L'opérateur de collage : gestion de plusieurs points de vue dans un contexte spatial

Géraldine Del Mondo, Michel Mainguenaud

Normandie Université,
Institut National des Sciences Appliquées de Rouen, LITIS-EA4108
BP08 avenue de l'université 76801 Saint-Etienne-du-Rouvray , France
{geraldine.del_mondo,michel.mainguenaud}@insa-rouen.fr

RÉSUMÉ. *De nombreux outils d'aide à la navigation s'appuient sur une représentation cartographique de l'espace couvert par le trajet effectué ou à effectuer. L'aspect sémantique (informations alphanumériques) associé aux éléments rencontrés lors de ce trajet est très faiblement étudié.*
Dans cet article, nous nous intéressons à la manipulation de l'information spatiale inhérente à ce type d'application et proposons un opérateur de Collage qui permet d'obtenir un unique point de vue à partir de deux points de vue différents. L'opérateur est basé sur la fusion de deux modèles alphanumériques (supposés alignés) réalisée à partir d'un ensemble de règles définies sur les caractéristiques des modèles associés à chacun des points de vue. La mise en œuvre de cet opérateur s'appuie sur un modèle de graphe et utilise des propriétés des deux modèles comme l'intersection, l'existence de sous-graphes quasi-fortement connexes ou l'existence de chemins pour définir les règles de construction du modèle de données résultat de l'application de l'opérateur de Collage.

ABSTRACT. *Many navigation aid tools rely on a cartographic representation of the space covered by the journey made or to be made. The semantic aspect (alphanumeric information) associated with the encountered elements during this trip is poorly studied in the literature.*
In this article, we propose a Stick (Collage) operator that provides a unique point of view from two points of view. It is based on two alphanumeric models we merge in order to obtain a single model using a set of rules defined on the characteristics of the models associated with each point of view.
We rely on a graph-oriented model and use characteristics such as the intersection of the two models or the existence of quasi-strongly connected sub-graphs, paths ... to define the construction rules for the result data model associated with the application of the Stick (Collage) operator.

MOTS-CLÉS : *information sémantique alphanumérique, niveau d'abstraction, système d'information géographique, fusion de modèles*

KEYWORDS: *alphanumeric semantic information, abstraction level, geographical information system, models fusion*

DOI:10.3166/RIG.x.1-36 © 2018 Lavoisier

1. Introduction

L'intérêt grandissant pour la problématique des Big Data vient renforcer les recherches dans le domaine de la géomatique. En effet, de nombreuses informations manipulées dans ce cadre possèdent une (géo-)localisation. La manipulation et l'analyse de l'information géographique devient donc un enjeu de plus en plus prégnant.Nous pouvons citer à titre d'exemples : en tant que VGI (Volunteered Geographic Information), dans le cadre de la remontée d'informations coopératives comme avec les boitiers "Coyotte" ou GPS dans les véhicules ou lors de l'utilisation des applications sur smartphone s'appuyant sur la géo-localisation.
Les niveaux d'abstraction sont très variables : du mondial à l'individu en passant par les région, département, ... A tous les niveaux, les applications utilisent de plus en plus des données multi-sources. Il convient alors de les rapprocher afin de permettre une analyse plus fine ou des services mieux adaptés. La problématique de la définition d'opérateurs multi-niveaux d'abstraction pour la manipulation de ces données, tant sur le plan sémantique que sur le plan de la spatialité, reste un problème scientifique d'actualité.
La mise à disposition d'opérateurs fermés (afin d'autoriser la composition d'opérations) permet d'offrir des opérateurs de haut niveau. Ils peuvent donc s'exprimer comme une combinaison d'opérateurs de plus bas niveaux. Il apparait alors deux approches. Une première approche, dite de type Bottom-up, part des éléments cartographiques pour remonter vers les niveaux d'abstraction d'un modèle de données. Une deuxième approche, dite Top-Down, part des modèles de données pour aller (éventuellement) jusqu'aux composants de représentation visuelle. Nous nous situons résolument dans ce dernier type d'approche.

A titre d'exemple, nous allons illustrer cette problématique pour une aide à la navigation. A l'image des logiciels embarqués à bord de véhicules, appelés par abus de langage GPS embarqués, la gestion des zones présentant une difficulté de navigation fait l'objet d'un traitement spécifique. A l'arrivée près d'une zone où le nombre d'échangeurs, rues ou autres zones de dégagement est important, la représentation cartographique (certes minimale) est améliorée afin de permettre une meilleure visibilité de l'environnement. Le conducteur dispose alors d'une représentation avec un meilleur niveau de détail afin de le guider dans sa prise de décision. Une fois cette zone où les risques d'erreur sont importants passée, l'affichage est de nouveau présenté avec une couverture spatiale plus importante et un niveau de détail plus faible. Cependant, que le déplacement soit effectué avec une vitesse relative importante (e.g. véhicules, trains à grande vitesse) ou plus faible (e.g. piétons), le problème de la mise à disposition d'informations de nature sémantique et non uniquement cartographiques (comme dans (Mustière, Devogele,2008)) reste ouvert (Raimond, Mustière,2008). De nombreux outils sont actuellement disponibles, tant dans les véhicules que sur les téléphones portables par exemple. Ils sont essentiellement basés sur une approche cartographique de la représentation de l'espace. La partie sémantique est relativement absente de ces propositions. Deux niveaux de traitement peuvent donc

être mis en avant. (a) Un premier niveau concerne la représentation cartographique de l'espace et de son traitement (e.g., généralisation, sémiologie graphique, placement de toponymie). (b) Un deuxième niveau concerne le modèle de données permettant de gérer la sémantique des informations (alphanumériques) associées aux objets présents dans cet espace. Nous nous positionnons résolument dans ce deuxième niveau.
Les outils de navigation embarqués fournissent des informations alphanumériques issues d'une mono source. Le développement de l'inter-opérabilité conduit à la composition de modèles différents. La composition de ces modèles fournit une information enrichie à l'utilisateur. Ainsi, l'ajout de sémantique peut provenir d'applications différentes travaillant au même niveau d'abstraction dans le cadre d'une démarche d'inter-opérabilité. C'est le cas avec le développement des serveurs de type WMS/WFS (Web Map Service/Web Feature Service). Ils ont pour objectif entre autres de mettre à disposition des représentations cartographiques gratuites ou payantes avec les informations alphanumériques associées. Ces ajouts peuvent aussi provenir d'un changement de niveau de granularité lors des manipulations de données. Des approches comme celles de (Stell, Worboys,1998) ou de (Aiken *et al.*,1996) en sont une illustration. Les ajouts de sémantique peuvent aussi provenir d'une approche mixte comme le rapprochement de plusieurs points de vue en une vision commune. Ce principe s'apparente à l'intégration de vues dans le contexte des bases de données relationnelles. Il peut être élaboré sous la forme d'une vision par inter-opérabilité (des schémas exogènes mis en commun pour offrir une vision d'un espace géographique) ou par rapprochement de point de vue d'un même système d'information (des schémas endogènes mis en commun pour offrir un complément d'information). Dans les deux cas de figure, nous retrouvons le besoin d'un opérateur permettant ce rapprochement de modèles.

