

Le thésaurus PACTOLS, système de vocabulaire contrôlé et partagé pour l'archéologie

The PACTOLS Thesaurus, a controlled and shared vocabulary system for archeology

Blandine Nouvel¹

¹ Aix Marseille Univ, CNRS, CCJ, Aix-en-Provence, France, nouvel@mmsh.univ-aix.fr

RÉSUMÉ. Créé par le réseau Frantiq (Fédération et Ressources sur l'Antiquité), initialement pour l'indexation documentaire de la littérature scientifique, PACTOLS est un réservoir unique de métadonnées thématiques réservé à l'archéologie. Le gestionnaire de thésaurus sur lequel il s'appuie, Opentheso, lui confère en outre des capacités techniques d'ouverture et d'interopérabilité qui complètent sa normalisation. De plus, l'attribution d'un identifiant pérenne à chaque concept fait de chacun un objet web, repérable et citable. PACTOLS constitue ainsi, de fait, un vocabulaire pivot qui répond aux standards du web sémantique et de la science ouverte. Il respecte les principes FAIR : il est consultable et accessible librement en ligne, il est interopérable et réutilisable. Ses caractéristiques lexicales et normatives représentent un outil au potentiel rare, proposé aujourd'hui à tous les acteurs de l'archéologie. Le projet de développement des PACTOLS, conduit dans le cadre du Consortium MASA de la Très Grande Infrastructure de Recherche Huma-Num, est destiné à offrir à la communauté des archéologues un vocabulaire de référence qui facilite le partage, sur le web, des données et des productions scientifiques en archéologie. Il s'oriente dans deux directions : il s'agit d'une part de consolider la structure sémantique qui aboutit à la réorganisation des domaines, facilitant la navigation dans le thésaurus. D'autre part, les contenus terminologiques sont enrichis, notamment par l'ajout de vocabulaires de spécialités et par l'alignement avec des vocabulaires et des programmes scientifiques du domaine, présents sur le web. La réorganisation du thésaurus se conduit en étroite collaboration avec les experts concernés. Pour cela, les modalités de contribution dynamique au thésaurus sont élargies à tous les acteurs de la discipline, sur la base de niveaux d'autorisation spécifiques. Une communauté élargie s'organise, formée et accompagnée à la prise en main partagée du thésaurus. Car PACTOLS, comme vocabulaire de référence, est préconisé dès le traitement des données de fouilles, par exemple à l'Inrap. Il est aussi associé à des programmes de valorisation des données de la recherche et intégré par plusieurs revues d'archéologie à leur production, via la chaîne de production éditoriale *Métopes* en XML-TEI. PACTOLS, réservoir commun pour les métadonnées en archéologie et archéoscience, constitue une brique à l'interopérabilité des données. Il s'insère aisément dans des systèmes simples de bases de données ou organisés plus puissamment avec des ontologies pour un web des données liées.

ABSTRACT. Created by the Frantiq network (Federation and Resources on Antiquity), initially for the documentary indexing of scientific literature, PACTOLS is a unique reservoir of thematic metadata dedicated to archaeology. The thesaurus manager on which it relies, Opentheso, also provides it with technical capabilities for openness and interoperability that complement its standardization. In addition, the assignment of a permanent identifier to each concept makes each one a web object, identifiable and citable. PACTOLS is therefore a pivotal vocabulary that meets the standards of the semantic web and open science. It respects the FAIR principles: it can be consulted and freely accessible online, it is interoperable and reusable. Its lexical and normative characteristics represent a tool with a rare potential, which is now available to all actors in archaeology. The PACTOLS development project, conducted as part of the MASA Consortium of the Huma-Num Very Large Research Infrastructure, is intended to provide the archaeological community with a reference vocabulary that facilitates the sharing of scientific data and productions in archaeology on the web. It is oriented in two directions: on the one hand, it consolidates the semantic structure that leads to the reorganization of domains, facilitating navigation in the thesaurus. On the other hand, terminology content is enriched, in particular by the addition of speciality vocabularies and by alignment with scientific vocabularies and programs in the field, which are available on the web. The reorganisation of the thesaurus is carried out in close collaboration with the experts concerned. To this end, the modalities of dynamic contribution to the thesaurus are extended to all actors in the discipline, on the basis of specific authorization levels. A larger community is organized, trained and supported in the shared handling of the thesaurus. Because PACTOLS, as a reference vocabulary, is recommended as soon as excavation data are processed, for example at Inrap. It is also associated with programs to promote research data and integrated by several archaeological journals into their production, via the editorial production chain *Metopes* in XML-TEI. PACTOLS, a common reservoir for metadata

