

# Interopérabilité des données géolinguistiques à l'exemple du projet VerbaAlpina

## 1. Les changements dans la façon de penser la recherche

Les modalités de recherche ont beaucoup changé dans les dernières décennies. Autrefois, chaque recherche représentait un projet conçu, développé et achevé en lui-même. En employant une métaphore, on pourrait imaginer chaque projet comme une île d'un immense archipel. Chaque projet se distinguait ainsi des autres. Malgré le fait que, même dans le passé, les connaissances se construisaient toujours sur des notions déjà acquises, on ne profitait pas directement des efforts réalisés par d'autres chercheurs, ou groupes de chercheurs, en termes de moyens et d'infrastructure de la recherche. En général on avait tendance à exploiter plutôt les connaissances, et pas les efforts accomplis pour rejoindre ce savoir du point de vue des structures, méthodes et outils. Ainsi, dans le cadre de la recherche, la notion de partage dans ce sens n'était pas encore vraiment développée comme elle peut l'être aujourd'hui. De plus, la communication scientifique s'effectuait seulement à travers la publication des livres sur papier, lesquels étaient conservés dans des lieux spécifiques comme les bibliothèques, ce qui, d'une certaine façon, ne facilitait pas l'accès aux données ; souvent celui-ci n'était même pas prévu<sup>1</sup>.

---

1 — Cf. T. KREFELD, « Communication scientifique ».

L'émergence des possibilités fournies par une numérisation de plus en plus rapide et le fait que les financements de la recherche restent limités ont pour conséquence le besoin d'une nouvelle approche dans la conception et dans l'organisation de la recherche.

La numérisation offre beaucoup de moyens par rapport aux instruments et aux moyens de présenter les contenus : le monde numérique permet un détachement du seul papier en intégrant d'autres types de médias pour la communication scientifique, comme, par exemple, la possibilité d'inclure des fichiers audio ou des liens vers d'autres ressources. Dans une perspective linguistique (et dans le cadre de *Verba Alpina*,) surtout en ce qui concerne les atlas linguistiques et les dictionnaires, le numérique permet un accès aux données qui est bidirectionnel, à savoir onomasiologique et sémasiologique au sein du même projet ou de la même œuvre, tandis que les publications sur papier ne permettaient qu'une seule perspective à la fois<sup>2</sup>. La numérisation offre les moyens pour une recherche vraiment « collective » et permet de mener des projets dans une optique de travail en réseau. De plus, grâce aux nouvelles technologies, la publication des résultats est simplifiée, ce qui permet à tous ceux qui s'y intéressent de pouvoir y accéder plus facilement. Aujourd'hui l'attention n'est donc pas seulement posée sur la conduite de la recherche, mais une importance relativement grande est aussi accordée à son accès dans une optique de grande diffusion du savoir. Néanmoins, entrer dans cet esprit de recherche ne signifie pas seulement appliquer des technologies numériques au travail scientifique. Il s'agit beaucoup plus de repenser les modalités de recherche : la numérisation peut ouvrir de grands horizons du point de vue du dynamisme, des automatismes et de la vitesse de certaines opérations, ce qui permet aussi de profiter au maximum de chaque investissement fait dans le monde académique. Toutefois, les possibilités offertes par le numérique doivent aussi être administrées : le défi principal dans ce sens est représenté par la gestion des données de la recherche de manière durable afin de pouvoir profiter le plus longtemps possible des connaissances acquises. Pour essayer de donner une réponse à la question de la gestion des données, un groupe de chercheurs du monde entier a dressé quatre principes servant comme point d'orientation dans le débat concernant le traitement des données, la communication et la coopération dans l'ère des médias<sup>3</sup>. Ces

2 — Cf. S. LÜCKE, « Einführung in die Geolinguistik ».

3 — Cf. FORCE11, « The Fair Data Principles » ; cf. M. WILKINSON, « The Fair Guiding Principles ».

quatre principes se résument dans l'acronyme anglais « FAIR ». Les données de la recherche devraient ainsi être :

- *findable* « (re)trouvables »
- *accessible* « accessibles »
- *interoperable* « interopérables »
- *reusable* « réutilisables »

Les trois premiers principes s'insèrent dans l'idée d'une création des connaissances qui soient à la fois interprétables par les humains et lisibles par les ordinateurs (*human readable* et *machine readable*), tandis que le dernier principe fait référence à une réutilisation concernant les ordinateurs<sup>4</sup>.

