



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral des finances DFF
Centrale de compensation



UPI : situation actuelle

Groupe d'accompagnement
Harmonisation des Registres

Caractères spéciaux

31.01.2017

J.P. Naef, CdC

Rappels sur ce qu'est UPI :

- Essentiellement : un registre miroir des registres fédéraux (+ consolidation des sources)
- Focus : identification administrative fiable d'une personne physique.
- En général : >1 variantes d'identité par personne (pluralité des sources ; historique temporel)
- Hiérarchie des sources :
Infostar > SYMIC & Ordipro > VERA > AVS > autres annonceurs
→ une unique identité «de référence» à chaque instant

Caractéristiques techniques du système UPI :

- 100% compliant avec Catalogue des Caractères OFS :
ISO-8859-15 est pleinement géré (noms et prénoms)
 - en interne (base de données)
 - dans les interfaces (UPIServices + UPIViewer)
 - Usage de majuscules non accentuées :
 - Interfaces «historiques» pour les utilisateurs de l'AVS/AI
 - Logique interne de recherche nominative
- Existence d'une table de translittération interne.

UPI «souffre» des circonstances suivantes :

1. Manque de coordination interne aux registres harmonisés :

Migration SYMIC → UTF-8 (2014) : nouveaux caractères autorisés hors du set ISO-8859-15 (č, ć, ě, ř, š...).

→ SYMIC translittère à la volée ces nouveaux caractères dans ses annonces à UPI (sans quoi l'annonce est refusée).

→ Les utilisateurs qui consultent UPI ne comprennent pas pourquoi UPI n'est plus le reflet fidèle de SYMIC dans certains cas !

UPI «souffre» des circonstances suivantes :

2. Discipline parfois défailante dans la saisie des données dans les registres fédéraux : la graphie dicte la saisie !

Cas archétypal :

Nom	Minuscule	Majuscule	Origine typique	Prononciation	Translittération ?
eth	ð	Ð	islandais	comme le « th » dans les mots anglais them ou that	d (ou dh)
d barré	đ	Đ	serbe, slovaque, vietnamien	Dépend de l'origine: - serbe, slovaque: « dj » - vietnamien: proche du d	Dépend de l'origine - slave du sud: dj - vietnamien: dd
d hameçon	ḑ	Ḑ	diverses langues africaines	Consonne occlusive rétroflexe voisée	d

Đoković ≠ Ðoković !!!

ISO-8859-15

ISO-8859-15
ISO-8859-2

Glyphes identiques mais objets (caractères) différents !

Conséquences :

- translittération erronée
- échec recherche dans UPI

UPI «souffre» des circonstances suivantes :

3. Situation du Ð dans UPI au 31.12.2016 :

«eth» est utilisée dans le nom ou le prénom :

- de 41 personnes de nationalité issue d'un pays scandinave 😊
- de 702 personnes de nationalité issue
 - d'un pays d'ex-Yougoslavie 😞 (668)
 - du Vietnam 😞 (34)

→ Taux d'erreur de 94% dans l'utilisation de la lettre eth !

D'autres cas similaires existent :
Í vs. Í (Arial)
Í vs. Í (Times) ; ž vs. ž ; ...
qui sèment la confusion entre acteurs de l'harmonisation (ex :
dialogues de sourds entre service clientèle UPI et registre d'habitants,
basés sur ce que chacun « voit » dans son interface).

UPI «souffre» des circonstances suivantes :

4. Variantes culturelles de transcription des alphabets non-latins :

Boris Eltsine	[transcription française]
Boris Jeltzin	[deutsche Umschrift]
Boris El'cin	[trascrizione italiana]
Boris Yeltsin	[english transcription]

Борис Ельцин



Pour garder le NAVS13 de M. Eltsine unique, UPI devrait être capable de reconnaître automatiquement que
Eltsine = Jeltzin = El'cin = Yeltsin !

Aucune algorithmique «simple» ne permet cela avec 100% d'efficacité, et nous n'avons pas la puissance de Google pour «résoudre» ces divergences par force brute (deep learning etc.) !

Souhaits / recommandations

Pour moins de problèmes (politiques, techniques...), au minimum :

- Un jeu de caractères identique et suffisant pris en charge par tous les systèmes (Infostar, SYMIC, UPI, reg. habitants, etc.)
- Une spécification suffisante et harmonisée entre institutions des règles
 - de saisie (basées sur caractères et non glyphes)
 - de translittération alphabétique
 - de transcription phonétique (alphabets non latins)
 - de traitement de telles opérations effectuées par des tiers (traducteurs dans des pays étrangers)
- Une meilleure sensibilisation et formation des acteurs chargés de la saisie des données dans les systèmes.