

Indexation des images numérisées des registres de matricules militaires

Recommandations et aide à la rédaction
des aspects techniques du cahier des charges

Février 2014

Cahier des clauses techniques particulières

Grille d'indexation

La prestation consistera en l'indexation des images issues de la numérisation des registres matricules en relevant 13 types d'information :

- 1. nom du soldat ;
- 2. prénoms ;
- 3. Année de naissance ;
- 4. commune de naissance ;
- 5. département de naissance ;
- 6. pays ou territoire de naissance (si autre que la métropole)¹ ;
- 7. profession ;
- 8. degré d'instruction générale ;
- 9. commune de résidence au moment de l'enregistrement ;
- 10. département de résidence au moment de l'enregistrement ;
- 11. bureau de recrutement ;
- 12. classe² ;
- 13. cote du registre.

Il sera demandé au prestataire de réaliser l'indexation des images numérisées des registres matricules en complétant la grille suivante :

cote du registre	classe	n° matricule ³	bureau de recrutement	nom	prénoms	année de naissance	commune de naissance	département de naissance	pays ou territoire de naissance	profession	degré d'instruction générale	commune de résidence	département de résidence	premier identifiant d'image	dernier identifiant d'image
1_r_2000	1895	432	Chambéry	Génevois	François Lucien	1875	Apremont	Savoie		agriculteur	3	Chambéry	Savoie	FRAD...	FRAD...

Les champs colorés en gris correspondent à des données qui peuvent être renseignées de façon automatique ou quasi-automatique par les archives départementales.

Il sera demandé au prestataire de prévoir une colonne supplémentaire, dédiée à un champ « alerte ». Lorsque le prestataire rencontrera une difficulté (exemples : noms sans matricules, informations illisibles, documents isolés et retombes décollées dont on ne sait plus à quel matricule ils appartiennent), il cochera simplement la case. Ceci permettra aux archives départementales de procéder au contrôle prioritaire des lignes d'indexation ainsi repérées.

Règles de saisie pour le traitement de l'indexation des champs

La pertinence des recherches futures dans les images numérisées des registres matricules dépend

¹ Ce champ permet de repérer les personnes nées à l'étranger ou originaires de l'Empire.

² Ce champ permet de repérer les éventuelles exceptions.

³ Ce champ ne constitue pas une information unique au niveau national et ne sera pas utilisée dans la base de recherches. Il constitue néanmoins une information utile pour l'exploitation locale et le recoupement avec la base des Morts pour la France.

étroitement de la qualité de l'indexation réalisée. Une information dont la saisie comporterait des erreurs ne pourrait en effet pas être retrouvée par le biais du moteur de recherche. Un objectif de qualité doit ainsi être recherché et atteint.

D'autre part, l'exploitation optimale des données d'indexation constituées par les départements au sein d'une base de recherche nationale repose sur la cohérence de ces données et l'harmonisation des règles de traitement. Les Archives de France ont donc édicté, pour chaque champ, des règles de saisie, qui devront être scrupuleusement suivies par le prestataire et contrôlées par les services d'archives départementales.

Les règles de saisie suivent les normes AFNOR NF Z 44-061 (juin 1986) « Documentation – Catalogage – Forme et structure des vedettes noms de personnes, des vedettes titres, des rubriques de classement et des titres forgés », 44-081 (septembre 1993) « Documentation – Catalogage des documents cartographiques – Forme et structure des vedettes noms géographiques » et NF Z44-070 (août 1986) « Documentation – Indexation analytique par matière », ainsi que les recommandations de la note d'information DITN/RES/2007/008 en date du 30 juillet 2007 (<http://www.archivesdefrance.culture.gouv.fr/static/1083>) et du « Guide pratique du catalogueur de la Bibliothèque nationale de France » (<http://guideducatalogueur.bnf.fr>).

– 1. nom du soldat.

La transcription se fera en lettres minuscules avec une majuscule initiale.

Tous les caractères, hormis les majuscules initiales, devront être si besoin accentués.
Exemples : « Albéritère ».

Pour les noms composés, les différentes parties du nom seront séparées par des espaces, même si ces parties ont été transcrites sur le matricule en étant séparées par des tirets courts.
Exemples : « Abel Duplan » ; « Bernard de Domsure ».