in archaeology and archaeoscience, is a brick to data interoperability. It can be easily integrated into simple database systems or organized more powerfully with ontologies for a web of linked data.

MOTS-CLÉS. archéologie, thésaurus, interoperabilité, web des données.

KEYWORDS. archaeology, thesaurus, interoperability, linked open data.

Une recherche d'information efficace suppose que la description des données s'appuie sur des métadonnées contrôlées et validées. Menus déroulants et listes à cocher des bases de données cachent alors tout un vocabulaire spécialisé élaboré au gré des besoins par les auteurs. Au moment où ces mêmes réservoirs de données s'affichent dans le web avec vocation à s'interconnecter, l'utilisation d'un vocabulaire commun et partagé devient essentiel. Le choix des mots justes constitue déjà tout un programme en soi. Mais leur organisation doit respecter les standards de la gestion de l'information et se conformer aux règles du web des données. Or, dans le paysage de l'archéologie française, peu de référentiels obéissent à ces impératifs : les listes de mots-clés restent captives des systèmes et des bases de données, compromettant les échanges entre réservoirs et limitant de fait la recherche d'information.

C'est dans ce contexte que la Fédération et Ressources sur l'Antiquité¹ met à disposition de la communauté des archéologues son thésaurus PACTOLS avec l'ambition d'en faire l'outil de référence national. Créé en 1987 pour servir à l'indexation du Catalogue collectif indexé (CCI) du jeune réseau Frantiq, PACTOLS sert depuis quelques années à la qualification des données de la recherche. Il est constitué de sept domaines propres à l'archéologie, depuis la Préhistoire jusqu'aux périodes contemporaines : Peuples et cultures, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux, Sujets. Il rassemble plus de 30000 concepts contrôlés, liés par des relations hiérarchiques et d'association et traduits en six langues.

The screenshot displays the PACTOLS thesaurus interface for the term 'fibule'. On the left is a hierarchical tree of categories, with 'fibule' selected under 'bijou' > 'fixation de vêtement'. The main content area is divided into several sections: 'Terme(s) générique(s)' (bijou (BT), fixation de vêtement (BT)), 'Terme(s) associé(s)' (broche), 'Terme(s) synonyme(s)', 'Notes' (definition: Agrafe ou broche articulée par un ressort ou une charnière...), 'Terme(s) spécifique(s)' (listing various types of fibules like 'fibule à arc (NT)', 'fibule ansée (NT)', etc.), 'Alignement' (Wikidata - Fibule, Wikipedia - Fibule), and 'Coordonnées Gps'. On the right, there are sections for 'Traduction' (listing translations in German, Arabic, English, Italian, Dutch, and Spanish) and 'Image' (showing a photograph of a fibula).

