

Beatrice Colcuc | Christina Mutter

Interopérabilité des données géolinguistiques à l'exemple du projet VerbaAlpina

<https://www.verba-alpina.gwi.uni-muenchen.de/>

Journée d'études: Quel dialogue numérique entre les atlas linguistiques galloromans? – Université de Lille, 26 septembre 2019

1) Introduction

- Changement dans la façon de penser la recherche
- Les principes FAIR
- L'interopérabilité

2) VerbaAlpina

- Présentation du projet
- Plan et organisation
- Domaine de recherche
- Les données



3) L'interopérabilité au sein de VerbaAlpina

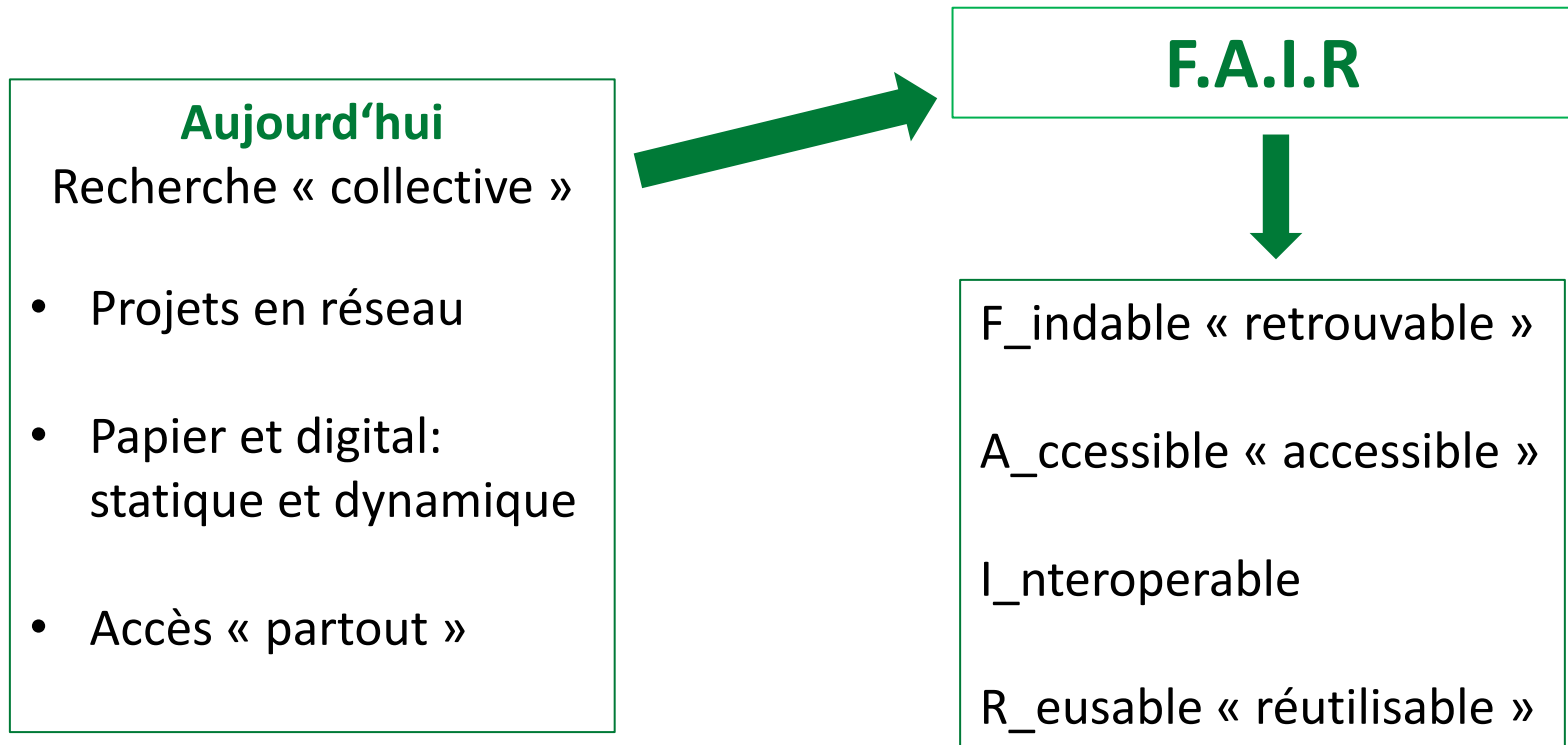
- Traitement structuré des données
- Enrichissement avec des métadonnées
- Attribution d'une norme standard (GeoNames, Q-ID, L-ID, GND)
- Référencement univoque aux données individuelles (URL, DOI)
- Accès aux données primaires et aux métadonnées (carte interactive, Lexicon Alpinum, API)

4) Les principes FAIR au sein de VerbaAlpina



Introduction

- Changement dans la façon de penser la recherche



Les principes FAIR

- Changement dans la façon de penser la recherche

F.A.I.R

F_indable « retrouvable » → A travers les catalogues des bibliothèques et des agrégateurs de données

