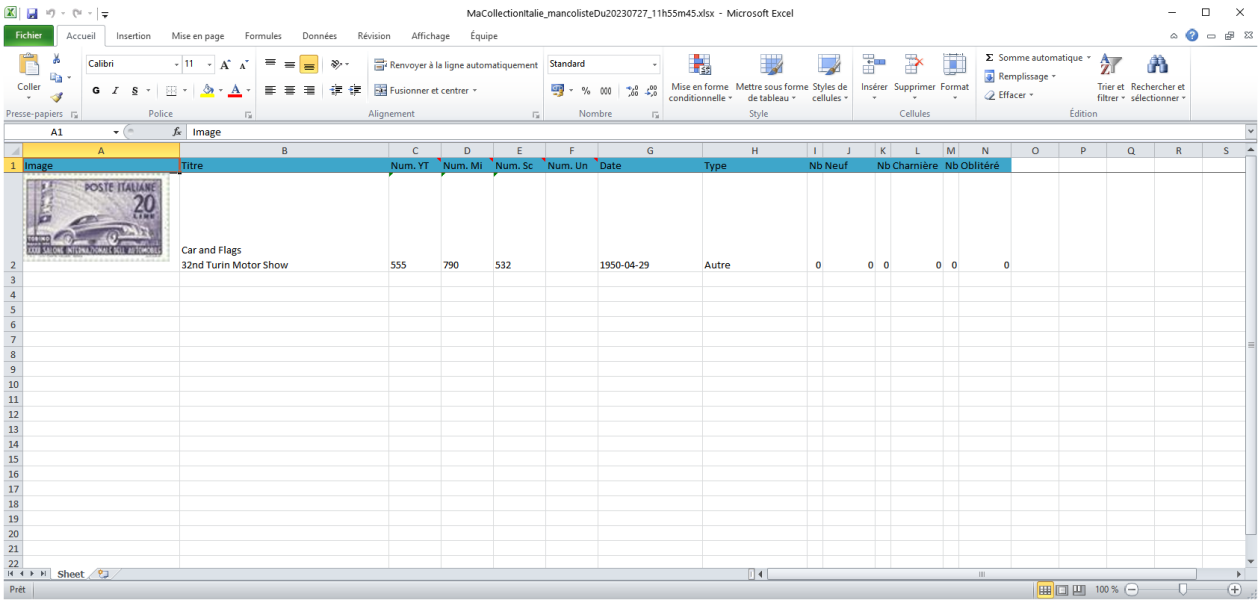
Résultat [Colorisation du fichier courant...] **sans** option Mancoliste détaillée



Résultat [Création d’un ficher Mancoliste] **avec** option Mancoliste détaillée



Résultat [Création d’un ficher Mancoliste] **sans** option Mancoliste détaillée



Génération d'un Fichier Excel côté

La génération se fait par intervalle d'année, mais vous pouvez revenir sur le même fichier Excel dans ce cas une fenêtre vous proposera d'écraser l'ancien fichier ou de l'enrichir d'un nouvel intervalle, dans ce cas le fichier Excel restera triable sous Excel au cas où la chronologie des années n'était pas respectée.

- ⇨ **pour la France**

La génération d'un fichier côté pour la France ne nécessite pas de reconnaissance d'image, si l'option « Valeur Marchande» (par défaut pour la France) est cochée. En revanche si vous décocher cette option vous opterez pour une valeur catalogue StampWorld et vous devrez suivre la même procédure que pour les autres pays.

- ⇨ **pour les autres pays que la France**

C'est là que les choses se compliquent un peu mais c'est aussi un peu l'originalité de ce logiciel, car pour retrouver une valeur catalogue sur le site StampWorld il va falloir utiliser la reconnaissance d'image ; ce site ne référençant les timbres que par une numérotation qui lui est propre.

3 méthodes de validation des timbres sont proposées :

- ⇨ **Pour les besogneux** : la validation manuelle vous propose un écran qui vous permet de valider visuellement le lien entre 2 timbres sur chaque site, cet écran vous facilite la tâche par des raccourcis claviers mais il faut un peu patience.

- ⇨ **Pour les pragmatiques** : <u>la validation interactive est la plus recommandée</u> vous propose d'utiliser la reconnaissance d'image mais d'une façon restrictive quand les seuils paramétrés pour chaque algorithme de reconnaissance d'image ne sont pas atteints.
  Seuls les timbres qui seront en dessous de ces seuils seront proposés à la validation manuelle, ce qui limite le nombre d'intervention manuelle.
  De plus cette saisie tentera de proposer en première intention le timbre qui a la meilleure similarité afin d'écourter cette phase.

- ⇨ **Pour les téméraires :** la validation peut se faire complètement automatiquement mais il va falloir régler par pays les critères de reconnaissances d'images avec perspicacité au risque d'avoir des erreurs de reconnaissances. Dans ce cas le fichier Excel résultant vous aidera à ajuster vos paramètres car il comporte une colonne image rapprochée et une colonne critère qui énonce les critères de rapprochement qui ont validé le rapprochement d'image.