Afin que les données soient retrouvables, il faudrait qu'elles soient pourvues de métadonnées (à savoir de données décrivant les données primaires, une sorte d'attributs) et que chaque donnée puisse être individuée à travers un identifiant unique et pérenne<sup>5</sup>. De plus, les bibliothèques pourraient jouer un rôle essentiel dans la mise en pratique de la repérabilité : il y aurait la possibilité de retrouver les données de la recherche de façon simple à travers les catalogues des bibliothèques<sup>6</sup>.

Les données accessibles sont compréhensibles à la fois par les humains et par les machines. De plus, il s'agit de renoncer aux droits d'auteur et d'appliquer aux données des licences qui suivent les règles du libre accès (*open access*).

Le troisième principe est celui de l'interopérabilité : il serait souhaitable que les données soient structurées de sorte qu'elles puissent être compatibles avec d'autres bases de données, afin de permettre leur interconnexion<sup>7</sup>. Il est question de permettre à la fois aux humains et aux machines de comprendre et d'interpréter les données de la recherche. Dans ce but, d'un côté il est nécessaire d'employer un vocabulaire et des ontologies communément utilisées et contrôlées, et, de l'autre côté, il faut bien structurer le modèle des données et des métadonnées. L'interopérabilité est garantie lorsque les données ou les outils de la recherche venant de ressources différentes sont capables de s'intégrer et de fonctionner ensemble. La pratique de l'interopérabilité se fait, au début, en créant des données bien décrites et organisées, toujours dans l'optique du partage et de la possibilité d'une exportation des données vers d'autres projets.

---

4 — Cf. T. Krefeld, « 54 Monate VerbaAlpina », p. 140.

5 — Cf. GoFAIR, « FAIR Principles ».

6 — Cf. LÜCKE, « Principes FAIR ».

7 — Cf. GoFAIR, « FAIR Principles ».

La réutilisation des données de la recherche résulte de l'application des trois premiers points. Pour faire en sorte que les données et les métadonnées soient réutilisables, elles devraient être décrites de façon à pouvoir être reproduites et/ou combinées dans des cadres différents<sup>8</sup>.

## 2. VerbaAlpina : étudier la complexité linguistique des Alpes

VerbaAlpina est un projet de recherche de l'Université Ludwig-Maximilian de Munich qui œuvre à étudier l'espace culturel alpin à la lumière de son plurilinguisme<sup>9</sup>. Le domaine de recherche de VerbaAlpina est limité aux frontières territoriales définies par la Convention Alpine<sup>10</sup> et concerne huit pays différents, à savoir la France, la principauté de Monaco, la Suisse, le Liechtenstein, l'Allemagne, l'Autriche, l'Italie et la Slovénie. D'un point de vue ethnographique et topographique, l'aire alpine présente une homogénéité constante dans l'ensemble du territoire – de la France à la Slovénie : l'élevage, la production laitière et fromagère, mais aussi la flore et la faune sont communes à toute la région montagnarde. Toutefois, en observant la composition linguistique des Alpes, une grande hétérogénéité se profile. Cette pluralité est donnée non seulement par le nombre de langues nationales standardisées (le français, l'italien, l'allemand, le romanche et le slovène) parlées dans le territoire, mais surtout par les différents parlers locaux, très bien conservés dans une grande partie de la région. L'objectif de VerbaAlpina est de montrer la complexité linguistique de l'espace alpin en mettant en exergue les points communs entre les différentes variétés et les familles linguistiques dans une perspective d'analyse lexicographique. VerbaAlpina a débuté en 2014 et est divisé en phases : dans la première phase du projet (octobre 2014 – octobre 2017), l'accent a été mis sur le lexique concernant la culture de l'alpage et le traitement du lait. Pendant la deuxième phase (novembre 2017 – octobre 2020), le projet se focalise sur le lexique de la nature, des formations des paysages, sur le temps météorologique, sur la faune et sur la flore. Dans une éventuelle troisième phase (novembre 2020 – octobre 2023), VerbaAlpina se penchera sur le lexique de la vie moderne, sur l'écologie et sur le tourisme, ainsi que sur la comparaison avec d'autres aires montagnardes.

---

8 — Cf. GoFAIR, « FAIR Principles ».

9 — Cf. KREFELD, « VerbaAlpina ».