A_ccessible « accessible » → Licences open-access

I_interoperable → Compatibilité de bases de données et leur interconnexion

R_eusable « réutilisable » → ceci résulte de F, A, I

Interopérabilité

- **Interopérabilité:** capacité des bases de données
 - à opérer ensemble
 - à se référer les unes aux autres

- **Défis et objectifs:**
 - création de données compatibles
 - création de liens persistants entre projets / bases de données
 - interconnexion de différentes bases de données



Le projet VerbaAlpina

Présentation du projet

- *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit* (« VerbaAlpina. L'espace culturel alpin à la lumière de son plurilinguisme »)
- Investigation de la région alpine:
 - Homogénéité ethnographique et topographique
 - Hétérogénéité linguistique

→ **But: montrer l'espace alpin dans son unité historique**



Le projet VerbaAlpina

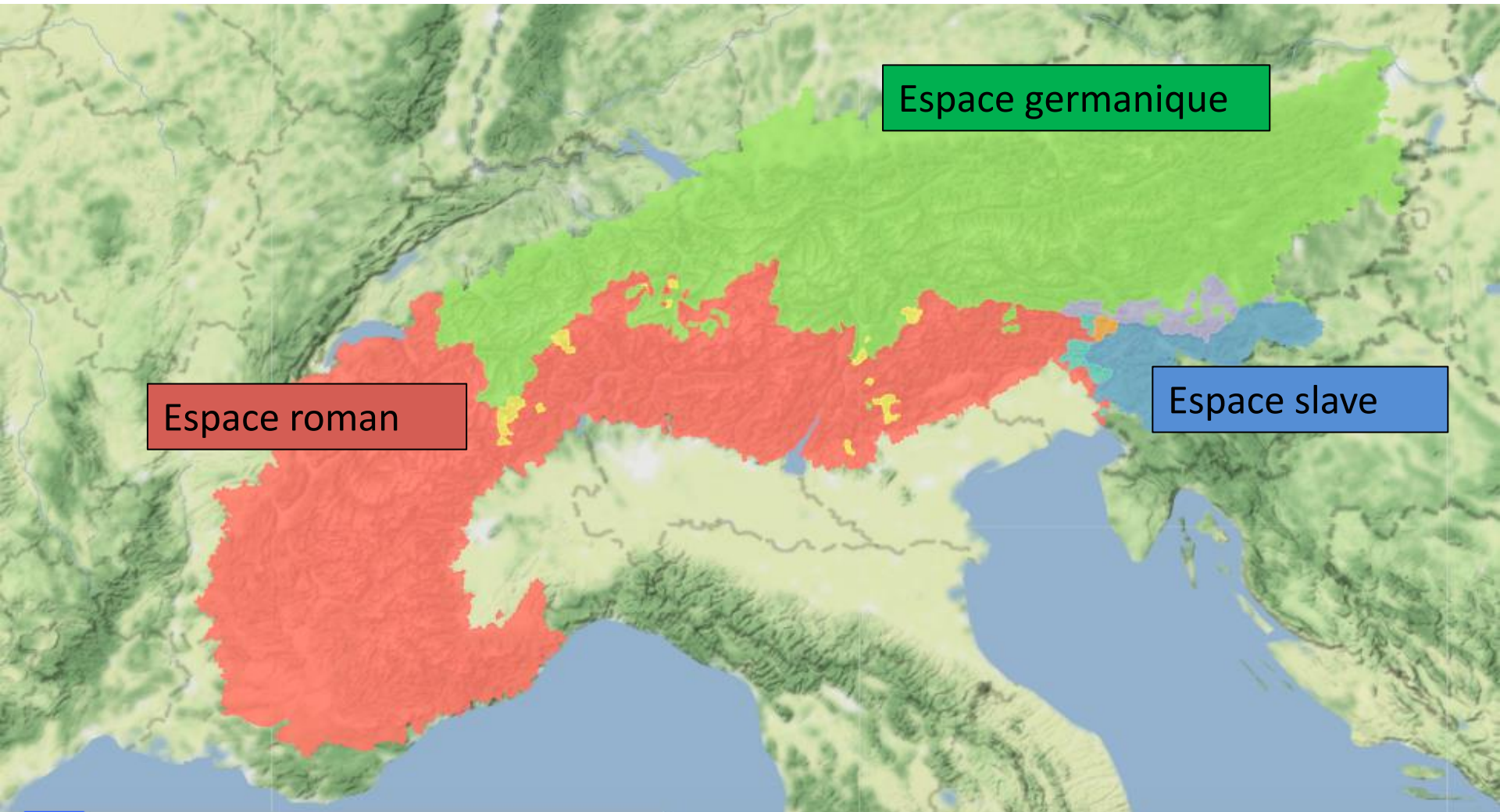
Plan et organisation:

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	culture <ul style="list-style-type: none"> • alpage • traitement du lait 			nature <ul style="list-style-type: none"> • formations des paysages • temps • faune • flore 			vie moderne <ul style="list-style-type: none"> • écologie • tourisme 			