Guide du rapprochement d'image

Le rapprochement automatique d'image repose sur 3 algorithmes paramétrables par seuil que l'on peut cumuler en une condition.

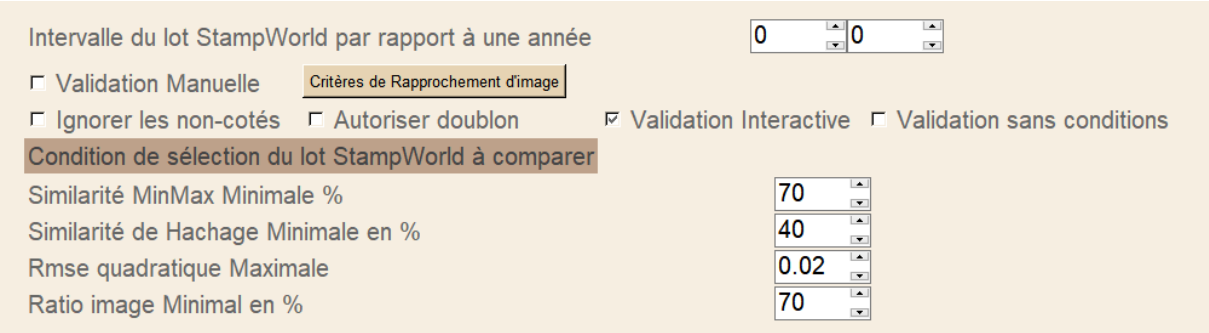
Une liste de conditions peut être établie par pays, accès par bouton [Critères de rapprochement d'image]

Pour chaque timbre Colnect à comparer, la liste de conditions est comparée à un lot de timbre StampWorld dont la taille varie selon [L'intervalle du lot StampWorld …] et selon la [Condition de sélection du lot StampWorld].

La liste est parcourue séquentiellement, dès qu'une condition est satisfaite le rapprochement est validé. En revanche si aucune n'est satisfaite l'image n'est pas rapprochée ce timbre n'aura pas de cotation hormis par validation manuelle.

<u>Une condition est considérée comme satisfaite quand les seuils de comparaisons des algorithmes choisis sont tous atteints</u>.

Condition de sélection du lot StampWorld à comparer

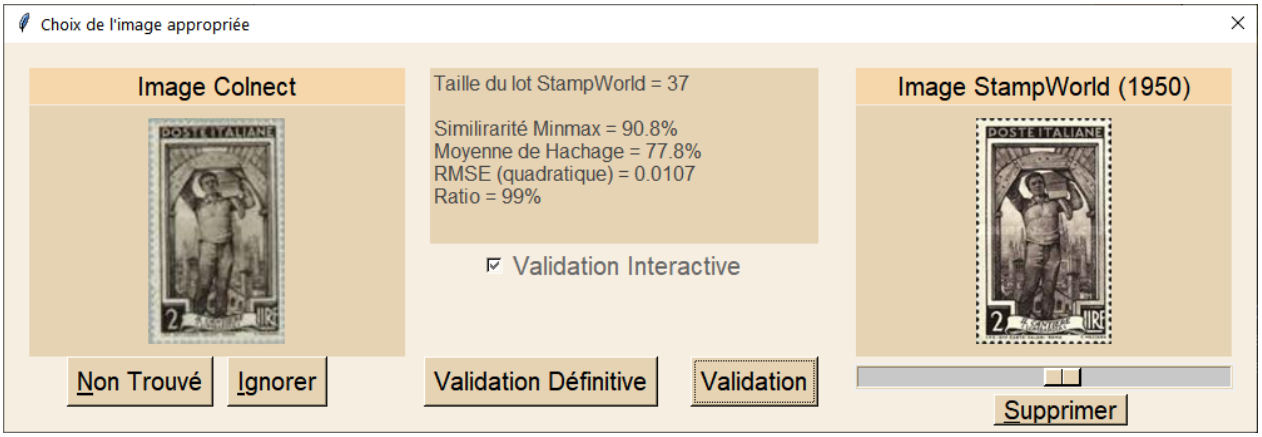


La comparaison de timbres se fait par lot d'années pour un timbre Colnect d'une année précise, on compare tout le lot de timbre StampWorld de l'année correspondante enrichie de [L'intervalle du lot StampWorld …] éventuel.

Si l'option [Validation sans Condition] est cochée le lot StampWorld à comparer sera complet.

Si l'option [Validation sans Condition] est non cochée
le lot StampWorld à comparer sera restreint au seuil fixé pour chaque critère de comparaison, plus ces critères seront restrictifs plus le lot sera petit ce qui améliorera la vitesse de traitement mais le risque de louper une association sera en revanche plus grand si les critères sont trop restrictifs. Les valeurs par défaut se veulent assez large pour simplement éliminer les timbres qui n'ont aucune similitude.