10 — Cf. CONVENTION ALPINE.

années du projet	1	2	3	4	5	6	7	8	9	
année civile	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
trimestre	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV	I, II, III, IV
phase du projet	I			II			III			
axe principal	<b>culture</b> <ul style="list-style-type: none"> <li>• alpage</li> <li>• traitement du lait</li> </ul>			<b>nature</b> <ul style="list-style-type: none"> <li>• formations des paysages</li> <li>• temps</li> <li>• faune</li> <li>• flore</li> </ul>			<b>vie moderne</b> <ul style="list-style-type: none"> <li>• écologie</li> <li>• tourisme</li> <li>• autres régions montagnardes</li> </ul>			

**Fig. 1** : Résumé du plan et de l'organisation du projet VerbaAlpina

Les données traitées par VerbaAlpina sont toujours géoréférencées et ont une double provenance : d'un côté elles viennent des atlas linguistiques et des dictionnaires analogues ou digitaux concernant les variétés alpines ; de l'autre côté, dans le but d'intégrer, compléter et éventuellement corriger les données venant des atlas et des dictionnaires, VerbaAlpina a développé un outil de *crowd-sourcing* à travers une plateforme en ligne qui s'adresse aux locuteurs des dialectes alpins. Le grand public est invité à participer à la récolte des données en envoyant des mots dialectaux concernant les divers concepts de la réalité alpine<sup>11</sup>.

### 3. L'interopérabilité au sein de VerbaAlpina

Dès sa naissance en 2014, VerbaAlpina a été conçu dans une optique de partage et d'interconnexion, avant que les principes FAIR soient concrétisés. Cela signifie que VerbaAlpina a toujours fait l'effort de ne pas penser le projet comme une fin en soi, mais de l'insérer dans un cadre plus large, non seulement en s'appropriant des méthodes innovantes offertes par le numérique, mais surtout en les développant davantage<sup>12</sup>.

Pour VerbaAlpina, en suivant les principes FAIR, l'interopérabilité est la capacité des bases de données à travailler ensemble et à se référer les unes aux autres. Cette conception fait en sorte que la création de données structurées et de liens persistants entre bases de données (mais aussi entre projets), dans le but de permettre une interconnexion, constitue l'un des défis les plus importants pour VerbaAlpina. L'interopérabilité est ainsi également la capacité

11 — Cf. KREFELD, « VerbaAlpina », [https://www.verba-alpina.gwi.uni-muenchen.de/en/?page\\_id=1741&db=192](https://www.verba-alpina.gwi.uni-muenchen.de/en/?page_id=1741&db=192).

12 — La programmation de la plateforme de VerbaAlpina a été conçue et mise en œuvre par les informaticiens du projet, David Englmeier et Florian Zacherl.

d'un système à œuvrer collectivement avec d'autres systèmes (et produits) et à communiquer dans différentes situations. L'objectif est finalement celui de la coopération, de l'interaction et de l'échange d'informations ou de services entre différentes bases de données de différents projets.

La première étape pour une pratique conséquente de l'interopérabilité est la stricte distinction entre les données primaires, à savoir les données linguistiques venant des atlas, des dictionnaires ou du *crowdsourcing*, et les métadonnées<sup>13</sup>, c'est-à-dire les données décrivant les données primaires afin de les identifier de façon univoque<sup>14</sup>.

L'interopérabilité au sein de VerbaAlpina se pratique dans cinq domaines concernant le traitement structuré des données, l'enrichissement avec des métadonnées, l'attribution des notices d'autorité, le référencement univoque des données individuelles et l'accès aux données primaires et aux métadonnées. Ces cinq nœuds sont liés entre eux et contribuent, dans l'ensemble, à faire de VerbaAlpina un projet qui répond aux exigences de la recherche moderne.

### *3.1 Traitement structuré des données et enrichissement avec des métadonnées*

Afin de pouvoir développer le projet dans le cadre de l'interopérabilité et étant donné que VerbaAlpina prend en compte des données provenant de multiples sources, le premier défi est représenté par l'uniformisation de ces données à travers leur transcription<sup>15</sup>. Puisque VerbaAlpina travaille avec des données qui viennent de traditions disciplinaires hétérogènes (études romanes, germaniques et slaves) et qui, par conséquent, suivent des traditions de recherche différentes et existent en différentes langues, la première étape du traitement des données est constituée par le processus de transcription. Ce premier travail vise à permettre la comparaison entre les données tout en restant fidèles à la source.

#### *3.1.1 Le système de transcription « Beta Code »*

Au sein de VerbaAlpina, la saisie du matériel linguistique dans la base de données se fait par un système de transcription basé uniquement sur les caractères ASCII. La transformation des attestations linguistiques est donc possible avec n'importe quel clavier et à tout moment. Le système utilisé par VerbaAlpina, en référence à la terminologie