Dans cet article nous proposons donc d'aborder la gestion des informations alphanumériques, en particulier celles qui sont liées à la dimension spatiale. Nous proposons un opérateur de *Collage* définissant un ensemble de règles qui permettent de conserver une sémantique spatiale consistante dans ces données. Cet opérateur a pour objectif de permettre le rapprochement entre différentes perceptions d'un espace sur le plan des informations sémantiques associées à une représentation spatiale. Il a pour vocation d'être complémenté par un opérateur de manipulation des informations cartographiques permettant la représentation visuelle de l'espace ainsi considéré (e.g. mise en correspondance des objets cartographiques ayant subi des généralisations différentes, ayant une légende différente, ...). Seul l'opérateur de *Collage* est défini dans cet article.
Afin de prendre en compte un volume toujours plus important de données (pas forcément structurées) deux grandes familles d'architectures sont mises en place : l'approche par médiation (?,? ; ?) et l'approche par fédération (?,? ; ?). L'architecture utilisée est orthogonale à l'approche que nous souhaitons mettre en oeuvre. Nous n'aborderons donc pas ce problème dans cet article. Au niveau des modèles de formalisation des informations, les travaux de la littérature portent sur deux grandes approches : le traitement des ontologies sur lesquelles les modèles de données sont

définis (Cruz, Xiao,2008) et les artéfacts d'intégration de modèle (Aiken *et al*.,1996), (Estublier *et al*.,2008). Les ontologies et le traitement de la sémantique repose sur les informations liées à la sémantique de l'application ou géométriques (e.g., au dessus, en dessous) mais au niveau des classes, entités, ... (le nom dépendant du formalisme retenu). Nous supposons ici le premier point résolu en utilisant des ontologies ayant été alignées. A l'image du modèle relationnel et de son processus de conception par intégration de vues, cette opération peut être faite dans un spectre qui va de l'intégration des vues deux à deux à l'intégration de l'ensemble des vues en une étape. Ainsi, les modèles de données, provenant éventuellement de différentes sources, ne présentent pas d'ambiguïté. Les artéfacts de modélisation proposent un cadre de résolution du problème sur lequel nous allons nous appuyer. A la différence des approches proposées dans la littérature, nous nous appuyons sur les propriétés définies au niveau des attributs.

Nous nous plaçons donc dans le cadre d'une manipulation binaire pour effectuer l'opérateur de collage et au niveau des attributs et non des classes, entités,... Nous pouvons alors nous focaliser sur les artéfacts d'intégration afin d'offrir le maximum d'informations pertinentes à un utilisateur à partir des données disponibles dans les deux modèles. Ces deux modèles peuvent être complémentaires ou peuvent présenter des concepts communs. Nous utilisons un modèle de données prenant en compte la pertinence spatiale des informations vis à vis de la représentation spatiale logique (et non cartographique) des informations manipulées. Une vision orthogonale à la perception spatiale est l'introduction de la composante temporelle dans l'aide à la navigation. Bien que cette composante soit importante, notre proposition se situe dans la mise à disposition d'informations à un instant t du parcours. Seule la composante spatiale est utilisée dans cet article.
Nous définissons un ensemble de règles permettant de construire le modèle de données associé au résultat de l'application de l'opérateur de *Collage*. Nous nous appuyons sur un modèle de classe représenté par un graphe étiqueté par des propriétés comme l'héritage, l'inclusion (stricte) spatiale ou des contraintes de cardinalité. A partir d'un graphe de référence, le résultat de l'opérateur de *Collage* entre deux sous-graphes est de nouveau un graphe dont l'ensemble de noeuds et d'arcs sont définis en prenant en compte des propriétés comme l'intersection vide (ou non) des deux sous-graphes, l'existence de chaîne/chemins (ou non) entre les deux sous-graphes considérés dans le graphe de référence.

Dans la section 2, nous présentons un état de l'art de l'approche des intégrations de modèle en liaison avec notre problématique. Dans la section 3, nous présentons le principe de l'opérateur de *Collage*. Dans la section 4, nous présentons le traitement de modèles indépendants. Dans la section 5, nous présentons le traitement de modèles non indépendants. Dans la section 6, nous présentons une conclusion de nos travaux et leurs perspectives.

2. Etat de l'art

Notre problématique se positionne au niveau du modèle de données (alphanumériques) associé aux informations manipulées dans le contexte d'une aide à la décision sur un environnement spatial. Dans le cadre de l'inter-opérabilité, de nombreux travaux chapeautés ou non par l'Open GeoSpatial Consortium ont eu pour objectifs de permettre l'échange de données/traitements entre organismes différents (Lieberman *et al.*,2007 ;OGC-WM,2006). L'une des principales directions suivies est l'alignement d'ontologies (Pavel, Euzenat,2013 ;Partyka *et al.*,2008). L'alignement d'ontologie couvre deux aspects : La découverte des correspondances entre deux ontologies différentes (équivalence de concepts ou de relations, subsomption, etc.) et l'expression de ces correspondances à travers un modèle de données. Les problèmes d'interprétation de sémantique relèvent de ce niveau (Harel, Rumpe,2004). Dans ce travail, nous considérons que cette expression est déjà disponible au travers de la mise à disposition d'un alignement cohérent des modèles.

L'Ingénierie Dirigée par les Modèles (IDM) est le domaine de l'informatique mettant à disposition des outils, concepts et langages pour créer et transformer des modèles. L'ajout d'information provenant de sources différentes ou d'une même source mais avec des points de vue différents (*i.e.* niveaux de granularité différents) se retrouve donc fusionné dans une seule et même problématique : la composition de modèle.

Nous proposons dans un premier temps les solutions proposées à partir d'une approche formelle de la modélisation de cette problématique, puis dans un deuxième temps la déclinaison opérationnelle de ces solutions, nous terminons par une synthèse de ces solutions.

2.1. Approche formelle

L'approche formelle repose sur la construction d'un modèle algébrique, sur la gestion de différents niveaux de granularité et sur l'intégration d'artéfacts.