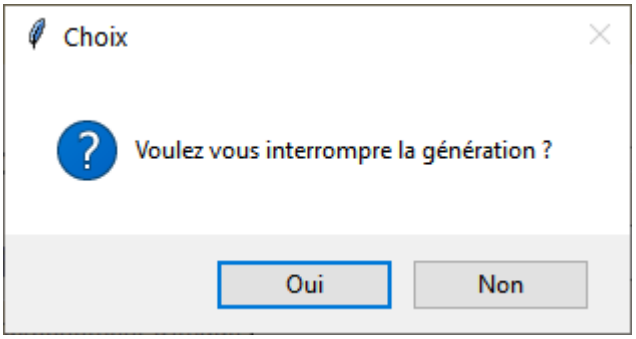
Ecran de Validation Interactive



A ce stade vous pouvez visuellement valider l'équivalence d'un timbre entre les 2 sites.
La navigation dans le lot se fait par la barre de défilement, les flèches ou par la molette.
Les lettres soulignées dans les boutons en sont les raccourcis.

[Validation] : l'équivalence est retenue
[Validation Définitive] : l'équivalence est retenue **et** le timbre est sorti du lot
[Supprimer] : l'équivalence n'est pas retenue **et** le timbre est sorti du lot
[Ignorer] : ce timbre ne sera pas dans la collection
[Non Trouvé] : Ce timbre est retenu dans la collection sans équivalence donc sans cotation
[Validation Interactive] : La validation interactive peut être interrompue après la sortie de saisie, si la case à est décochée.

Si cette fenêtre est fermée par X en haut à droite, la génération peut être interrompue.



Les algorithmes de similarité d’image

Les algorithmes de similarité d’image (Minmax, Hash, Rmse) ont été retenus sur un compromis lié à leur vitesse de traitement de l'ordre du centième de seconde pour une comparaison, en les combinant judicieusement on atteint un taux de reconnaissance satisfaisant mais n'excluant pas une validation manuelle.

Des algorithmes plus discriminants existent mais sont coûteux en temps de traitements certains dépassent la minute pour une comparaison !

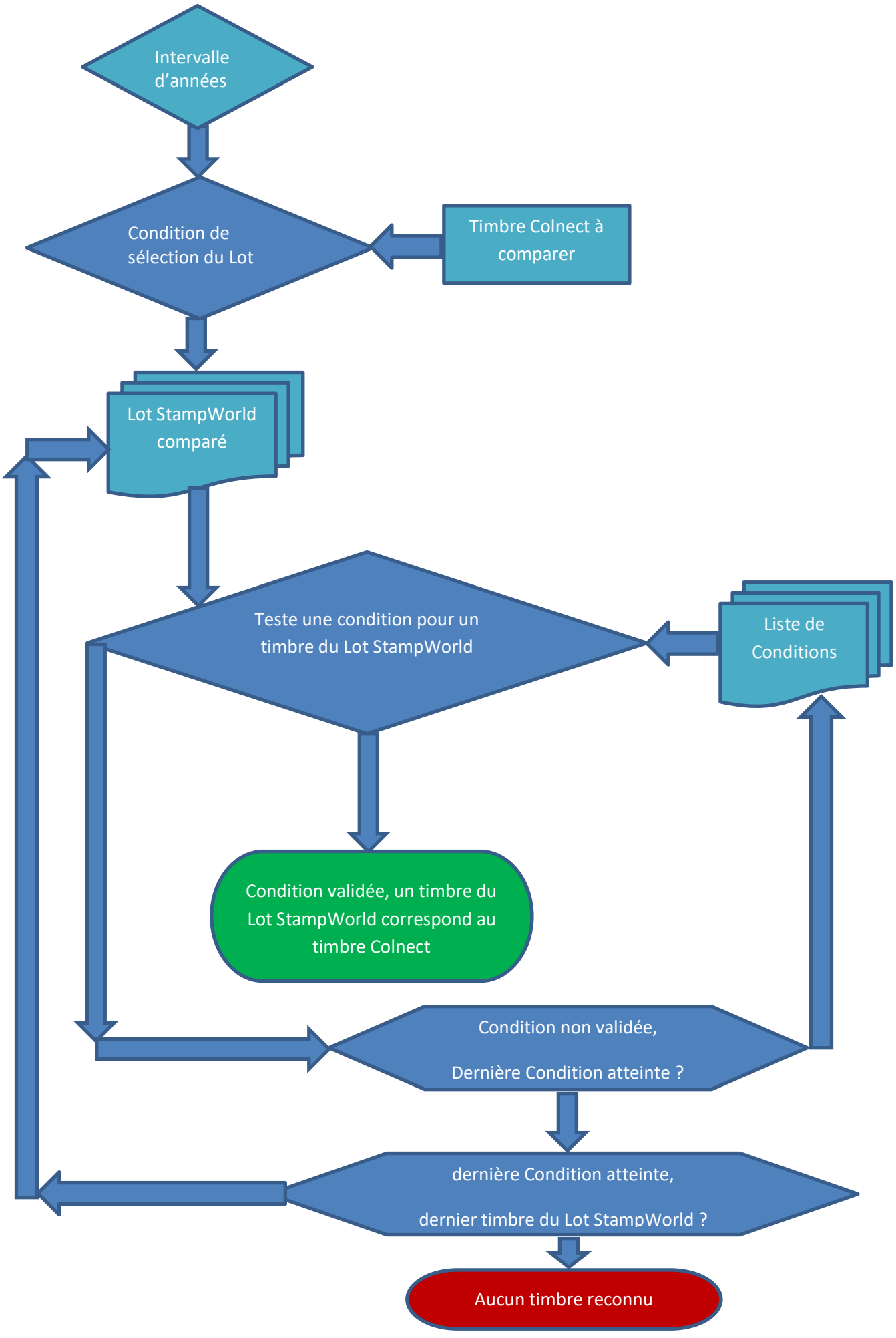
La similarité MinMax est de 100% quand l'image est identique, elle est relativement fiable à partir de 92%, à 0% cette similarité ne sera pas testée !