Le projet VerbaAlpina

Domaine de recherche



Espace roman

Espace germanique

Espace slave



Le projet VerbaAlpina

Les données:

- Double provenance des données:

Atlas linguistiques et dictionnaires

- Analogues et digitaux
- Données géoréférencées

→ non homogènes

Crowdsourcing

- En ligne
- Intégrer, compléter et corriger les données des atlas/dictionnaires

→ Défi: Unifier les différents types de transcription



Le projet VerbaAlpina

Les données:

- Géoréférencement des données
- Distinction entre:
 - **Données primaires:** données linguistiques venant des atlas linguistiques, des dictionnaires et du crowdsourcing
 - **Métadonnées:** Toutes les données décrivant les données primaires
→ 'attributs'/'caractéristiques'



Grâce à:

1. Traitement structuré des données
2. Enrichissement avec des métadonnées
3. Attribution d`une norme standard (GeoNames, Q-ID, L-ID, GND)
4. Référencement univoque aux données individuelles (URL, DOI)
5. Accès aux données primaires et aux métadonnées (carte interactive, Lexicon Alpinum, API)

Interopérabilité au sein de VerbaAlpina

- Transcription

- Tokenisation

- Typisation

- Enrichissement avec des données sociolinguistiques et sociodémographiques

Traitement structuré des données

Enrichissement avec des métadonnées

L'interopérabilité au sein de VerbaAlpina

1. Traitement structuré des données

a. Transcription: Saisie du matériel des atlas et dictionnaires dans la base de données

erhebung	beta	ipa	Hex_ipa	Art
AIS			 	Trennzeichen
AIS	/	'	ˈ	Akzent
AIS	//	'	ˈˈ	Akzent
AIS	\	,	ˌ	Akzent
AIS	a	a	a	Vokal
AIS	a!	a	a	Zeichen
AIS	a%	añ	ah̆	Zeichen
AIS	a%0	añ	ah̆	Zeichen
AIS	a&	e	ɐ	Zeichen
AIS	a(ɑ	ɑ	Zeichen
AIS	a*	ɑ	ɑ	Zeichen
AIS	a(-	ɑ:	ɑː	Zeichen
AIS	a(-{i}	ɑĩ	ɑːĭ	Zeichen
AIS	a(-	e:	ɐː	Zeichen
AIS	a(:	æ	æ	Zeichen
AIS	a({i)-	ɑĩ	ɑːĭ	Zeichen
AIS	a(-	e:	ɐː	Zeichen
AIS	a)	ɑ	ɑ	Zeichen
AIS	a)-	ɑ:	ɑː	Zeichen
AIS	a*	a	a	Zeichen
AIS	a*(ɑ	ɑ	Zeichen
AIS	a*-	a:	aː	Zeichen
AIS	a*-?	a:	aː	Zeichen
AIS	a*:	æ	æ	Zeichen
AIS	a*?-	a:	aː	Zeichen
AIS	a*	e	ɐ	Zeichen
AIS	a-	ɑ:	aː	Zeichen
AIS	a-(ɑ:	ɑː	Zeichen

Système de

Alphabet phonétique international (A)

Transformation automatique

„Beta code“

Règles de transcription

AIS: Böhmer-Asco

ère étape de
uniformisation
Différentiation
entre caractère
de base et
diacritique
transcription
forme pour
les sources



b. Tokenisation: Fragmentation du matériel transcrit en mots individuels

2. Enrichissement avec des métadonnées

c. Typisation: Regroupement des mots en types

- de base
- morpho-lexicaux
- phonétiques

d. Enrichissement avec des données sociolinguistiques et sociodémographiques

→ But: montrer la cohésion du matériel linguistique au-delà des frontières

c. Typisation: Regroupement des mots en types

- **de base**: racine lexicale commune, appartenance à plusieurs familles linguistiques possible

Exemple

deu. *Butter*
fra. *beurre*
ita. *burro* } gr.-lat. *butyru(m)*

Interopérabilité et Verba Alpina

- **morpho-lexicaux**: mot prototypique pour différentes expressions; caractéristiques grammaticales communes

Exemple

roa. *vatsa* }
roa. *vatfo* } différents variantes d'un seul et même
roa. *vatfa* } type morpho-lexical (*vache*)

- **phonétiques**: concernent surtout la tradition dialectologique germanique; pour VA intérêt marginal



d. Enrichissement avec données sociolinguistiques et sociodémographiques

- Langue, commune

(si disponible: minorité, âge de l'informateur, sexe)