2.1.1. Modèle algébrique

(Herrmann *et al.*,2007) propose une approche algébrique pour la composition de modèle. Dans sa forme la plus simple, la composition de modèles est le mécanisme de combinaison de deux modèles pour en produire un troisième. Le point de départ est le lien entre un modèle et un domaine d'application. Il s'appuie sur une fonction, sm, appliquée sur un langage de modélisation (M) et un domaine sémantique (D) tels que tout modèle, $m \in M$, doit être mis en correspondance avec un domaine D :

$$\mathrm{sm} : \mathrm{M} \rightarrow \mathrm{D}$$

A titre d'exemple, le langage de modélisation peut être UML et le domaine sémantique peut être une application d'aide à la navigation. Le langage permettra de modéliser des composants comme une ville et les voies de communications.

L'hypothèse de départ est la consistance des modèles vis à vis des applications et un modèle m_2 précise un modèle m_1 si :

$$sm(m_2) \subseteq sm(m_1)$$

Les systèmes implémentant le modèle m_2 sont moins nombreux que les systèmes implémentant le modèle m_1. Hermann et al définissent un opérateur binaire de base ($\otimes$) qui à partir de deux modèles produit un modèle composé :

$$\otimes : \text{M x M} \rightarrow \text{M}$$

Partant de cet opérateur, ils définissent trois propriétés concernant la préservation des propriétés : "Property Preserving" (PP), "Full Property Preserving" (FPP) et "Consistency Preserving" (CP).

La première propriété (PP) repose sur :

$$\text{PP} : \forall m_1, m_2 \in M : sm(m_1 \otimes m_2) \subseteq sm(m_1) \ \wedge \ sm(m_1 \otimes m_2) \subseteq sm(m_2)$$

Cette propriété assure qu'aucune information présente dans le premier (resp. deuxième) modèle n'est perdue lors de la composition.

La notion de $\subseteq$ est liée au fait que des informations issues de la composition peuvent être ajoutées alors qu'elles ne faisaient parties d'aucun des modèles m_1 ou m_2.

La deuxième propriété (FPP) repose sur :

$$\text{FPP} : \forall m_1, m_2 \in M : sm(m_1 \otimes m_2) = sm(m_1) \cap sm(m_2)$$

Cette propriété permet d'analyser le modèle résultat à partir de chacun des modèles source puisque la définition est construite à partir de l'égalité. Les modifications associées à un modèle opérande n'impactent que la partie concernée par ce modèle dans le modèle résultat de la composition.

La troisième propriété (CP) repose sur :

$$\text{CP} : \forall m_1, m_2 \in M : sm(m_1) \cap sm(m_2) \neq \emptyset \Rightarrow sm(m_1 \otimes m_2) \neq \emptyset$$

L'ajout d'information, liée à la composition, peut rendre le modèle résultat inconsistant (*i.e.*, $sm(m_1 \otimes sm(m_2)) = \emptyset$) même si initialement les modèles pris individuellement n'étaient pas inconsistants (*i.e.*, $sm(m_1) \cap sm(m_2) \neq \emptyset$).

Partant de cette définition formelle de la composition de modèles, deux grandes approches peuvent être définies pour rendre opérationnelle cette composition : la structuration par niveau de granularité ou la composition par l'intégration d'artéfacts.

2.1.2. Niveaux de granularité

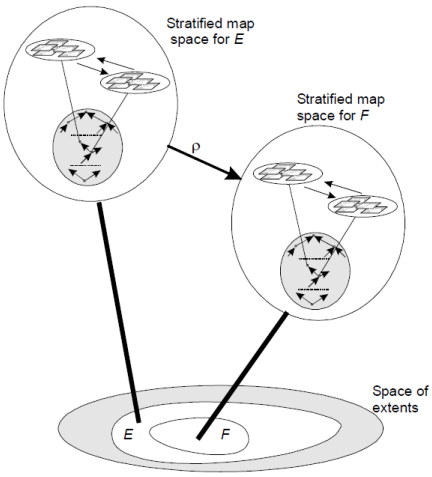

Figure 1. Illustration de Sheaf of Stratified map space (Stell, Worboys,1998)

Les approches liées à la gestion d'information sémantiques à plusieurs niveaux d'abstraction sont essentielles pour la compréhension globale d'un phénomène (Longo, Medeiros,2013 ;El Adnani *et al.*,2001). Les travaux de (Stell, Worboys,1998) sont dans cette philosophie et ont l'avantage de décrire un cadre conceptuel de haut niveau permettant de représenter plusieurs niveaux de granularité (*i.e.* les m_i de l'approche section 2.1.1). Les auteurs de ces travaux définissent une fonction ρ qui effectue le passage d'un niveau à l'autre (figure 1). Ce cadre n'impose pas de spécificité de modélisation des données comme dans (Fonseca *et al.*,2002) et repose sur les concepts suivants :

– *Map* : Collection de données (de préférence ayant des attributs sémantiques et géométriques), c'est un état de la base de données (BD). A titre d'exemple, une carte papier représente une Map au sens de ces travaux.

– *Map space* : Ensemble de *Map*, c'est à dire toutes les instances d'une base de données selon un certain schéma fixé. Elles sont partiellement ordonnées selon leur précision sémantique ou spatiale (dans le sens où toutes les informations du schéma fixé ne sont pas forcément disponibles pour une carte). A titre d'exemple, la BD Carto propose un schéma de données et représente le Map Space. Chacune des cartes définies selon ce schéma sera un élément de la Map Space.

– *Granularity lattice* : Ordonne des *Maps* dont les niveaux de granularité spatiale et sémantique varient entre un seuil minimum et maximum (deux treillis différents). A titre d'exemple, comme toutes les informations (à partir du schéma initial fixé) ne sont pas forcément disponibles pour une carte donnée, les cartes peuvent être rangées selon un ordre (partiel) définissant le niveau d'informations qu'elles offrent.

– L'étendue (*extent* E) d'une représentation se définit par la totalité des éléments qui la composent. L'étendue peut-être sémantique ou géométrique. A titre d'exemple, elle représente l'ensemble des éléments formant le schéma de la base de données et qui seront renseignés pour cette carte (sur le plan sémantique) et l'ensemble des éléments représentés sur la carte (zone, points, ... auxquels une légende est attribuée).

– *Stratified map space (SMS)* $\Sigma(E)$: Ce concept permet de gérer les différents niveaux de définition tant sémantique que spatial. Un SMS est un *granularity lattice* dont les éléments sont des *Map space*. L'idée ici est d'ordonner au sein d'un treillis les différentes *Map space* selon leurs niveaux de granularité (spatiale et) sémantique. A titre d'exemple, la définition de deux schémas de base de données, BD Carto vs. BD Topo va permettre d'introduire différents niveaux de détail dans la définition (sémantique et spatiale) des objets manipulés. Le SMS représente donc le treillis formé par ces deux schémas. Il est possible de faire varier le niveau de granularité en utilisant deux opérations définies entre *Map space*: la généralisation Gen et son opposé $Lift$. Un SMS n'est associé qu'à une seule *étendue* (i.e., l'*étendue* est fixe).