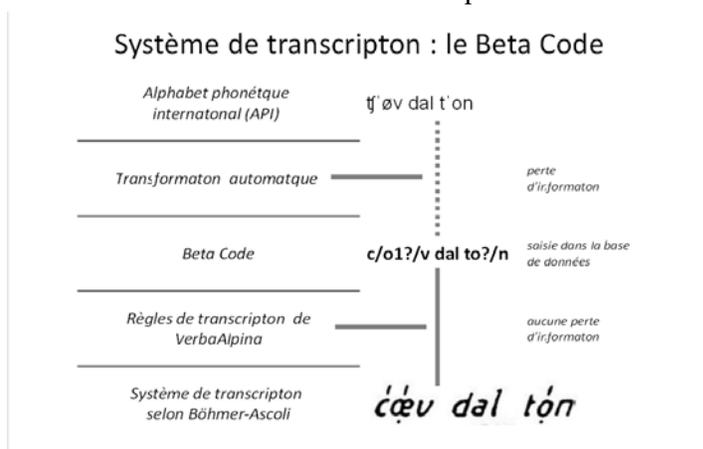
---

13 — Il convient de souligner que dans la communauté scientifique, il ne s'est pas encore établi de schéma de métadonnées standardisé. VerbaAlpina a donc opté pour un concept d'interface flexible (cf. point 3.1.6) qui permet d'utiliser pratiquement n'importe quel schéma de métadonnées.

14 — Cf. S. LÜCKE, « Métadonnées ».

15 — Cf. T. KREFELD, « Transcription ».

utilisée par le *Thesaurus Linguae Graecae*<sup>16</sup>, a été nommé *Beta Code*<sup>17</sup>. Dans ce système, à chaque caractère présent dans les atlas et dans les dictionnaires de référence de VerbaAlpina correspond une combinaison de caractères du clavier. La différenciation entre les caractères de base (c'est-à-dire les lettres) et les diacritiques marqués au-dessus et en dessous est très importante. D'abord, il s'agit de transcrire le caractère de base et après lui, tous les diacritiques connectés du bas à gauche vers le haut à droite. Par la suite, la séquence de caractères est saisie dans la base de données relationnelles. Pour réaliser l'uniformisation nécessaire, des listes dans la base de données sont créées afin que chaque caractère issu d'une source corresponde à un caractère de l'Alphabet Phonétique International (API)<sup>18</sup>. Grâce aux règles de transcription, le passage du papier au numérique ne souffre aucune perte d'information. Cependant, la transformation en API peut éventuellement comporter une perte, puisque la transcription phonétique selon les systèmes Böhmer-Ascoli ou Teuthonista<sup>19</sup>, à partir desquels est transcrite la majorité des atlas pris en considération par VerbaAlpina, prévoit, par exemple, des degrés d'ouverture des voyelles qui ne sont pas prévus dans l'API. Une fois les attestations transcrites, elles sont fragmentées en mots individuels à travers le processus de *tokenisation*.



**Fig. 2 :** Fonctionnement du système de transcription de VerbaAlpina (lld. *ciöf dal ton* = rhododendron)

16 — Cf. T. F. BRUNNER, « *Thesaurus Linguae Graecae* ».

17 — Cf. T. KREFELD, « *Beta Code* ».

18 — Cf. S. LÜCKE, « *Page de code* ».

19 — Le système Böhmer-Ascoli trouve son application surtout dans la dialectologie romane, tandis que le système Teuthonista est utilisé pour les dialectes allemands ; cf. E. BÖHMER, « *De sonis* » ; A. ASCOLI, « *Saggi Ladini* », p. XLII-XLVIII ; P. WIESINGER, « *Das phonetische Transkriptionssystem* ».

### 3.1.2 La typisation

L'interopérabilité au sein de VerbaAlpina se pratique aussi dans la classification des mots en types. La typisation fait partie des activités centrales de VerbaAlpina, dont l'un des buts est celui de montrer les zones de cohésion entre les différentes langues et les différentes familles linguistiques des Alpes<sup>20</sup>. Pendant le processus de typisation, les mots individuels sont regroupés en types lexicaux et en types de base, et sont décrits à travers l'application de métadonnées concernant la famille linguistique d'appartenance, la catégorie et le genre grammaticaux, ainsi que la racine lexicale. Le type de base regroupe toutes les attestations qui peuvent aussi appartenir à des familles linguistiques différentes, mais qui ont la même racine commune. Sous un même type morpho-lexical sont regroupées les différentes formes d'un seul type de base appartenant à une seule famille linguistique. Pour être regroupées sous un même type morpho-lexical, les attestations doivent présenter des caractéristiques grammaticales fondamentales en commun comme la catégorie grammaticale, le genre et les éléments de formation des mots.