Figure 1. Par exemple, le concept "fibule" tel qu'il se présente dans l'interface de consultation du thésaurus PACTOLS (au 1er août 2018)

Son lexique s'enrichit de façon continue. Ainsi, entre 2014 et juin 2018, 1895 concepts ont été créés, plus de 450 modifiés. La Chronologie a été mise à jour et réorganisée en quatre dossiers. Mais pour

¹ Frantiq, GDS3378 du CNRS, <https://www.frantiq.fr>

l'adapter à d'autres usages que ceux exclusifs à la documentation et l'ouvrir sur le web, il convenait de repenser complètement son organisation. Le programme lancé en 2017 est soutenu par le Consortium MASA, Mémoire des archéologues et des sites archéologiques². Si la préfiguration de la version à venir des PACTOLS liste désormais une vingtaine de domaines de premier niveau³ et s'inspire du *BackBone Thesaurus* de DARIAH⁴, les relations entre les concepts sont dès à présent rigoureusement corrigées, les contenus enrichis. Les alignements avec GéoNames, Wikidata et partiellement avec le *Art and Architecture Thesaurus* du Getty produisent un maillage propre à élargir et à enrichir toute recherche d'information basée sur des données qualifiées. De nombreux ajouts terminologiques, issus de vocabulaires de spécialités étoffent et rééquilibrent certaines thématiques. Car une mise à niveau lexical s'avèrait nécessaire pour les besoins précis des éditeurs et des archéologues. S'ils constatent qu'en moyenne 80 à 90% de leurs propres listes de mots-clés se retrouvent dans PACTOLS, il convient de compléter le lexique du restant, au bénéfice final et réciproque de tous les usagers.

L'alimentation de PACTOLS se fait en ligne sur autorisation et suivant un workflow largement ouvert à la collaboration. Il permet propositions, discussions et insertion (ou non) des concepts. Les contributeurs sont accompagnés dans le *mapping* de leur vocabulaire par les équipes de Frantiq et formés à l'utilisation des PACTOLS.

Le respect des normes relatives à son domaine et l'utilisation du logiciel libre de gestion de thésaurus Opentheso confèrent à PACTOLS des capacités d'interopérabilité et, nous venons de l'évoquer, de gestion collaborative. Structuré et contrôlé selon la norme ISO 25964 pour l'organisation et l'interopérabilité des thésaurus, sous licence libre ODbL, conforme au format XML-SKOS, interopérable via un webservice ouvert de type REST et exportable dans des formats non propriétaire (SKOS et JSON-LD), le thésaurus a poursuivi son évolution normative en intégrant des identifiants pérennes (URI Ark), rendant ainsi chaque concept unique et citable dans le web. De la sorte, PACTOLS répond aux principes FAIR⁵ d'accessibilité, de lisibilité, d'interopérabilité et de réutilisation pour le partage des données dans le web.

Reste la question du *webmapping*. Comme thésaurus, PACTOLS n'a pas vocation de départ à constituer une cartographie en ligne. Et pourtant aujourd'hui, un tiers de ses concepts sont déjà géo localisés. Il s'agit principalement des Lieux : la France en totalité et bientôt l'Espagne et l'Italie sont alignés *a minima* sur GéoNames et l'alignement sur Wikidata concerne l'ensemble du thésaurus. Il serait donc tout à fait envisageable de produire une carte des localités et des sites indexés avec PACTOLS. De plus, le système prendrait toute sa valeur si les corpus, en autorisant le retour d'information comme le fait aujourd'hui le CCI⁶, permettait au thésaurus de pointer les autres réservoirs liés. Ainsi partagé, PACTOLS pourra devenir le vocabulaire pivot à partir duquel l'internaute rebondira de corpus en corpus au sein du nuage des données liées.

² <https://masa.hypotheses.org>

³ Cf. Présentation de la feuille de route aux Journées du Consortium MASA « Interopérabilités », Tours, novembre 2017. <https://masa.hypotheses.org/481>

⁴ THESAURUS MAINTENANCE WORKING GROUP, 2016. *DARIAH Backbone Thesaurus (BBT): Definition of a model for sustainable interoperable thesauri maintenance, version 1.2*. DARIAH EU : 35 p. [En ligne] http://www.backbonethesaurus.eu/sites/default/files/DARIAH_BBT%20v%201.2%20draft%20v4_0.pdf

⁵ <https://www.force11.org/fairprinciples>

⁶ Pour chaque concept, PACTOLS comptabilise les notices et renvoie vers le catalogue