– *Sheaf of Stratified map space* : Permet de ne plus être limité par une seule étendue et de mettre en relation des SMS. A chaque SMS est associé une étendue. Un Sheaf of Stratified map space est un ensemble de SMS et les auteurs définissent une fonction ρ qui permet de passer d'un SMS à l'autre. A titre d'exemple, l'objectif ici serait de manipuler des cartes provenant de différents fournisseurs de données et de les manipuler sur l'espace commun qu'elles couvrent.

L'aspect statique du modèle est fourni par les différentes structurations (*extent*, *SMS*). L'aspect dynamique est fourni par l'opérateur ρ. Dans cet article nous allons étudier plus précisément la dynamique sous son aspect sémantique (niveau d'étude non présent dans la proposition de (Stell, Worboys,1998)). Nous nous positionnons au niveau des *Sheaf of Stratified map space* permettant de mettre en liaison différentes *étendues*.

2.1.3. L'intégration d'artéfacts

L'intégration d'artéfacts (*i.e.* les m_i de l'approche section 2.1.1) se base sur le principe de l'utilisation de modèles hétérogènes concernant le niveau de détail (*i.e.* niveau de granularité) dans le but d'offrir un modèle homogène, appelé domaine (Estublier *et al.*,2008). Elle peut utiliser une orientation du haut vers le bas (top-down) ou du bas vers le haut (bottom-up). Surtout appliquée dans le génie logiciel, elle permet dans une orientation du haut vers le bas d'identifier les artéfacts de programmation indépendamment des technologies utilisées (ici l'équivalent des signatures publiques et donc du modèle d'interface). Dans une approche du bas vers le haut, le concepteur doit définir les artéfacts logiciels et la consistance de leurs interactions. Quelle que soit

l'approche, la synchronisation entre les artéfacts logiciels et leurs modèles se traduit par des actions effectuées par les artéfacts logiciels.

Les travaux de (Aiken *et al.*,1996) prennent en compte l'aspect sémantique en proposant à l'utilisateur de créer ses propres programmes dans un langage dédié afin de visualiser les données selon le niveau de granularité souhaité. L'idée est de pouvoir naviguer au travers de données multidimensionnelles en créant des vues différentes (*i.e.* dimensions ou niveaux de granularité différents) selon les besoins de l'utilisateur. Cependant, la réalisation des programmes, bien qu'utilisant un langage graphique, n'est pas triviale et le maintien de la cohérence des contextes entre deux niveaux de granularité n'est pas clairement défini.

Un moyen de garantir cette cohérence de contexte (illustré figure 2), est de proposer des opérateurs qui à partir de modèles potentiellement distincts (e.g. m_1, m_2) génèrent des modèles (e.g. m_3) qui respectent les propriétés citées au section 2.1.1 et une structure cohérente.

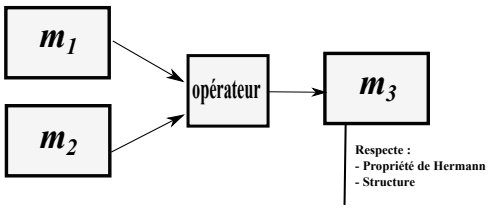

Figure 2. Objectif général

2.2. Déclinaison opérationnelle

Le développement de l'inter-opérabilité conduit à la composition de modèles différents. La composition de ces modèles fournit une information enrichie à l'utilisateur. Il apparait alors deux grandes familles d'approches pour représenter une liaison entre deux niveaux de perception de l'espace sur le plan des informations sémantiques. (a) Une première approche, dénommée interprétation, repose sur la construction ad-hoc de l'espace sur le point d'intérêt concerné. (b) Une deuxième approche, dénommée compilation, repose sur la construction globale des interactions possibles entre les deux niveaux de perception de l'espace et ce sur l'intégralité des points d'intérêt potentiels (*i.e.*, de leur zone de recouvrement).

2.2.1. Interprétation

Cette approche est similaire au traitement graphique des logiciels d'aide à la navigation embarqués dans les véhicules mais sur le plan du modèle de données et non de la représentation de l'espace. Pour un modèle de données et à une localisation spatiale (*i.e.*, une instance de ce modèle de données), le modèle associé à cette instance est enrichi grâce à un apport externe sous la forme d'une "fusion" des modèles de données et des instances qui y sont associées.

Compte tenu du fait que l'approche interprétée utilise un point d'intérêt spécifique, la signature publique de cette opération de rapprochement de modèles, appelée ici $Collage$ sera différente de l'approche compilée plus généraliste. Ce focus permet de ne pas tout recalculer. Le principe est donc basé sur :

$$Collage : \text{Modèle1 x Modèle2 x pointIntérêt} \rightarrow \text{ModèleRésultat}$$

La validité associée au modèle résultat est donc locale au point d'intérêt concerné. A titre d'exemple la jonction entre deux autoroutes pour entrer dans une ville, A, définie sur deux points d'intérêt spatialement différents mais sémantiquement identiques, B et C (par exemple une arrivée par autoroute au nord et au sud d'une même ville) ne donnerait pas le même modèle associé pour B et C.

Une alternative à cette démarche est de proposer une vision plus globale de l'opération en généralisant la notion de point d'intérêt à l'ensemble des interactions possibles entre deux modèles. Cette démarche est nommée compilation.

2.2.2. Compilation

L'approche par compilation se propose de rapprocher deux modèles de données quel que soit le nombre de points d'intérêt aux deux modèles. La proposition de (Stell, Worboys,1998) suit cette philosophie. Deux niveaux peuvent être définis pour la manipulation de modèle. (a) Une approche par inter-opérabilité prend plusieurs modèles et essaie de les rapprocher. (b) Une approche par schémas externes, définis sur un même modèle de référence, sert de support à ce rapprochement.

La première approche soulève la problématique plus générale du multi-sources et de l'alignement ontologique nécessaire qui en découle. De nombreux travaux ont étudié cette question, notamment ceux de (Cruz *et al.*,2004 ;Zhao *et al.*,2008). Ces travaux évoquent les problèmes qui se posent lorsque l'on tente d'aligner deux ontologies provenant de différentes sources (e.g. homonymie, synonymie, un concept dans une ontologie couvrant un ensemble de concept dans l'autre). (Cruz *et al.*,2004) propose des solutions de mapping semi-automatique entre ces ontologies en caractérisant par exemple la qualité du rapprochement entre deux concepts (exact, approchant, aucun lien sémantique).

Nous supposons donc que dans le cas d'inter-opérabilité, cette étape d'alignement a été réalisé et nous pouvons nous focaliser sur les manipulations de modèles dans un cadre unifié.