ATTESTATION	TYPE MORPHO-LEXICAL
roa. <i>frumagi</i>	roa. <i>fromage</i>
roa. <i>fromagio</i>	
roa. <i>fruomag</i>	

ATTESTATION	TYPE DE BASE
deu. <i>Butter</i>	gr.-lat. <i>butyru(m)</i>
fra. <i>beurre</i>	
ita. <i>burro</i>	

Tab. 1 et 2 : Schématisation du processus de typisation de VerbaAlpina<sup>21</sup>

### 3.1.3 Enrichissement avec données sociolinguistiques et sociodémographiques

La description des données primaires initiée pendant la typisation continue, à travers leur enrichissement par des informations sociolinguistiques et sociodémographiques dans la base des données. Parmi les données décrivant les données primaires, VerbaAlpina dispose toujours des informations concernant la

20 — Cf. T. KREFELD, « Réduction à types ».

21 — Les abréviations pour les langues sont indiquées selon le système ISO 639-3 (cf. SIL INTERNATIONAL).

langue et la commune d'appartenance de l'informateur. De plus, selon la disponibilité, des informations concernant, par exemple, la minorité linguistique d'appartenance, l'âge de l'informateur et le sexe sont ajoutés. Cet enrichissement avec des informations supplémentaires est fait dans le but de créer une base de données riche et bien décrite.

### 3.1.4 Attribution des notices d'autorité

Afin que les données puissent être liées et mises en relation non seulement entre elles, mais aussi avec des données d'autres projets, VerbaAlpina applique des notices d'autorité<sup>22</sup> aux données primaires, ce qui permet une identification précise et finement granulée des données de recherche. Pour ce faire, aux données primaires sont assignés certains identifiants. Une distinction est faite entre les identifiants générés par VerbaAlpina, c'est-à-dire les notices d'autorité spécifiques de VerbaAlpina, et les identifiants des instances externes.

Dans le cadre du projet, VerbaAlpina distingue essentiellement trois catégories de données auxquelles des identifiants individuels sont attribués. Ces identifiants peuvent être récupérés avec les données. Ces trois catégories de données sont les suivantes :

- types morpho-lexicaux
- concepts
- communes politiques.

Chaque objet numérique de ces trois catégories est doté d'un identifiant unique et persistant. La lettre L est réservée aux types morpho-lexicaux, C aux concepts, A aux communautés, comme dans les exemples qui suivent :

- L1435, « babeurre (m.) (roa.) »
- C1777, « PIN CEMBRO »
- A53891, « Montauroux »

En outre, VerbaAlpina intègre actuellement les identifiants suivants provenant des instances externes qui se sont également engagées à respecter les critères FAIR et notamment l'interopérabilité :

- *Gemeinsame Normdatei* (GND, « norme standard commune ») de la Bibliothèque Nationale Allemande
- Q- et L-IDs de Wikidata
- GeoNames de la page web [www.geonames.org](http://www.geonames.org)

---

<sup>22</sup> — Cf. WIKIPEDIA, « Autorité ».

- codages des langues selon la norme ISO 639-3
- divers dictionnaires de référence (actuellement 43 titres, dont Vocabolario Treccani, FEW en ligne, DRG (Dicziunari Rumantsch Grischun), Georges 1913 [1998], Duden, etc.)

L'identification des concepts est faite en utilisant les identifiants Q-IDs de Wikidata, mais également, si elles sont disponibles, en appliquant les GNDs conçues par la Bibliothèque Nationale Allemande. Par exemple, le concept PIN CEMBRO a le Q-ID 147278<sup>23</sup> et le GND 4068873-2<sup>24</sup>. En ce qui concerne les types morpho-lexicaux, VerbaAlpina est en train d'assigner les identifiants L-IDs<sup>25</sup> de Wikidata. Les Q-IDs et les GNDs permettent le référencement univoque des objets et des entités extralinguistiques, qui peuvent être des objets, des collectivités, des associations, etc. Les L-IDs, en revanche, permettent le référencement unique des lexèmes. Pour l'identification des communes nous utilisons ce qu'on appelle GeoNames de la page web <https://www.geonames.org/>. Selon ce standard, le lieu « Montauroux » a le numéro d'identification 6613115<sup>26</sup>. En outre, les différents types morpho-lexicaux et les types de base qui sont attribués à une attestation sont dotés de codes de langue de la norme ISO 639-3. Les types morpho-lexicaux sont également liés aux entrées correspondantes des divers dictionnaires de référence. Le graphique suivant montre comment les identifiants individuels sont liés à une attestation sur la carte interactive de VerbaAlpina en utilisant l'exemple d'une attestation pour le concept LAIT.



**Fig. 3** : Visualisation des informations concernant une attestation du concept LAIT sur la carte interactive de VerbaAlpina

23 — Cf. <https://www.wikidata.org/wiki/Q147278>.

24 — Cf. <http://d-nb.info/gnd/4068873-2>.