La similarité de Hachage est de 100% quand l'image est identique, elle est relativement fiable à partir de 60%, à 0% cette similarité ne sera pas testée !

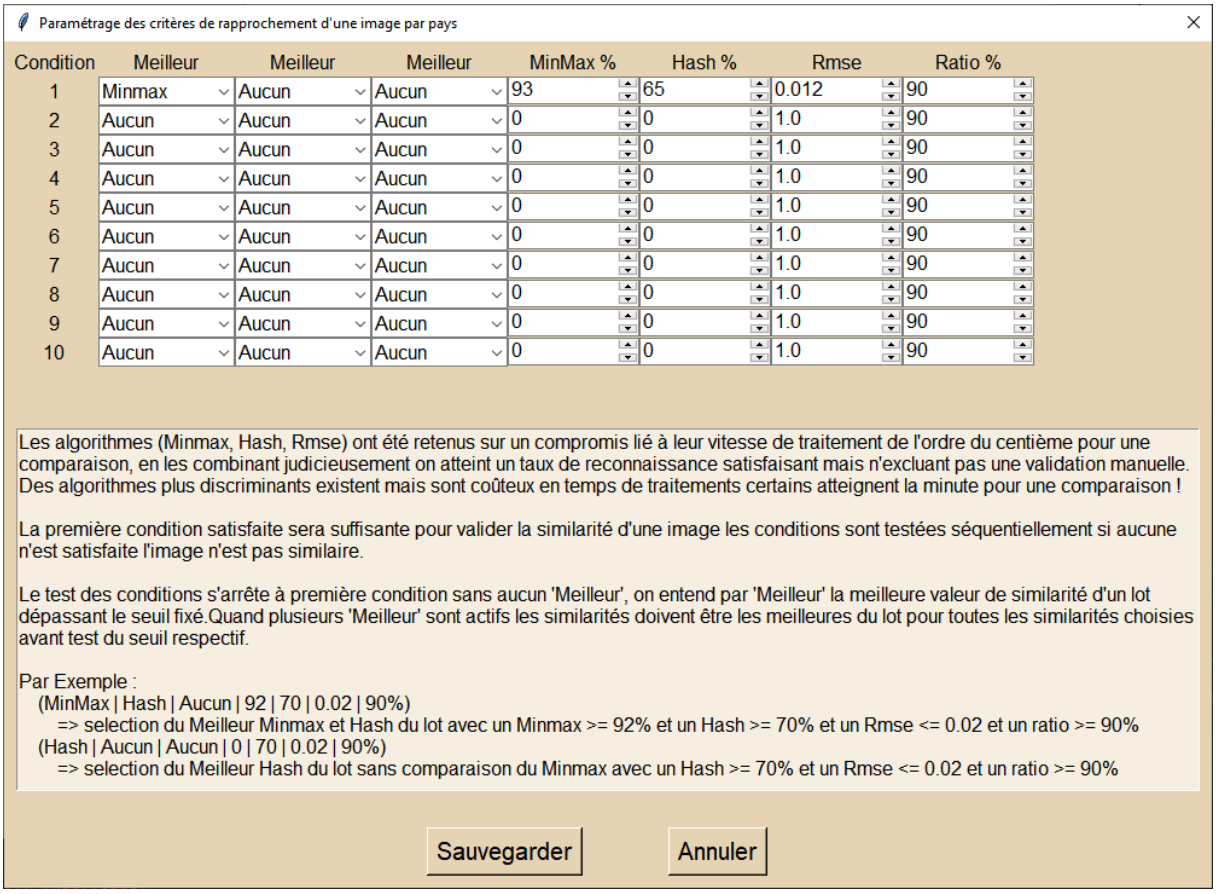
La similarité Rmse donne un taux de différences entre les images, attention le seuil est inversé !, elle est de 0 pour des images identiques et relativement fiable sous 0.015, à 1 cette similarité ne sera pas testée !

Le Ratio est n'est pas un critère de reconnaissance d'image, mais un critère sélectif qui permet d'éliminer les timbres qui n'ont pas les mêmes proportions. Il est de 100% pour des timbres de même proportion. Mais il faut toujours accepter une marge car les images des 2 sites sont différentes pour le même timbre, et aussi car dans certains pays le timbre possède un TAB attaché qui change radicalement sa proportion !