L'approche par compilation va donc généraliser la notion de signature publique de l'opérateur de $Collage$ de l'approche interprétation par :

$$Collage : \text{Modèle1 x Modèle2} \rightarrow \text{ModèleRésultat}$$

2.3. Synthèse

Le modèle algébrique proposé section 2.1.1 permet de définir le positionnement d'une approche dans le cadre de la construction de modèles par ajout ou non d'information lors de l'utilisation d'opérateurs spatiaux. L'approche définie par (Stell, Worboys,1998) s'appuie sur plusieurs niveaux de granularité mais ne transfère pas dynamiquement d'information d'un niveau à l'autre puisqu'elle est entièrement compilée.

En s'appuyant sur le modèle de Hermann et al, celui de Stell, et en prenant une approche interprétée (*i.e.* l'évaluation du modèle résultat a lieu à la demande), il devient possible de définir des opérateurs qui peuvent être utilisés dans une philosophie proche de celle proposée par Aiken *et al*. (par intégration dynamique d'artéfacts) pour apporter des informations lors d'une variation de ces niveaux.

Aiken *et al*. laisse la responsabilité de la cohérence du modèle résultat à l'utilisateur. La définition de règles cohérentes de transfert d'information d'un niveau d'abstraction à un autre apporte donc une réelle plus value lorsque son application est (semi-)automatique.

En nous appuyant ici aussi sur la notion de graphe pour être homogène avec la présentation de l'opérateur de $ZoomIn$ [notreRef], et parce que les modèles de graphes sont particulièrement interessants dans ce contexte (Hoel *et al*.,2015), l'opérateur de $Collage$ peut se formaliser comme une fusion de modèle.

3. Principe de l'opérateur de collage

Dans cette partie nous présentons les éléments structurant l'opérateur de $Collage$ en définissant le contexte d'application avec un modèle de référence jouet sur lequel s'appuieront nos exemples, le principe du collage et la présentation du formalisme.

3.1. Contexte

Soit un modèle de référence m (qui peut-être issu de différents modèles pour lesquels l'alignement ontologique a été réalisé). Ce modèle est représenté sous la forme d'un graphe $\mathcal{G} = (N, E, \psi, \nu, \epsilon)$ orienté, avec N l'ensemble des nœuds, E l'ensemble des arcs, ψ la fonction d'incidence qui associe nœuds et arcs, ν(resp. ϵ) représente la fonction d'étiquetage des nœuds (resp. arcs), gérée de manière ensembliste. Soit $\mathfrak{g}$ un sous-graphe de $\mathcal{G}$ (*i.e.* une vue de m ou un modèle m_i de la section 2.1.1), sans circuit dans lequel des nœuds et des arcs sont potentiellement supprimés vis à vis du graphe initial (e.g. figure 3a). Dans ce cas, il existe une perte d'information (en terme d'attributs) entre $\mathcal{G}$ et $\mathfrak{g}$. Pour pallier ce problème nous avons proposé dans [notreRef] une étape de propagation des attributs qui doit avoir lieu avant la création de la vue (figure 3b). Cette étape est réalisée en respectant un ensemble de règles $\mathbb{R}_P$ définies par la relation entretenue entre les nœuds, autrement dit, l'étiquetage du chemin entre deux nœuds ($\mathtt{ch}_i$). Cet étiquetage est obtenu en calculant une expression

régulière qui est associée à ce chemin. Ces règles permettent de décider si un attribut est propagé d'un nœud à l'autre, et de modifier éventuellement sa caractéristique (*i.e.* SUBSET, GLOBAL, FUZZY) le cas échéant. Pour mémoire, un attribut est caractérisé de GLOBAL si sa validité sémantique concerne l'ensemble de la représentation spatiale initialement disponible (e.g. une population définie en extension pour une ville). Un attribut est caractérisé de SUBSET si sa validité sémantique concerne n'importe quel sous-ensemble de la représentation spatiale initialement disponible (e.g. le nom du maire d'une ville). Enfin, un attribut est caractérisé de FUZZY si sa pertinence vis à vis d'une représentation spatiale n'est pas assurée. Sa validité sémantique est associée à une fonction d'appartenance $\mathcal{A}$ définie dans $[0, 1]$. La propagation permet non seulement le transfert des attributs, mais elle permet aussi au travers de la modification de leurs caractéristiques, de garantir leur cohérence avec le nouveau nœud où ils sont propagés. Cette étape réalisée, l'application d'opérateurs est possible.

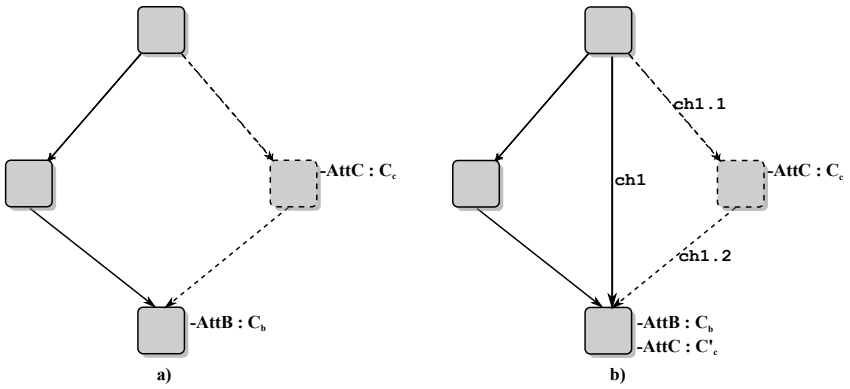

*Figure 3. Suppression d'un nœud dans $\mathcal{G}$ a) sans propagation, l'attribut **AttC** de caractéristique $\boldsymbol{C}_c$ est perdu b) avec propagation, **AttC** est conservé (respect des règles $\mathbb{R}_p$); il change éventuellement de caractéristique ($\boldsymbol{C}'_c$)*

Nous prenons comme domaine applicatif pour illustrer l'article la gestion administrative de ressources culturelles. Le modèle de référence est présenté figure 4. Les rôles des associations, lorsqu'ils sont utilisés dans l'article, sont précisés sur les arcs entre crochets. Les étiquettes associées aux arcs (ϵ) modélisent trois relations : l'héritage, l'inclusion spatiale (stricte) et la cardinalité. Ces trois relations sont respectivement représentées par l'étiquette des arcs avec H, $\circledcirc$ et $1..*$.

Dans nos précédents travaux nous avions défini l'opérateur de $ZoomIn$ qui s'apparentait à l'opérateur de projection (Π) de l'algèbre relationnelle (figure 5a). En résultat de son application à un graphe $\mathfrak{g}_1$, un sous-graphe $\mathfrak{g}_2$ respectant la structure (*i.e.* les classes) de projection choisie est obtenu. Grâce à l'étape de propagation, $\mathfrak{g}_2$ contient un maximum d'information y compris celles qui appartenaient au(x) nœud(s) supprimé(s) par l'application de l'opérateur de $ZoomIn$. Nous présentons maintenant un nouvel opérateur, l'opérateur de $Collage$ qui permet de rapprocher deux modèles.