S'il y a une particule, celle-ci sera rejetée après le nom sans virgule. Exemples : « Aboville d' », « Abzac de Ladouze d' ».

S'il y a un surnom, celui-ci suivra le nom en étant précédé de « dit ». Exemple : « Leblanc dit Revel ».

– 2. prénoms ;

La transcription se fera en lettres minuscules avec une majuscule initiale.

Tous les caractères, hormis les majuscules initiales, devront être si besoin accentués.
Exemples : « Benoît ».

Pour les prénoms composés, les différents prénoms seront séparés par des espaces, même si ces prénoms ont été transcrits sur le matricule en étant séparés par des tirets courts. Cette règle s'applique aussi pour les prénoms de baptême qui ont pu être transcrits dans les documents.
Exemples : « Marie Joseph Victor », « Aimé Jean Baptiste ».

Si une mention de surnom existe, celle-ci sera saisie après les prénoms, en étant précédée de « dit ». Exemple : « Martial Joseph Charles dit Pierre ».

– 3. année de naissance ;

L'année de naissance sera saisie au format AAAA représentant l'année tout en chiffre, en conformité avec la norme ISO 8601 « Éléments de données et formats d'échange – Échange d'information – Représentation de la date et de l'heure ».

Exemple : « 1886 ».

– 4. commune de naissance ;

La transcription se fera en lettres minuscules en maintenant la majuscule initiale selon l'usage typographique français.

Tous les caractères, hormis les majuscules initiales, devront être si besoin accentués.

Lorsque le nom est composé de deux éléments séparés par un tiret, celui-ci est maintenu.

Exemple : « Saint-Affrique ».

Les formes aberrantes ne seront pas retenues dans l'index : s'il y a discordance entre l'usage ancien et l'usage actuel, c'est la forme la plus moderne qui devra être adoptée. On se référera à l'orthographe en usage aujourd'hui adoptée dans le *Code officiel géographique* de l'INSEE, repris dans le *Dictionnaire des communes* (<http://www.insee.fr/fr/methodes/nomenclatures/cog/>).

Exemples : « Rambervillers » et non pas « Rambervilliers », « Vathiménil » et non pas « Vathiménille », « Mirecourt » et non pas « Remirecourt », « Thaon-les-Vosges » et non pas « Thaon-lès-Vosges », « Saint-Dié-des-Vosges » et non pas « Saint-Dié ».

L'article initial est conservé lorsqu'il fait partie intégrante du toponyme (quand on ne peut employer le nom sans l'article).

Exemples : « La Mulatière », « Le Leuy », « Les Mayons ».

Pour les lieux ayant changé de nom, on citera à la fois le nom du lieu de l'époque et le nom actuel, qui sera précisé entre parenthèses.

Exemples : « Alfort (Maisons-Alfort) », « Altroff (Bettelainville) », « Saïgon (Hô-Chi-Minh Ville) ».

Pour les communes fusionnées, on citera à la fois le nom du lieu de l'époque et le nom de la commune actuelle, qui sera précisé entre parenthèses.

Exemple : « La Capelle (La Canourgue) ».

Si un lieu-dit est indiqué à la suite de la commune, il peut également être saisi. Il n'y a pas de forme normalisée à respecter et il suffit de transcrire exactement l'information du registre, entre parenthèses.

Exemple : Mende (Chabrits).

Les toponymes étrangers sont cités sous une forme française lorsque celle-ci est attestée. Exemples : « Aix-la-Chapelle » et non pas Aachen, « Florence » et non pas « Firenze ».

À défaut, le nom est cité sous sa forme nationale, en translittération quand la forme nationale est en caractères non latins.

Exemple : « Aïn Beida ».

– 5. département de naissance ;

La transcription se fera en lettres minuscules en maintenant la majuscule initiale selon l'usage typographique français.

Tous les caractères, hormis les majuscules initiales, devront être si besoin accentués.

On indiquera le nom du département actuel de rattachement.

Exemples : « Seine-Maritime » et non pas « Seine-Inférieure », « Val-d'Oise » et non pas « Seine-et-Oise », « Loire-Atlantique » et non pas « Loire-Inférieure ».