Processus de validation d'un timbre selon les critères de validation



Ecran de configuration de la liste de conditions



Exemple de conditions :

(MinMax | Hash | Aucun | 92 | 70 | 0.02 | 90%) :

Cette condition est satisfaite si le timbre du lot StampWorld a :
le meilleur Minmax et Meilleur Hash du lot avec
Minmax >= 92% et Hash >= 70% et Rmse <= 0.02 et un ratio >= 90%

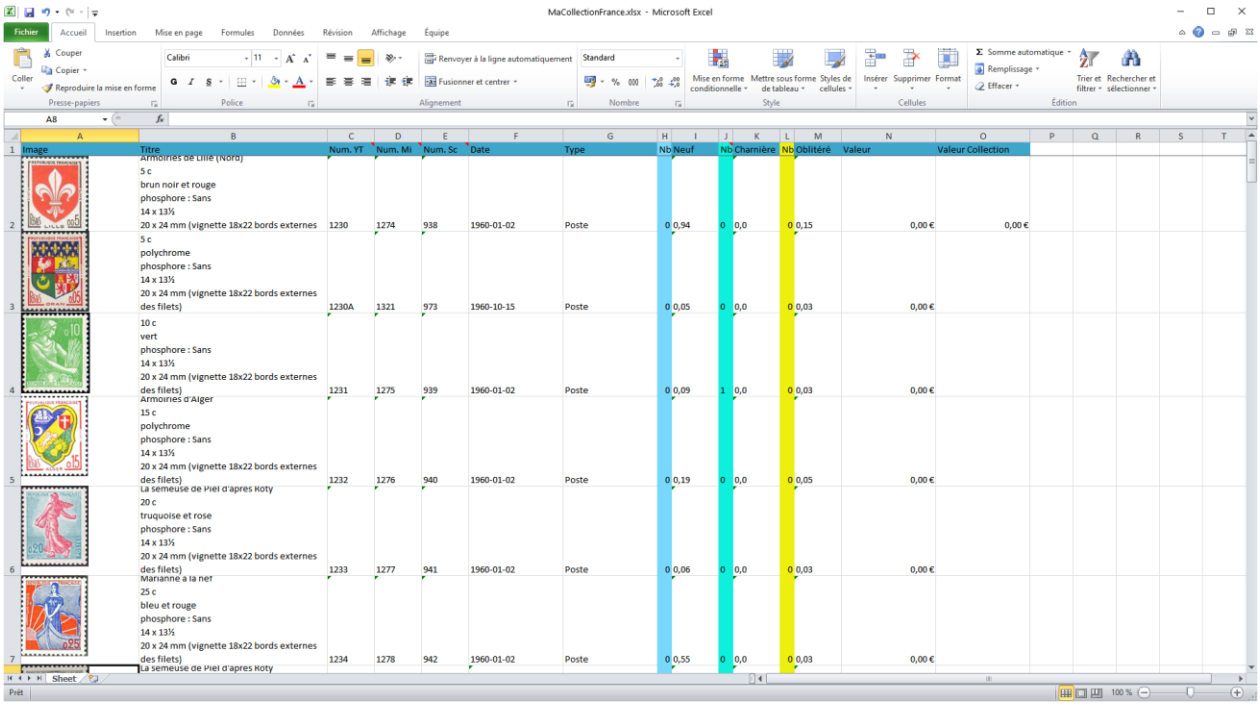
(Hash | Aucun | Aucun | 0 | 70 | 0.02 | 90%) :

Cette condition est satisfaite si le timbre du lot StampWorld a :
le meilleur Hash du lot avec
Hash >= 70% et un Rmse <= 0.02 et un ratio >= 90%
!!! A noter pas de comparaison du Minmax car il est à 0 !!!

(Aucun | Aucun | Aucun | - | - | - | -) :

Cette condition marque la fin de la liste des conditions

Génération d’un Fichier Excel côté pour la France



MaCollectionFrance.xlsx - Microsoft Excel

	Image	Titre	Num. YT	Num. Mi	Num. Sc	Date	Type	Nb	Neuf	Nb	Charnière	Nb	Oblitéré	Valeur	Valeur Collection
2		Armoiries de Lille (Nord) 5 c brun noir et rouge phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes	1230	1274	938	1960-01-02	Poste	0	0,94	0	0,0	0	0,15	0,00 €	0,00 €
3		5 c polychrome phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes des filets)	1230A	1321	973	1960-10-15	Poste	0	0,05	0	0,0	0	0,03	0,00 €	
4		10 c vert phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes des filets)	1231	1275	939	1960-01-02	Poste	0	0,09	1	0,0	0	0,03	0,00 €	
5		Armoiries d'Alger 15 c polychrome phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes des filets)	1232	1276	940	1960-01-02	Poste	0	0,19	0	0,0	0	0,05	0,00 €	
6		La semeuse de Piel d'après Roty 20 c truquoise et rose phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes des filets)	1233	1277	941	1960-01-02	Poste	0	0,06	0	0,0	0	0,03	0,00 €	
7		Marianne à la nef 25 c bleu et rouge phosphore : Sans 14 x 13½ 20 x 24 mm (vignette 18x22 bords externes des filets)	1234	1278	942	1960-01-02	Poste	0	0,55	0	0,0	0	0,03	0,00 €	
		La semeuse de Piel d'après Roty													