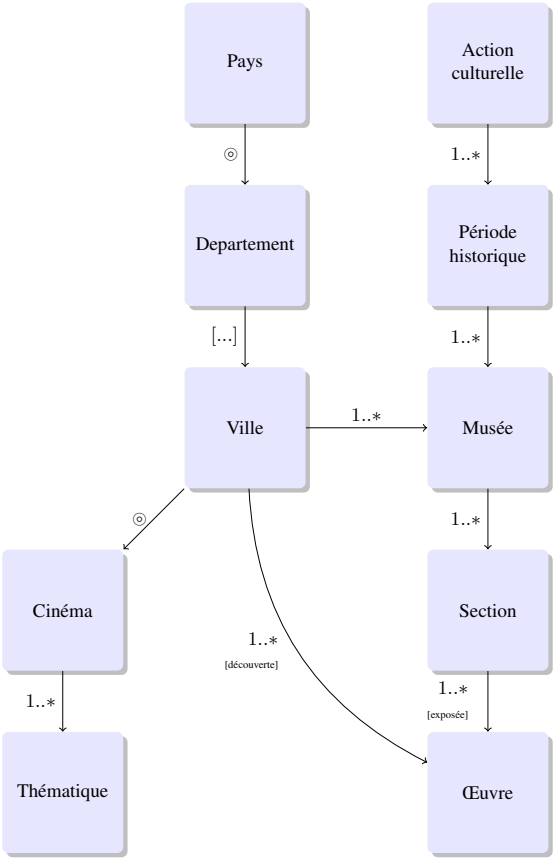

Figure 4. Modèle de référence

3.2. L'opérateur de $Collage$

L'opérateur de $Collage$ s'apparente à l'opérateur d'union ($\cup$) entre deux graphes (figure 5b). Nous travaillons au niveau du modèle, nous définissons cet opérateur comme binaire, symétrique, fermé (pour autoriser la composition d'opérations) et basé sur une approche ensembliste.

L'opérateur de $Collage$ permet d'ajouter un complément d'information qui serait nécessaire à l'analyse d'un phénomène.

Il existe deux familles de cas pour la spécification de cet opérateur : l'application dans le cadre d'un seul graphe de référence et l'application dans un contexte d'interopérabilité utilisant au moins deux graphes de références distincts mais alignés. Dans les faits, ces deux approches peuvent être généralisées.

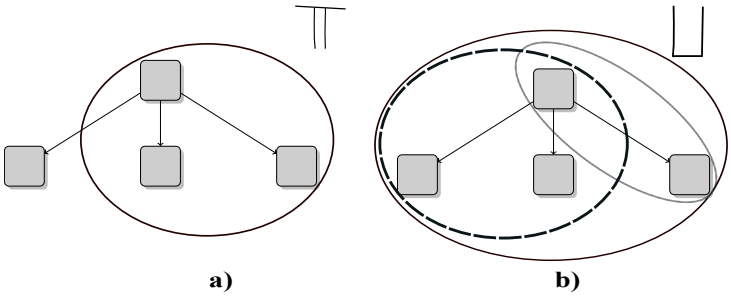

Figure 5. a) $ZoomIn$ (projection) b) $Collage$ (union)

3.2.1. Un unique modèle de référence

Dans cette famille, $\mathfrak{g}_1$ et $\mathfrak{g}_2$ sont issus d'un unique graphe $\mathcal{G}$ et leur appliquer l'opérateur de $Collage$ revient à rechercher dans $\mathcal{G}$ au moins un chemin qui relie un nœud du graphe $\mathfrak{g}_1$ à un nœud du graphe $\mathfrak{g}_2$. L'étiquetage du chemin est formalisé par une expression régulière composée des étiquettes des arcs formant le chemin. Deux types de cas sont à étudier : liés au positionnement (*i.e.* dépend des relations) et liés à la sémantique (*i.e.* choix des attributs et leur caractéristique). Le graphe $\mathcal{G}$ étant unique l'étape de propagation a déjà été effectuée lors de la création de $\mathfrak{g}_1$ et $\mathfrak{g}_2$ sur $\mathcal{G}$. La cohérence est donc garantie.

3.2.2. Deux modèles de référence

Cette famille rend possible l'application de l'opérateur de $Collage$ dans le cadre de la fusion d'informations provenant de modèles indépendants (e.g. inter-opérabilité) mais dont les ontologies ont été alignées. Dans ce cas il faut réaliser la liaison entre les deux graphes de référence en utilisant la signature suivante pour l'opérateur :

$$Collage(\mathcal{G}_1, \mathfrak{g}_1, \mathcal{G}_2, \mathfrak{g}_2) \rightarrow (\mathcal{H}, \mathfrak{h})$$

avec $\mathcal{H}$ résultat de l'union ontologique ($\cup_o$) entre $\mathcal{G}_1$, $\mathcal{G}_2$ respectivement graphes de référence de $\mathfrak{g}_1, \mathfrak{g}_2$ et $\mathfrak{h}$ le graphe résultat de l'opération de collage. Dans ce cas, l'étape de propagation réalisée une seule fois en amont dans le cas précédent doit être itérée dans $\mathcal{H}$ compte tenu du fait que les modèles $\mathcal{G}_1$ et $\mathcal{G}_2$ étaient initialement indépendants. Les étapes opérationnelles sont donc les suivantes :

- Calcul de $\mathcal{H} = \mathcal{G}_1 \cup_o \mathcal{G}_2$ (eq_1)
- Propagation via les règles $\mathbb{R}_P$ sur $\mathcal{H}$

3.2.3. Généralisation des deux familles : l'opérateur de $Collage$

Le graphe de référence $\mathcal{H}$ correspond, sur la base d'un unique modèle de référence ou de deux modèles, respectivement au graphe $\mathcal{G}$ ou à $\mathcal{G}_1 \cup_o \mathcal{G}_2$. Ainsi nous pouvons généraliser l'opérateur de $Collage$ entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ par :

$$Collage(\mathcal{H}, \mathfrak{g}_1, \mathfrak{g}_2) \rightarrow (\mathcal{H}, \mathfrak{h})$$

Avec $\mathfrak{h}$, graphe résultat du collage.