25 — Cf. WIKIDATA, « Lexicographical data ».

26 — Cf. <https://www.geonames.org/6613115/montauroux.html>.

Grâce à ces identificateurs, les données lexicales peuvent, dans le monde entier, être reliées sans ambiguïté les unes aux autres dans des bases de données séparées et indépendamment de la langue à laquelle elles appartiennent.

### 3.1.5 *Référencement univoque aux données individuelles*

VerbaAlpina pratique également l'interopérabilité à travers le référencement univoque aux données individuelles. Pour pouvoir référer directement à un objet, chaque donnée (attestations individuelles, types morpho-lexicaux, concepts et municipalités) dispose d'une URL (*Uniform Resource Locator*) individuelle. Chaque URL contient le numéro de la version de VerbaAlpina comme paramètre URL. Une nouvelle version de l'ensemble de données de VerbaAlpina est créée tous les 6 mois.

Dans une optique de durabilité, à chaque URL est assigné aussi un DOI<sup>27</sup> (*Digital Object Identifier*, « identificateur d'objet numérique ») par la bibliothèque de l'Université de Munich. Un *Digital Object Identifier* est une adresse unique et invariable à l'échelle mondiale à travers laquelle on peut accéder aux ressources électroniques. L'accessibilité est également garantie si, par exemple, l'URL d'une ressource change. Le principal avantage du système DOI est donc la citation durable des ressources électroniques. En même temps, cela permet que toutes les données puissent être trouvées facilement. De même, il est possible d'inclure entièrement VerbaAlpina dans les catalogues des bibliothèques<sup>28</sup>. En plus des attestations individuelles, des types morpho-lexicaux, des concepts et des municipalités, les articles thématiques publiés sur le site web de VerbaAlpina sous les onglets « Lexicon alpinum », « Méthodologie » et « Contributions » sont également dotés d'un DOI et peuvent donc être cités directement<sup>29</sup>.

### 3.1.6 *Accès aux données primaires et aux métadonnées*

Un autre aspect de l'interopérabilité au sein de VerbaAlpina est l'accès aux données primaires et aux métadonnées. D'une part, l'accès aux données est possible via la carte interactive<sup>30</sup> et la section « Lexicon Alpinum »<sup>31</sup> sur la page web de VerbaAlpina. La carte interactive offre à la fois une fonction de recherche générale et divers filtres (concepts, types morpho-lexicaux, types de base,

27 — Cf. S. LÜCKE, « Digital Object Identifier (DOI) ».

28 — Cf. [http://dbis.uni-regensburg.de/frontdoor.php?titel\\_id=101965&bib\\_id=ub\\_m](http://dbis.uni-regensburg.de/frontdoor.php?titel_id=101965&bib_id=ub_m).

29 — Cf. T. KREFELD, « butyru(m) ».

30 — Cf. [https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page\\_id=18&noeredirect=fr\\_FR&db=192](https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=18&noeredirect=fr_FR&db=192).

31 — Cf. [https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page\\_id=84&db=192](https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=84&db=192).

informateurs, données complémentaires, surfaces et régions) qui permettent l'accès aux données collectées par VerbaAlpina. Dans le « Lexicon Alpinum », les différents concepts et types morpho-lexicaux sont listés avec un commentaire sur l'étymologie des termes. De l'autre côté, les données sont également accessibles via une interface de programmation d'application (API, *application programming interface*). L'API permet de consulter des contenus spécifiques de la base de données VA (VA\_DB) dans des formats définis à travers un navigateur web. La sélection des données et le format de sortie sont contrôlés par des paramètres URL. L'API de VerbaAlpina est disponible à cette adresse :

<https://www.verba-alpina.gwi.uni-muenchen.de/?api=1>. La documentation de la syntaxe à utiliser se trouve à l'adresse suivante : [https://www.verba-alpina.gwi.uni-muenchen.de/fr/?-page\\_id=128&noredirect=fr\\_FR&db=191](https://www.verba-alpina.gwi.uni-muenchen.de/fr/?-page_id=128&noredirect=fr_FR&db=191). Les APIs sont spécifiquement conçues pour un accès machine sans navigateur.

#### 4. Les autres principes FAIR au sein de VerbaAlpina : Findable - Accessible - Reusable

Après avoir décrit en détail au chapitre 3 comment l'interopérabilité est assurée au sein de VerbaAlpina, pour conclure, nous décrirons comment VerbaAlpina essaie de se conformer aux trois autres principes FAIR : *findable*, *accessible*, *reusable*. En général, il convient toutefois de noter d'abord que les différents principes FAIR ne peuvent pas toujours être clairement séparés les uns des autres, car ils sont fonctionnellement entrelacés et se chevauchent donc partiellement dans leur contenu. Dans la pratique, cependant, VerbaAlpina fait référence à la définition des critères FAIR formulés par FORCE 11<sup>32</sup>.