Génération d’un Fichier Excel côté pour l’Italie

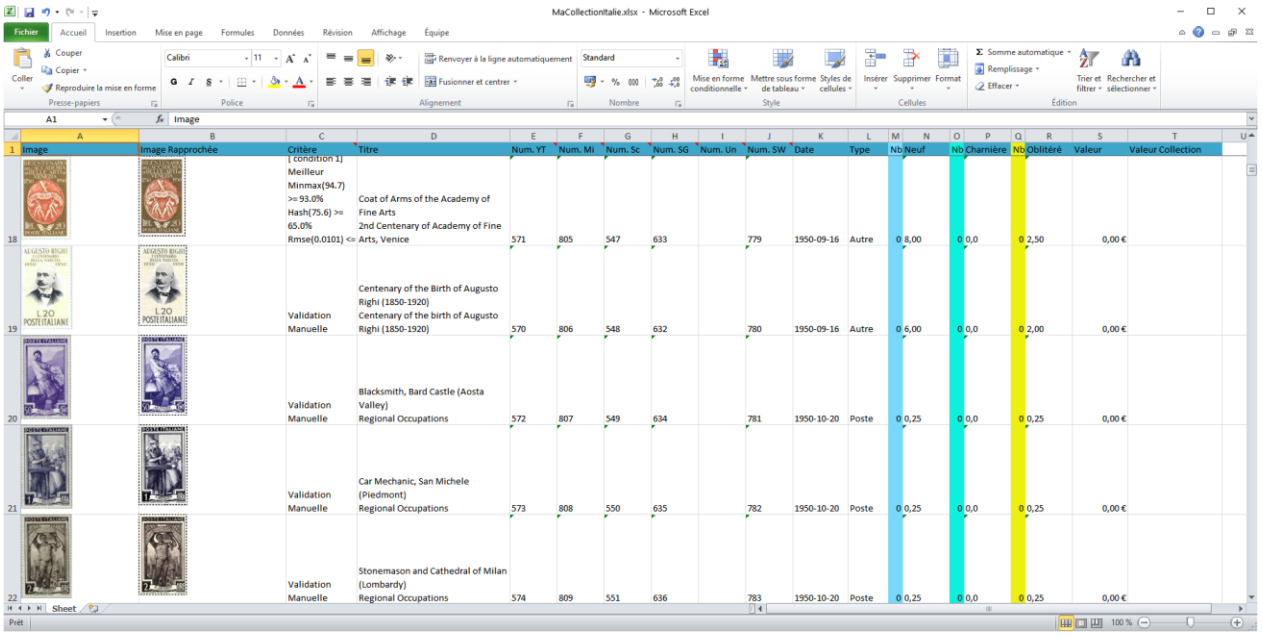


Image	Image Rapprochée	Critère	Titre	Num. YT	Num. MI	Num. Sc	Num. SG	Num. Un	Num. SW	Date	Type	Nb Neuf	Nb Charnière	Nb Oblitéré	Valeur	Valeur Collection
18		[condition 1] Meilleur Minmax(94.7) >= 93.0% Hash(75.6) >= 65.0% Rmse(0.0101) <=	Coat of Arms of the Academy of Fine Arts 2nd Centenary of Academy of Fine Arts, Venice	571	805	547	633		779	1950-09-16	Autre	0 8,00	0 0,0	0 2,50	0,00 €	
19		Validation Manuelle	Centenary of the Birth of Augusto Righi (1850-1920) Centenary of the birth of Augusto Righi (1850-1920)	570	806	548	632		780	1950-09-16	Autre	0 6,00	0 0,0	0 2,00	0,00 €	
20		Validation Manuelle	Blacksmith, Bard Castle (Aosta Valley) Regional Occupations	572	807	549	634		781	1950-10-20	Poste	0 0,25	0 0,0	0 0,25	0,00 €	
21		Validation Manuelle	Car Mechanic, San Michele (Piedmont) Regional Occupations	573	808	550	635		782	1950-10-20	Poste	0 0,25	0 0,0	0 0,25	0,00 €	
22		Validation Manuelle	Stonemason and Cathedral of Milan (Lombardy) Regional Occupations	574	809	551	636		783	1950-10-20	Poste	0 0,25	0 0,0	0 0,25	0,00 €	