informanten (16×7.686)													
Id_Informant	Id_Gemeinde	Erhebung	Nummer	Sprache	Alter...	Gesc...	Minderheit	Ortsname	Georeferenz	Alpenkonven...	Position	Id_AK_Gemeinde	
1	60.227	AIS	1	rom	51	m	Bündnerromanisch	Brigels-Breil	RíÓñ` `@Ñ-`l`bG@	1	1	1.626	
2	60.232	AIS	3	rom	59	m	Bündnerromanisch	Pitasch	+êäiin`@j`úce>[G@	1	2	1.607	
3	60.127	AIS	5	rom	64	m	Bündnerromanisch	Ems-Domat	%w@Dfæ"@ygP`+gG@	1	3	1.506	
4	60.133	AIS	7	rom	26	m	Bündnerromanisch	Ardez	éD,çff\$@Bzš`bG@	1	4	1.512	
5	60.140	AIS	9	rom	zwisc...	w	Bündnerromanisch	Remüs - Ramosch	Njō#EÁ\$@Ic`Ž`gG@	1	5	1.518	
6	60.148	AIS	9_1	rom			Bündnerromanisch	Schleins-Tschlin	'`%`aú\$@W`ÉêVoG@	1	6	1.520	
7	60.261	AIS	10	rom	44	m	Bündnerromanisch	Camischolas (Tavet...	ì`qóó`~!`@v8JwWG@	1	7	1.631	
8	60.256	AIS	11	rom	84	m	Bündnerromanisch	Surrhein (Somvix)	Ûž`±YÍ`@Ü,İf>[G@	1	8	1.630	
9	60.267	AIS	13	rom	62	m	Bündnerromanisch	Vrin	,JÁ**@QLp`3SG@	1	9	1.616	
10	60.113	AIS	14	rom	41	w	Bündnerromanisch	Dahn (Präz)	ŋy`tīi`@Üž`±YÍ`G@	1	10	1.485	
11	60.104	AIS	15	rom	75	m	Bündnerromanisch	Mathon - Matón	~`TÉôKÓ`@8EGRQG@	1	11	1.501	
12	60.116	AIS	16	rom	68	w	Bündnerromanisch	Scharans - Scharons	l#žfæ"@đ`À->[G@	1	12	1.478	
13	60.071	AIS	17	rom	63	w	Bündnerromanisch	Lenz - Lantsch	≠šÉ!`#`@ç`äcwWG@	1	13	1.461	
14	60.150	AIS	19	rom	73	m	Bündnerromanisch	Zernez	oe`%`i23\$@üi`ò`~`Yg@	1	14	1.517	
15	60.914	AIS	22	rom	29	w		Olivone	è.`%`"á!`@Á`%`è`xÖCG@	1	15	1.658	

3. Attribution d`une norme standard

OBJECTIF: lier et mettre en relation les données entre elles

- **GeoNames** de www.geonames.org (pour les communes)
- **Q-ID** de Wikidata (pour les concepts)
- En partie aussi: **L-ID** de Wikidata (pour les types morpho-lexicaux) + **GND** de la Bibliothèque Nationale Allemande (pour les concepts)
(« Gemeinsame Normdatei », Norme standard commune)