En ce qui concerne la *findability* des données, celle-ci est, d'une part, assurée par des propriétés qui rendent les données en même temps interopérables, et qui ont déjà été mentionnées au chapitre 3. Cela comprend également la mise en relation des données avec des identifiants (IDs) aussi finement granulés et uniques que possible, ainsi qu'avec des notices d'autorité et des métadonnées (voir chapitre 3.1.4.), afin que les données puissent être identifiées sur l'Internet. En outre, pour rendre *findable* les données collectées et traitées, VerbaAlpina entretient une coopération avec deux projets de recherche actuellement en cours, lesquels abordent également le problème de l'absence d'un schéma de

---

32 — Cf. FORCE11, « The Fair Data Principles ».

métadonnées standard uniforme et contraignant dans le monde scientifique (mentionné au chapitre 3.1.2.). Les deux projets sont l'initiative GeRDI<sup>33</sup> (*Generic Research Data Infrastructure*) du LRZ (*Leibniz-Rechenzentrum*, « Centre informatique de Leibniz ») et le projet « eHumanities - interdisziplinär »<sup>34</sup> de la bibliothèque universitaire de l'Université de Munich. Dans le cadre de ces coopérations, le stock de données centrales est enrichi avec des métadonnées version par version. Puis les données sont transférées sous plusieurs formes à la bibliothèque universitaire de la LMU, où elles sont stockées dans le référentiel Open Data. Ensuite, les métadonnées sont intégrées à l'index, qui est en cours de création dans le cadre du projet GeRDI. De cette façon les données sont accessibles de manière centrale via le catalogue de la bibliothèque universitaire ainsi que via le portail de recherche du projet GeRDI, lequel est encore en cours de développement.

Afin de rendre les données *accessibles*, nous utilisons des licences *Creative Commons* (CC)<sup>35</sup> compatibles avec *open access* pour toutes les données créées par VerbaAlpina. Jusqu'à la version 18/1, nous avons utilisé une licence CC BY SA 3.0 ; ensuite, avec la version 18/2 nous sommes passés à la licence CC BY SA 4.0.

Comme nous l'avons mentionné au début de cet article, le principe de réutilisation des données est garanti seulement si les trois premiers principes (*findable, accessible, interoperable*) sont respectés. Au cours de cette contribution, nous avons montré comment VerbaAlpina travaille dans le respect de ces trois principes en rendant, par conséquent, les données du projet tout à fait réutilisables.

## 5. Conclusion

À l'exemple du projet géolinguistique VerbaAlpina, cette contribution voulait montrer que l'interopérabilité des données de la recherche, ainsi que l'orientation générale des activités de recherche vers les principes FAIR, apportent des avantages décisifs concernant surtout l'interconnexion et la réutilisation permanente des données de la recherche provenant de disciplines et de domaines différents. La valeur ajoutée d'une orientation uniforme des projets de recherche au niveau international vers les principes FAIR serait donc que les projets individuels pourraient être fusionnés dans un stock de données qui ne cesse de croître et qui

---

33 — Cf. <https://www.gerdi-project.eu/communitieFzs/verbaalpina/>.

34 — Cf. <https://www.fdm-bayern.org/ehumanities-interdisziplinaer/>.

35 — Cf. S. LÜCKE, « Concession d'une licence »

peut être réutilisé de façon généralisée par un grand nombre de personnes. Il en résulte des *workflows* plus efficaces et des résultats optimisés pour l'ensemble de la communauté de recherche.