L'utilisation du code sur trois chiffres (001 pour l'Ain, 972 pour la Martinique) est également autorisée.

– **6. pays ou territoire de naissance (si autre que la métropole) ;**

La transcription se fera en lettres minuscules en maintenant la majuscule initiale selon l'usage typographique français.

Le nom saisi est celui du pays actuel.

Tous les caractères, hormis les majuscules initiales, devront être si besoin accentués.

Pour bénéficier des fonctionnalités de recherche avancée et de filtres dans les résultats, il est nécessaire que les pays indexés soient les pays actuels.

Les pays peuvent être indexés soit sous la forme de leur nom actuel (formes retenues dans le référentiel « Pays et villes du monde » du Conseil national de l'information géographique (CNIG)⁴, soit sous la forme de leur code ISO (norme ISO 3166 « Codes des noms de pays »)

Si, pour des raisons de coût ou de compétence, le prestataire qui indexe ne peut assurer cette qualité de données, il est fortement recommandé aux services d'archives de procéder à une vérification des données du champ « pays » ou « territoire de naissance » et à leur harmonisation.

Si aucune de ces opérations n'est possible, le pays « incorrect » sera affiché dans les résultats de recherche, mais il ne sera pas interrogeable, ni en recherche avancée ni en filtre.

– **7. profession ;**

La profession du conscrit sera retranscrite en minuscules, sans majuscule à la lettre initiale, et au singulier, tel qu'il apparaît dans le matricule, en respectant les règles actuelles d'orthographe.

Tous les caractères devront être si besoin accentués.

– **8. degré d'instruction générale ;**

Le degré d'instruction générale est inscrit dans les registres matricules sous forme de code, composé d'un chiffre entre 0 et 5.

Le champ sera donc renseigné par le chiffre tel qu'il figure sur le document numérisé.

Exemple : « 1 ».

Dans le cas où plusieurs niveaux sont renseignés, on ne saisira que le plus élevé. Par exemple : « 1, 2, 3 » ou « 1-2-3 » seront retranscrits « 3 ».

⁴ http://archives.cnig.gouv.fr/Front/docs/cms/cntpvm_enrichi_13avril2010_127540209260628000.pdf

– **9. commune de résidence au moment de son enregistrement ;**

Ce champ sera saisi selon les mêmes règles que celles du champ 4. commune de naissance.

– **10. département de résidence au moment de son enregistrement ;**

Ce champ sera renseigné de façon semi-automatique par les archives départementales. Attention, le ressort des régions militaires ne correspond pas au ressort des départements, ce champ ne comprend donc pas qu'une seule possibilité de nom de département.

Il sera saisi selon les mêmes règles que celles du champ 5. département de naissance.

– **11. bureau de recrutement ;**

Ce champ sera renseigné de façon semi-automatique par les archives départementales.

Seul le nom de la commune siège du bureau de recrutement sera renseigné, suivant les mêmes règles de saisie édictées pour le champ 4. commune de naissance.

Exemple : « Chambéry » pour le bureau de recrutement de Chambéry.

– **12. classe ;**

Ce champ sera renseigné de façon automatique par les archives départementales.

Seule l'année de la classe sera renseignée, sous la forme AAAA représentant l'année tout en chiffre, en conformité avec la norme ISO 8601 « Éléments de données et formats d'échange – Échange d'information – Représentation de la date et de l'heure ».

– **13. cote du registre.**

Ce champ sera renseigné de façon automatique par les archives départementales.

La cote sera saisie en séparant ses éléments par un tiret « _ » underscore. La lettre de la série R est saisie en minuscule.

Exemple : « 1_r_2000 ».

Préconisations pour les modalités de contrôle qualité.

Le contrôle qualité de l'indexation veillera à relever :

- la lisibilité du fichier d'indexation et sa conformité aux standards définis ;
- les champs manquants et non indexés ;
- le non respect des règles de saisie des champs ;
- les erreurs de transcription par rapport au document original.

Le contrôle sera effectué par échantillonnage, avec un seuil d'erreur de 2,5 % (soit 10 erreurs d'indexation pour un lot compris entre 3 200 et 10 000 fiches).

On se reportera au guide des Archives de France pour la définition de l'échantillon et du seuil d'erreur en fonction du nombre total d'items.