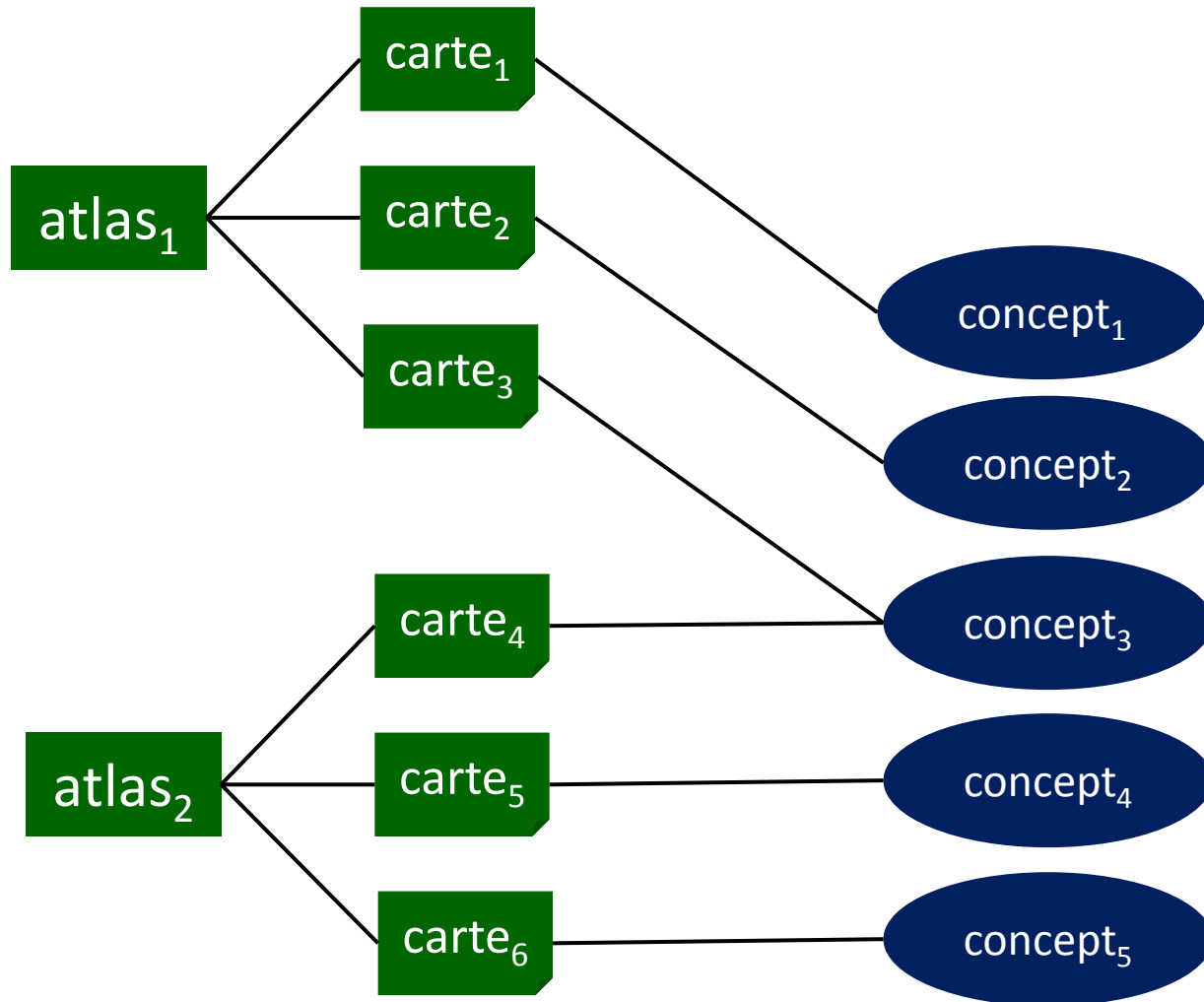
Lille
2998324

Q54050
(Q-ID)

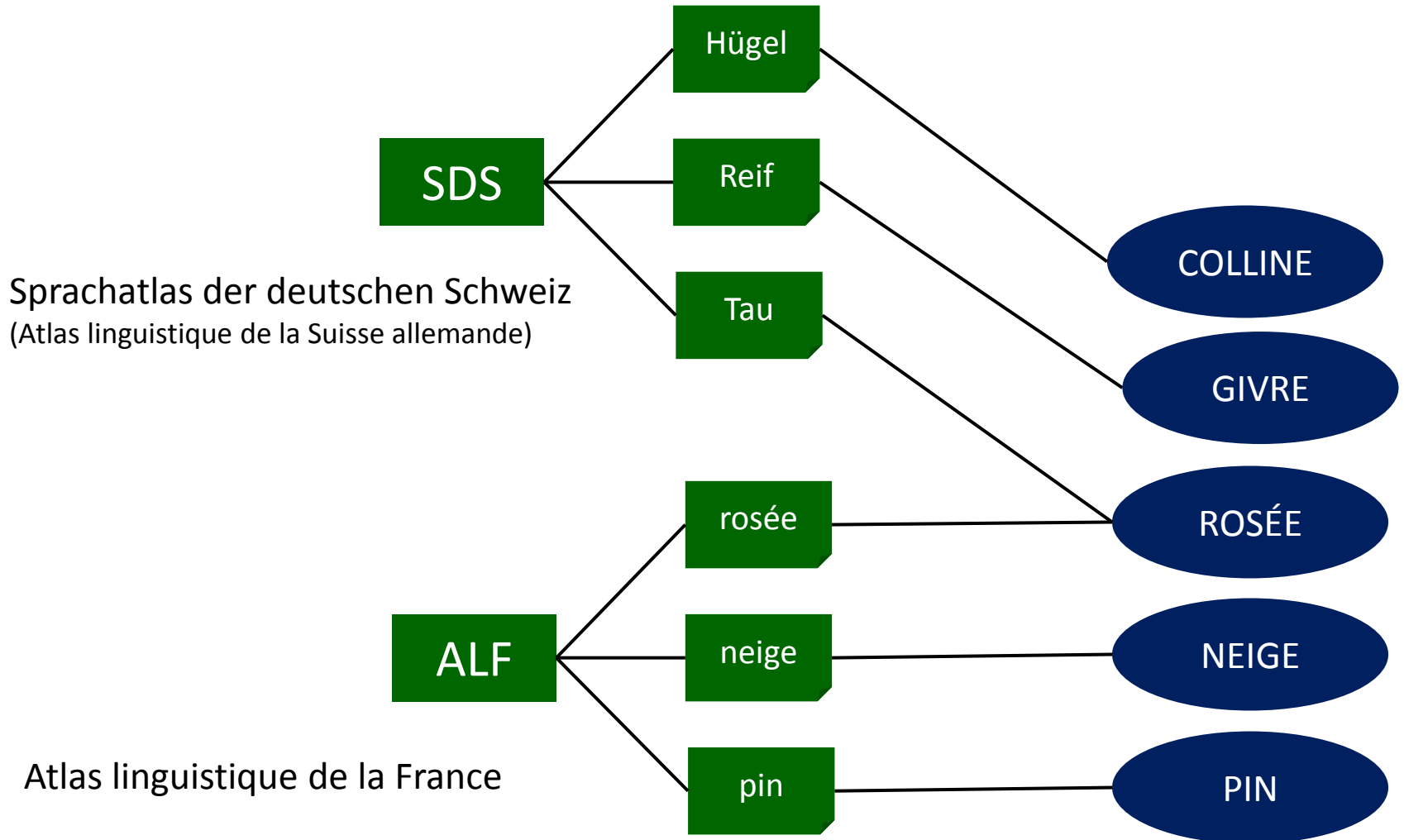
L73260
(L-ID)