Beatrice COLCUC | Christina MUTTER  
Ludwig-Maximilians-Universität München

## Références bibliographiques

- G. I. ASCOLI, « Saggi ladini », *Archivio Glottologico italiano*, vol. 1, 1873, p. 1-556.
- E. BOEHMER, « De sonis grammaticis accuratius distinguendis et notandis », *Romanische Studien*, vol. 1, 1875, p. 295-301.
- TLG = T. F. BRUNNER (éd.), *Thesaurus Linguae Graecae*, University of California, Irvine, 1972-. <http://stephanus.tlg.uci.edu/> [page consultée le 29/01/2020].
- CONVENTION ALPINE (éd.), <https://www.alpconv.org/fr/home> [page consultée le 27/01/2020].
- FORCE11 (éd.), *The Fair Data Principles*, 2011-2017, <https://www.force11.org/group/fairgroup/fairprinciples> [page consultée le 21/01/2020].
- GoFAIR (éd.), *FAIR Principles*, <https://www.go-fair.org/fair-principles> [page consultée le 21/01/2020].
- T. KREFELD et S. LÜCKE (éd.), *Verba Alpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit, München*, (2014-), online, <http://dx.doi.org/10.5282/verba-alpina>
- T. KREFELD et S. LÜCKE, « 54 Monate Verba Alpina - auf dem Weg zur FAIRness », *Ladinia* XLIII, 2019, p. 139-155.
- T. KREFELD et S. LÜCKE, « Beta Code », *Verba Alpina-fr* 19/2, (créé : 16/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D191%26letter%3DB%237](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D191%26letter%3DB%237)
- T. KREFELD et S. LÜCKE, « butyru(m) », *Verba Alpina-fr* 19/2, *Lexicon alpinum*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D84%26db%3D192%23B128](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D84%26db%3D192%23B128) [page consultée le 19/02/2020].
- T. KREFELD et S. LÜCKE, « Réduction à types », *Verba Alpina-fr* 19/2 (créé : 16/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DR%2358](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DR%2358)

- T. KREFELD, « Communication scientifique sur le web », *VerbaAlpina-fr* 19/2, (créé : 16/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DC%2362](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DC%2362)
- T. KREFELD, « Transcription », *VerbaAlpina-fr* 19/2 (créé : 16/1, dernière modification : 16/2), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D191%26letter%3DT%2357](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D191%26letter%3DT%2357)
- SIL INTERNATIONAL (éd.), « ISO 639 Code Tables », *ISO 639-3 Registration Authority*, [https://iso639-3.sil.org/code\\_tables/639/data](https://iso639-3.sil.org/code_tables/639/data) [page consultée le 19/01/2020].
- S. LÜCKE et J. SCHULZ « Digital Object Identifier (DOI) », *VerbaAlpina-fr* 19/2, (créé : 17/2), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DD%2373](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DD%2373)
- S. LÜCKE, « API » *VerbaAlpina-fr* 19/2 (créé : 19/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DA%23138](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DA%23138)
- S. LÜCKE, « Concession d'une licence », *VerbaAlpina-fr* 19/2 (créé : 16/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DC%2341](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DC%2341)
- S. LÜCKE, « Einführung in die Geolinguistik », *Lehre in den Digital Humanities*, Version 3, (13.01.2020, 10:19), 2020, <https://www.dh-lehre.gwi.uni-muenchen.de/?p=78314&v=3>
- S. LÜCKE, « Métadonnées », *VerbaAlpina-fr* 19/2 (créé : 19/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DM%23140](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DM%23140)
- S. LÜCKE, « Notices d'autorité », *VerbaAlpina-fr* 19/2, *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DD%23114](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DD%23114)
- S. LÜCKE, « Page de code », *VerbaAlpina-fr* 19/2 (créé : 16/1, dernière modification : 19/1), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DP%2311](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DP%2311)
- S. LÜCKE, « Principes FAIR », *VerbaAlpina-fr* 19/2 (créé : 18/2), *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DP%23128](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DP%23128)

- C. MUTTER, « Wikidata », *Verba Alpina-fr* 19/2, *Méthodologie*, [https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage\\_id%3D13%26db%3D192%26letter%3DW%23105](https://doi.org/10.5282/verba-alpina?urlappend=%2Ffr%3Fpage_id%3D13%26db%3D192%26letter%3DW%23105)
- P. WIESINGER, « Das phonetische Transkriptionssystem der Zeitschrift 'Teuthonista'. Eine Studie zu seiner Entstehung und Anwendbarkeit in der deutschen Dialektologie mit einem Überblick über die Geschichte der phonetischen Transkription im Deutschen bis 1924 », *Zeitschrift für Mundartforschung*, vol. 31, 1964, p. 1-20.
- WIKIDATA, *Lexicographical data*, [https://www.wikidata.org/wiki/Wikidata:Lexicographical\\_data](https://www.wikidata.org/wiki/Wikidata:Lexicographical_data) [page consultée le 19/02/2020].
- WIKIPEDIA, *Autorité*, [https://fr.wikipedia.org/wiki/Autorit%C3%A9\\_\(sciences\\_de\\_l%27information\)](https://fr.wikipedia.org/wiki/Autorit%C3%A9_(sciences_de_l%27information)) [page consultée le 19/02/2020].
- M. WILKINSON, M. DUMONTIER, I. AALBERSBERG *et al.*, « The FAIR Guiding Principles for scientific data management and stewardship », *Sci Data* 3, 2016, doi:10.1038/sdata.2016.18 [page consultée le 19/01/2020].