Les opérateurs définis sur la structure d'un graphe de référence sont fermés (*i.e.* le résultat de l'application d'un opérateur peut-être utilisé comme opérande d'un nouvel opérateur). L'intérêt de l'opérateur de $Collage$ est renforcé dans le cadre de la manipulation d'attributs en intention (*i.e.* dont la valeur au niveau d'une instance est le résultat d'une fonction de calcul). La définition d'un attribut calculé peut-être basée sur l'utilisation d'autres attributs et/ou fonctions spatiales type ISO 13249 (e.g. ST_Area). Des opérateurs qui diminuent le nombre d'attributs (e.g. $ZoomIn$) peuvent conduire un attribut à être classifié comme "figé". Les attributs permettant l'évaluation de sa fonction de calcul ne sont plus tous disponibles transformant la notion de vue en notion de cliché (*i.e.* la variation de valeur suite à une modification des opérandes devient impossible). Cette dernière peut potentiellement récupérer, via l'ajout d'information, les attributs nécessaires à son calcul.

L'opérateur de $Collage$ nous permet donc de passer d'une propriété PP de Hermann restreinte à l'égalité à une propriété PP sur la base de $\subseteq$. En effet, des attributs en intention dont la fonction de calcul n'étaient plus évaluables peuvent de nouveau devenir évaluables grâce à l'ajout d'attributs issus du collage. La classification de l'attribut en "figé" disparait donc. L'union ontologique de graphe ne fait pas disparaitre d'attribut par (eq_1).

La deuxième propriété de Hermann (FPP) n'est garantie qu'en l'absence d'attributs en intention puisque l'indépendance n'est plus garantie dans le cas d'une fonction de calcul utilisant des attributs provenant des deux modèles. La présence d'un attribut dans le modèle $m2$ (*i.e.* l'ajout d'un attribut défini comme étant le seul manquant à l'évaluation d'un attribut en intention de $m1$) modifie l'interprétation de $m1$.

La troisième propriété de Hermann (CP) est garantie puisqu'établie sur des modèles initiaux cohérents. Les attributs n'intervenant pas dans les fonctions de calcul maintiennent l'indépendance des modèles et ceux intervenants dans les fonctions de calcul, soit ne changent rien car la fonction de calcul est toujours non évaluable, soit la rendent évaluable mais cohérente par la propriété PP.

Autrement dit, si l'opérateur de $ZoomIn$ peut rendre des attributs figés car l'ensemble des paramètres nécessaires à la fonction de calcul ne sont plus présents, par l'opérateur de $Collage$ ces paramètres peuvent devenir disponibles et ainsi rendre la mise à jour de l'attribut à nouveau possible.

3.3. Présentation du formalisme

L'opérateur de $Collage$ nécessite la définition de règles d'étiquetage des nœuds et des arcs de $\mathfrak{h}$ (fonction de l'étiquetage et des caractéristiques des attributs dans le modèle de référence associé).

Cet article s'attache à la description des cas de collage entre deux graphes $\mathfrak{g}_1 = (N_1, E_1, \psi_1, \nu_1, \epsilon_1)$, $\mathfrak{g}_2 = (N_2, E_2, \psi_2, \nu_2, \epsilon_2)$ selon deux hypothèses. Nous illustrons ces hypothèses à partir du modèle de référence présenté figure 4, et les figures associées à leur description utilisent la légende définie dans la figure 6. Dans la première hypothèse, ces graphes sont indépendants c'est à dire que $\mathfrak{g}_1 \cap \mathfrak{g}_2 = \emptyset$. Dans la deuxième hypothèse, soit il existe une chaine entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ dans $\mathcal{H}$ (*i.e.* d'un ancêtre commun à) et l'absence de chemin entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ ou $\mathfrak{g}_2$ et $\mathfrak{g}_1$ (*i.e.* le graphe est quasi-fortement connexe) soit il existe au moins un chemin entre un nœud de $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) avec un nœud de $\mathfrak{g}_2$ (resp. $\mathfrak{g}_1$) dans $\mathcal{H}$.

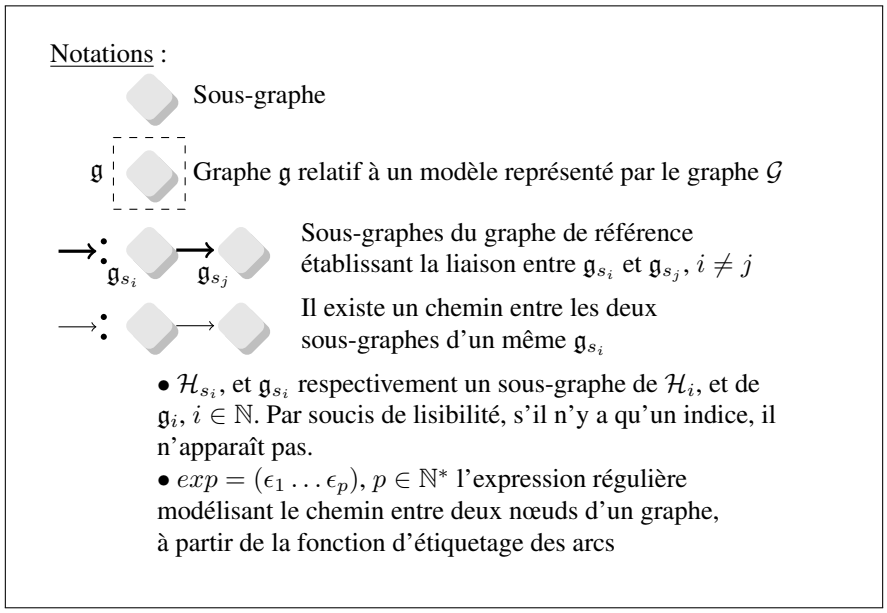

Figure 6. Légende des figures

4. Graphes indépendants

En présence de graphes indépendants, deux configurations vis à vis du modèle de référence $\mathcal{H}$ sont envisageables. Dans la première configuration, il existe un ancêtre commun de $\mathfrak{g}_1$ et $\mathfrak{g}_2$, dans la seconde il n'en existe pas.

4.1. Graphes indépendants : lien conceptuel entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ [C1]

[C1] est construit sous les hypothèses de la figure 7a). Ce cas illustre l'opérateur de $Collage$ entre deux graphes (*i.e.* $\mathfrak{g}_1$ et $\mathfrak{g}_2$) indépendants mais rattachés à un concept commun dans le graphe de référence (ils ont donc conceptuellement un lien). Dans l'exemple présenté dans la figure 7b), l'opération de collage est appliquée entre le graphe $\mathfrak{g}_1$ (qui comprend les nœuds **Cinéma** et **Thématique**) et le graphe $\mathfrak{g}_2$ (qui

comprend les nœuds **Musée**, **Section** et **Œuvre**). Compte tenu des hypothèses (figure 7a)), la première étape consiste à rechercher un ancêtre commun dans $\mathcal{H}$ permettant de relier ces deux graphes (ici **Ville**). Pour simplifier la présentation le nœud **Ville** est relié à $\mathfrak{g}_1$ et $\mathfrak{g}_2$ par un arc mais la présence d'un chemin ne nuit pas à la généralisation de la gestion de [C1].