4135744-9
(GND)

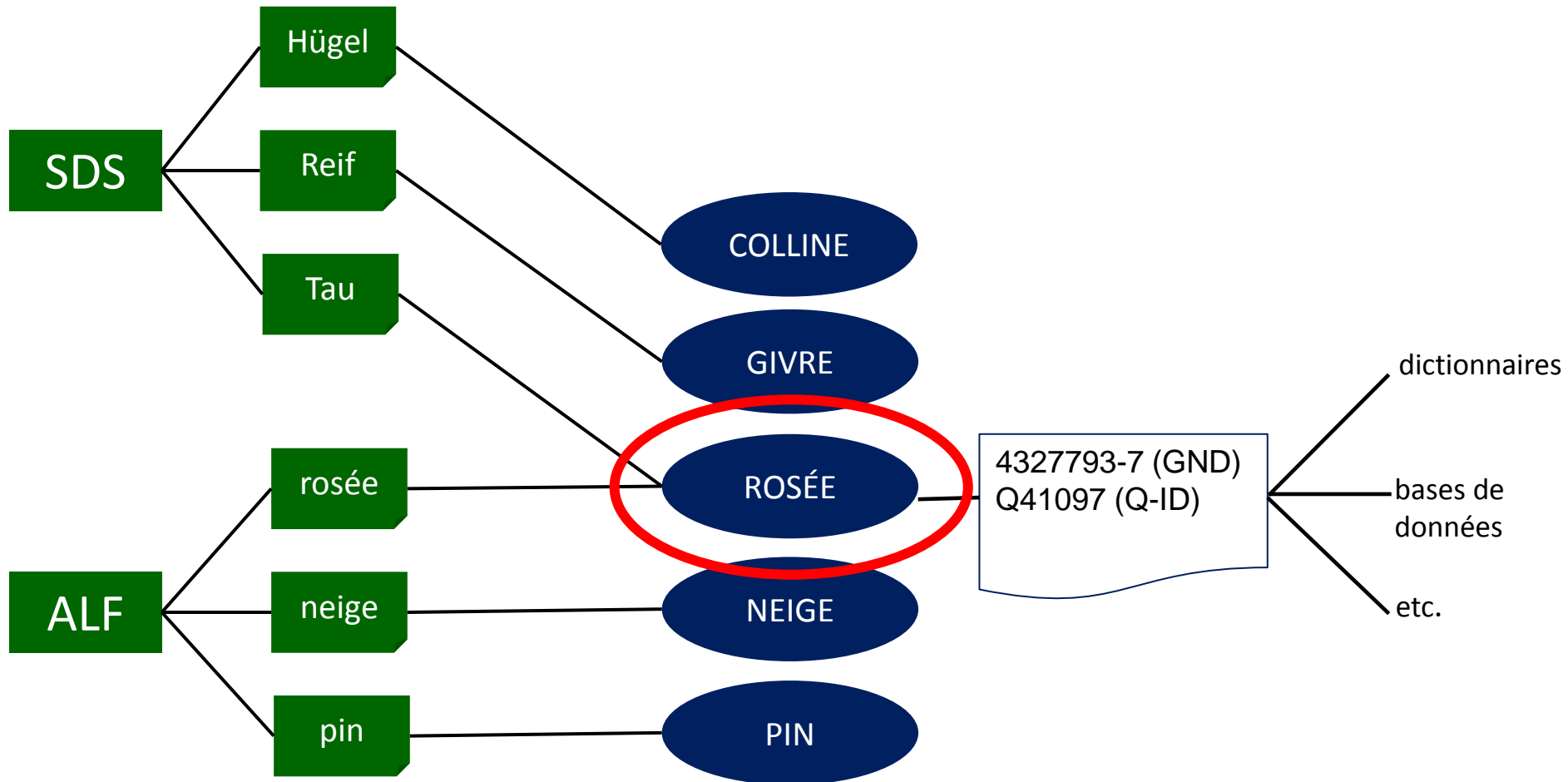
Interopérabilité au sein de VerbaAlpina



Interopérabilité au sein de VerbaAlpina



Interopérabilité au sein de VerbaAlpina



Interopérabilité au sein de VerbaAlpina



- Main page
- Community portal
- Project chat
- Create a new Item
- Create a new Lexeme
- Recent changes
- Random Item
- Query Service
- Nearby
- Help
- Donate