Idem après saisie des timbres possédés

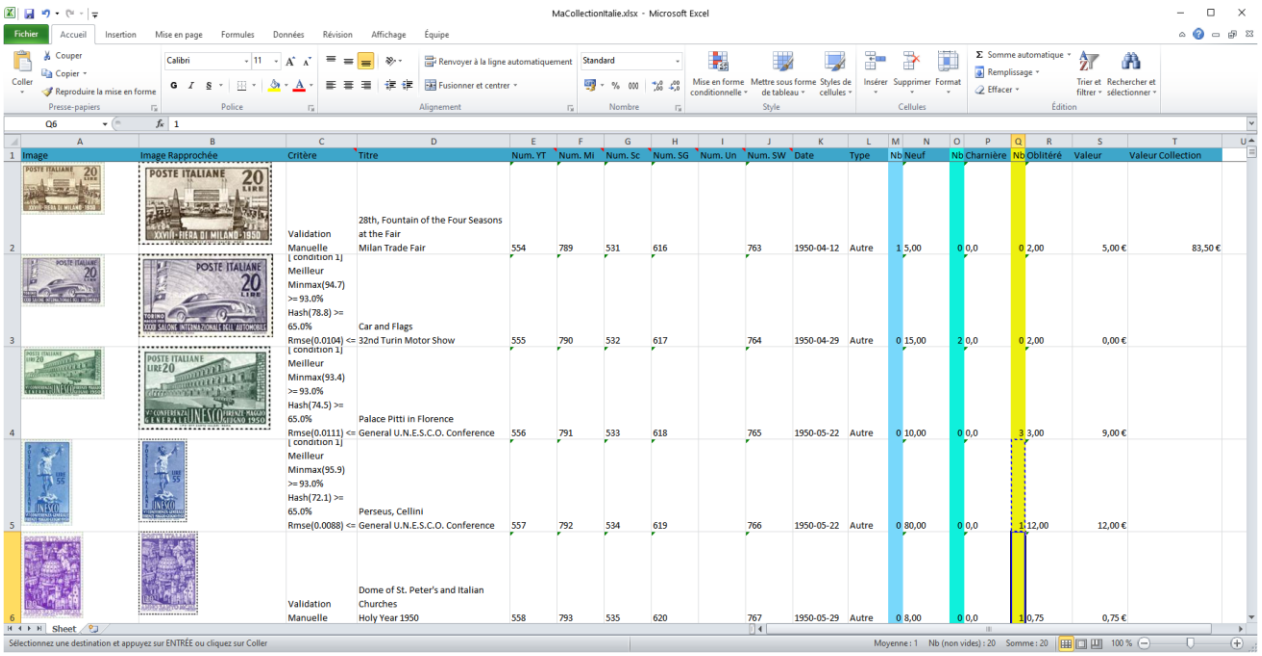


Image	Image Rapprochée	Critère	Titre	Num. YT	Num. MI	Num. Sc	Num. SG	Num. Un	Num. SW	Date	Type	Nb Neuf	Nb Charnière	Nb Oblitéré	Valeur	Valeur Collection
2		Validation Manuelle	28th, Fountain of the Four Seasons at the Fair Milan Trade Fair	554	789	531	616		763	1950-04-12	Autre	1 5,00	0 0,0	0 2,00	5,00 €	83,50 €
3		[condition 1] Meilleur Minmax(94.7) >= 93.0% Hash(78.8) >= 65.0% Rmse(0.0104) <=	Car and Flags 32nd Turin Motor Show	555	790	532	617		764	1950-04-29	Autre	0 15,00	2 0,0	0 2,00	0,00 €	
4		[condition 1] Meilleur Minmax(93.4) >= 93.0% Hash(74.5) >= 65.0% Rmse(0.0111) <=	Palace Pitti in Florence General U.N.E.S.C.O. Conference	556	791	533	618		765	1950-05-22	Autre	0 10,00	0 0,0	3 3,00	9,00 €	
5		[condition 1] Meilleur Minmax(95.9) >= 93.0% Hash(72.1) >= 65.0% Rmse(0.0088) <=	Perseus, Cellini General U.N.E.S.C.O. Conference	557	792	534	619		766	1950-05-22	Autre	0 80,00	0 0,0	1 12,00	12,00 €	
6		Validation Manuelle	Dome of St. Peter's and Italian Churches Holy Year 1950	558	793	535	620		767	1950-05-29	Autre	0 8,00	0 0,0	1 0,75	0,75 €	