- Print/export
- Create a book
- Download as PDF
- Printable version

- Tools
- What links here
- Related changes
- Special pages
- Permanent link

Item [Discussion](#)



butter (Q34172)

dairy product

edit

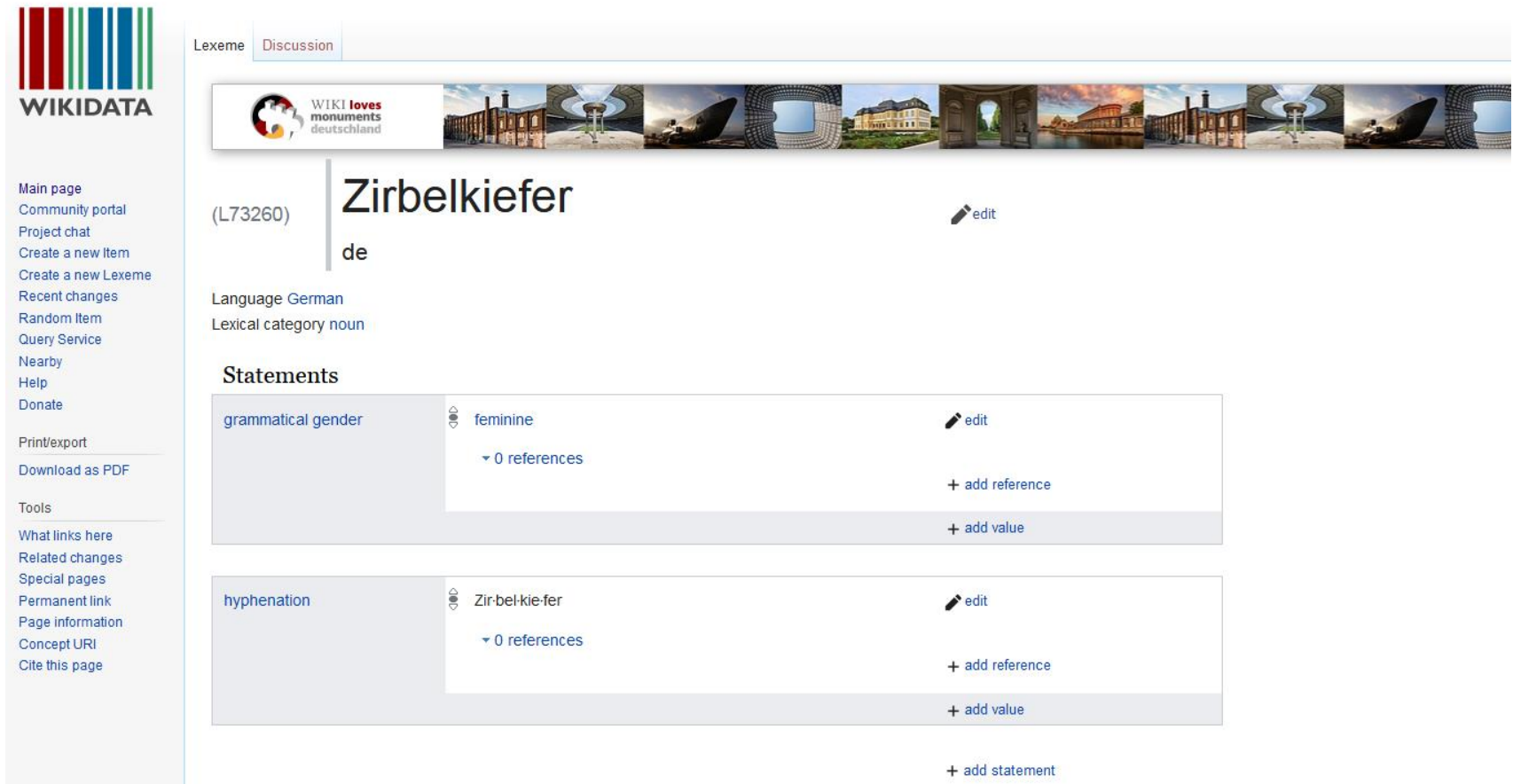
[In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	butter	dairy product	
German	Butter	aus Milch hergestelltes Streichfett	Salzbutter Tafelbutter Buttern Rohmilchbutter Sauerrahmbutter Kirnen Süßrahmbutter
French	beurre	aliment obtenu à partir de la matière grasse du lait	8029-34-3 Beurre d'escargot
Bavarian	Budan	No description defined	



[All entered languages](#)




„Zirbelkiefer“ – Pin montagnard



Lexeme [Discussion](#)




 




(L73260) **Zirbelkiefer**  edit

de

Language German
Lexical category noun

Statements

grammatical gender	 feminine  edit
	 0 references
	+ add reference
	+ add value

hyphenation	 Zir-bel-kie-fer  edit
	 0 references
	+ add reference
	+ add value

[+ add statement](#)

Interopérabilité au sein de VerbaAlpina



[Kontakt](#) | [A-Z](#) | [Träger / Förderer](#) | [Datenschutz](#) | [Impressum](#) | [Hilfe](#) | [Mein Konto](#) | [English](#)

↓ **Katalog**

→ **Einfache Suche**

→ Erweiterte Suche

→ Browsen (DDC)

→ Suchverlauf

→ Meine Auswahl

→ Hilfe

→ **Datenshop**

→ Mein Konto

→ Ablieferung von
Netzpublikationen

→ Informationsvermittlung

Login →

→ Über die Deutsche
Nationalbibliothek

KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

Gesamter Bestand

Musikarchiv

Exilsammlungen

Buchmuseum

→ **Suchformular zurücksetzen**

sw all "Butter"

Finden →

Expertensuche ?

eingeschränkt auf

- Materialarten: Elektronische Datenträger

- Normdaten: Sachbegriffe

Ergebnis der Suche nach: sw all "Butter"

← Zurück zur Trefferliste

Treffer 1 von 11

GND	
Link zu diesem Datensatz	http://d-nb.info/gnd/4009236-7
Sachbegriff	Butter
Quelle	M
Oberbegriffe	Milchprodukt Streichfett
DDC-Notation	637.2 641.372
Systematik	32.7 Milchwirtschaft ; 31.11 Lebensmitteltechnologie
Typ	Allgemeinbegriff (saz)
Andere Normdaten	LCSH: Butter RAMEAU: Beurre LCSH: Cooking (Butter) RAMEAU: Cuisine (beurre)

4. Référencement univoque aux données individuelles (URL, DOI)

- chaque donnée dispose d'un URL (Uniform Resource Locator) individuel (attestations individuelles, types morpho-lexicaux, concepts et municipalités)
- à chaque URL est assigné aussi un DOI (Digital Object Identifier)

5. Accès aux données primaires et aux métadonnées

→ à travers la **carte interactive** et le **Lexicon Alpinum**

→ à travers une **API**: « application programming interface »

(interface de programmation d'application)

→ permet de consulter des contenus spécifiques de la base de données VA (VA_DB) dans des formats définis à travers un navigateur Web

Interopérabilité au sein de VerbaAlpina

→ disponible à l`adresse:

<https://www.verba-alpina.gwi.uni-muenchen.de/?api=1>

→ documentation de la syntaxe à utiliser:

https://www.verba-alpina.gwi.uni-muenchen.de/fr/?page_id=128&noredirect=fr FR&db=191



Les principes FAIR au sein de VerbaAlpina

F_indable

- Coopération avec la bibliothèque universitaire (UB) de l'Université de Munich (dans le cadre du projet „e-humanities-interdisziplinär“) et le projet GeRDI (Generic Research Data Infrastructure)

OBJECTIF: rendre les données accessibles de manière centrale via le catalogue de la bibliothèque universitaire ainsi que via le portail de GeRDI

A_ccessible

- Licence „Creative Commons“ (compatible avec open-access) pour toutes les données gérées de VerbaAlpina (jusqu'à la version 18/1: CC BY SA 3.0, depuis 18/2: CC BY SA 4.0)



Les principes FAIR au sein de VerbaAlpina

I_nteroperable

- à travers une granulation fine du stock de données
 - traitement structuré des données
(transcription, tokenisation, typisation)
 - attribution d'un fichier de metadonnées au format DataCite à chaque fichier
 - connexion avec des données standard (*norm data*) déjà existantes
(Q-ID, L-ID, GND, GeoNames)
 - attribution des DOI individuels

R_eusable

- ceci résulte de F, A, I



Références bibliographiques

- Convention alpine (eds.), <https://www.alpconv.org/fr/home/> [Page consultée le 26 août 2019]
- Force11 (eds.), “The Fair Data Principles”, <https://www.force11.org/group/fairgroup/fairprinciples> [Page consultée le 26 août 2019]
- Krefeld, T.: s.v. “Transkription”, in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 16/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DT%2357
- Krefeld, T. / Lücke, S.: s.v. “Betacode”, in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 16/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DB%237
- Krefeld, T. / Lücke, S.: s.v. “Typisierung”, in: VerbaAlpina-de 19/1 (Erstellt: 16/1, letzte Änderung: 18/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DT%2358
- Kümmer, S. / Lücke, S. / Schulz, J. / Zacherl, F.: s.v. “Forschungsdatenmanagement”, in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DF%23112
- Lücke, S.: s.v. “API”, in: VerbaAlpina-de 19/1, Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DA%23138
- Lücke, S.: s.v. “Codepage”, in: VerbaAlpina-de 19/1 (Erstellt: 16/1), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DC%2311
- Lücke, S.: s.v. “FAIR-Prinzipien”, in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DF%23128
- Lücke, S.: s.v. “Metadaten”, in: VerbaAlpina-de 19/1, Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DM%23140
- Lücke, S.: s.v. “Normdaten”, in: VerbaAlpina-de 19/1 (Erstellt: 18/2, letzte Änderung: 18/2), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DN%23114
- Lücke, S. / Schulz, J.: s.v. “Digital Object Identifier (DOI)”, in: VerbaAlpina-de 19/1 (Erstellt: 16/1), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DD%2373
- Mutter, C.: s.v. “Wikidata”, in: VerbaAlpina-de 19/1 (Erstellt: 18/1), Methodologie, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DW%23105

Merci pour votre attention!

www.verba-alpina.gwi.uni-muenchen.de