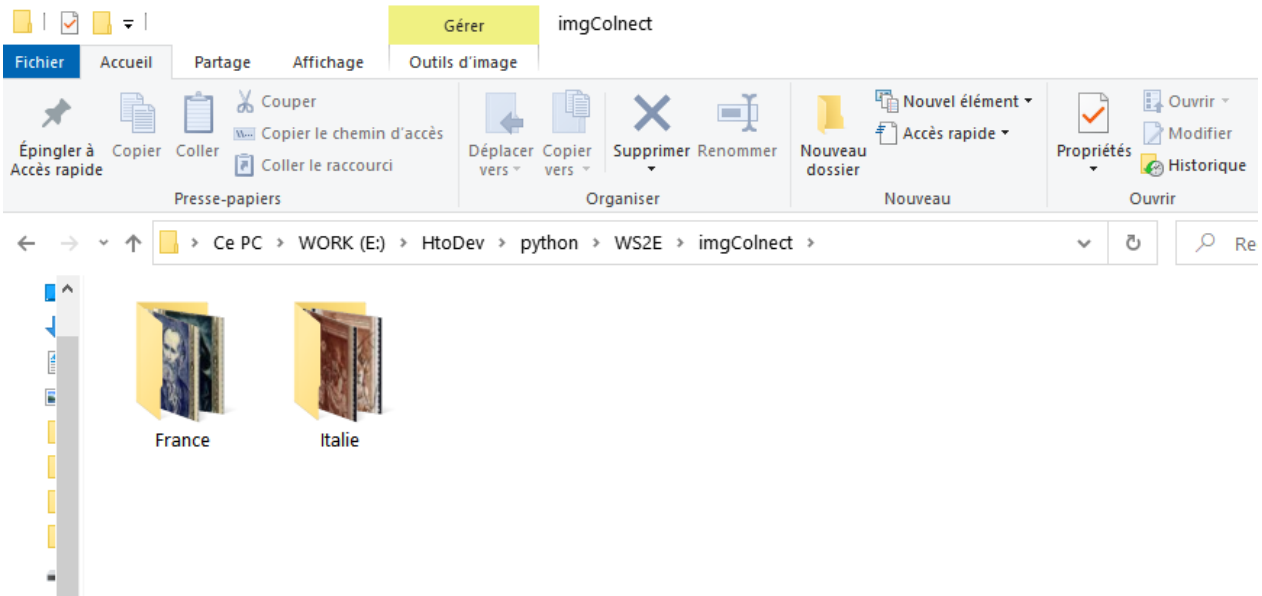
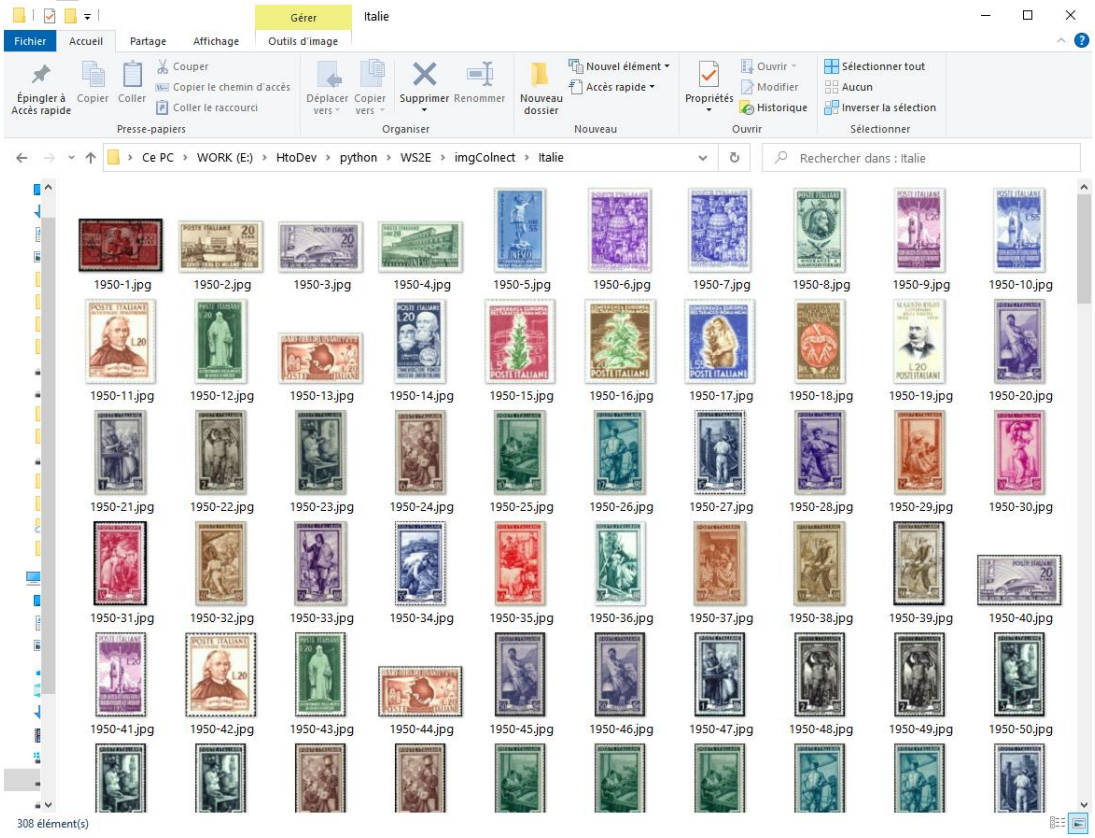
Mise à jour des cotations

La mise à jour des cotations permet de mettre à jour les cotations dans un fichier Excel courant déjà généré avec cotations.

Après chaque modification du fichier une copie datée du fichier précédente est générée.

Les « cotations protégées » par la retouche du fichier ne seront pas modifiées.

Extraction des images

Extraction uniquement des images de timbres des 2 sites respectifs par pays dans les répertoires imgColnect et imgStampWorld selon l'intervalle d'année spécifié.

Les extractions sont cumulatives car l'indexation se fait par année.