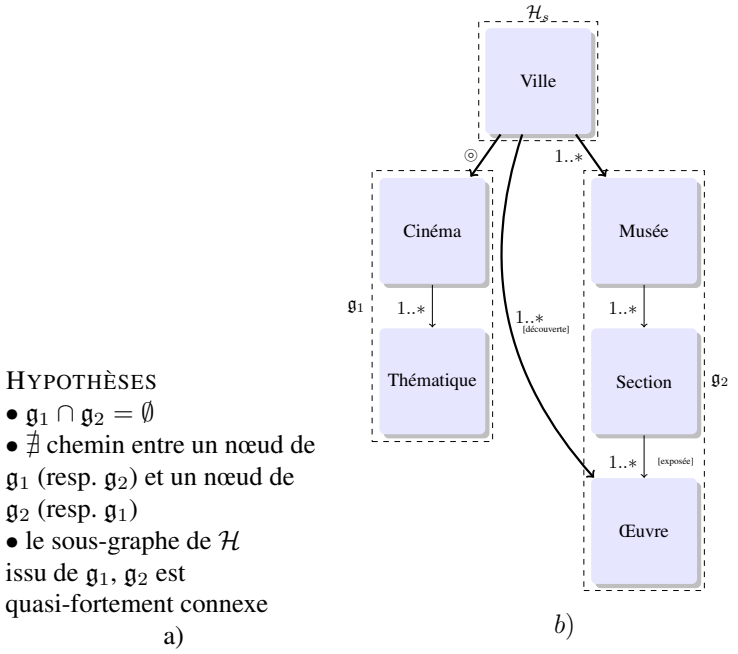

Figure 7. [C1] *: a) Hypothèses b) Exemple de sous-graphe de $\mathcal{H}$*

4.1.1. Le(s) nœud(s)

L'objectif est de déterminer dans le graphe du modèle de référence, $\mathcal{H}$, le concept partagé appelé nœud ancêtre entre au moins un nœud provenant du graphe $\mathfrak{g}_1$ et un nœud provenant du graphe $\mathfrak{g}_2$.

Soit la fonction $lg : \mathcal{G} \times N \times N \rightarrow \mathbb{N}$ ici le nombre d'arc d'un chemin entre deux nœuds dans un graphe (sa valeur est 0 si le chemin n'existe pas). Dans le cas général lg peut être basée sur des métriques différentes du nombre d'arcs suivant le contexte, il s'agit donc d'une fonction paramétrable suivant le domaine traité (e.g. réseaux sociaux, analyse spatiale).

Cela revient à chercher un sous-graphe quasi-fortement connexe de $\mathcal{H}$. Un nœud vérifiant cette propriété et minimisant la distance (en nombre d'arcs selon la définition retenue pour lg) n'est pas forcément unique compte tenu du fait que le graphe de référence n'est pas forcément un arbre. Les nœuds ancêtres sont donc définis par un ensemble nommé A (figure 8).

$A = \{a \in N_{\mathcal{H}}\ a \notin N_1, a \notin N_2\ \wedge\ \exists\, n_1 \in N_1 | \min_{n_i \in N_1} lg(\mathcal{H}, a, n_i) > 0 \wedge\ \exists\, n_2 \in N_2 | \min_{n_i \in N_2} lg(\mathcal{H}, a, n_i) > 0\}$. Sur la figure 8, $A = \{a_1, a_2\}$.

Soit D1 (resp. D2) l'ensemble des nœuds d'entrée dans le graphe $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) venant d'un nœud de A (*i.e.* D_1 (resp. D_2) est l'ensemble des nœuds qui sont en liaison avec (au moins) un des ancêtres communs). Comme il peut exister des longueurs minimales équivalentes entre un même ancêtre et plusieurs nœuds de $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) cet ensemble lui aussi n'est pas obligatoirement réduit à un singleton (figure 9).

$$D_1 : \quad \{n_i \in N_1 | \min_{n_i \in N_1} lg(\mathcal{H}, a, n_i) > 0, a \in A\}$$
$$D_2 : \quad \{n_i \in N_2 | \min_{n_i \in N_2} lg(\mathcal{H}, a, n_i) > 0, a \in A\}$$

Par exemple sur la figure 9, $D_1 = \{n_1, n_{11}\}$ car $lg(\mathcal{H}, a_1, n_1) = lg(\mathcal{H}, a_1, n_{11})$, et concrètement sur l'exemple de la figure 7b) $A = \{Ville\}$, $D_1 = \{\text{Cinema}\}$ et $D_2 = \{\text{Musée}, \text{Œuvre}\}$.

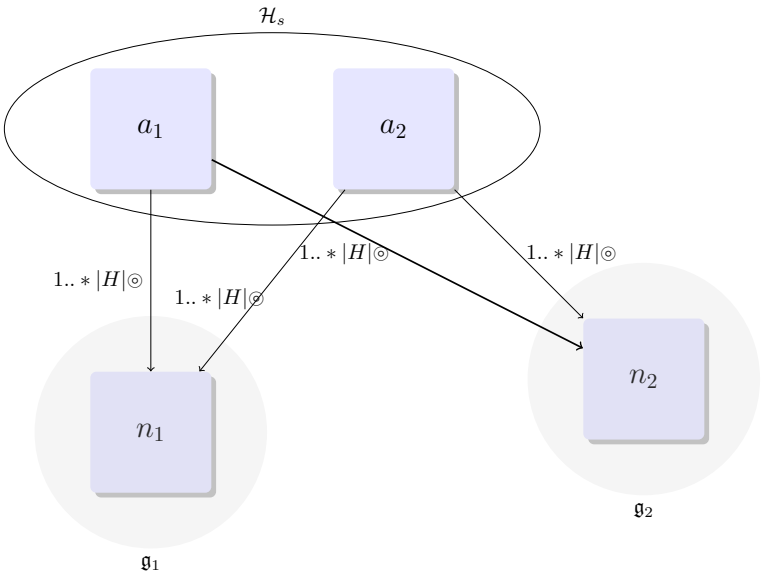

Figure 8. $A = \{a_1, a_2\}$ ensemble des ancêtres non réduit à un singleton

Les concepts associés aux nœuds présents dans A peuvent avoir l'inconvénient si la longueur lg est trop importante, de ne pas apporter d'information pertinente vis à vis du collage réalisé. Par exemple si nous essayons de rapprocher des informations concernant les cinémas d'une ville et celles liées à des musées, des informations sur l'Union Européenne qui se trouverait être le niveau d'abstraction de l'ancêtre commun ne sont pas forcément pertinentes. La fonction lg fournit donc une mesure de proximité des graphes impliqués dans l'opération de collage, elle informe sur le degré de corrélation entre les concepts présents dans $\mathfrak{g}_1$ et $\mathfrak{g}_2$.

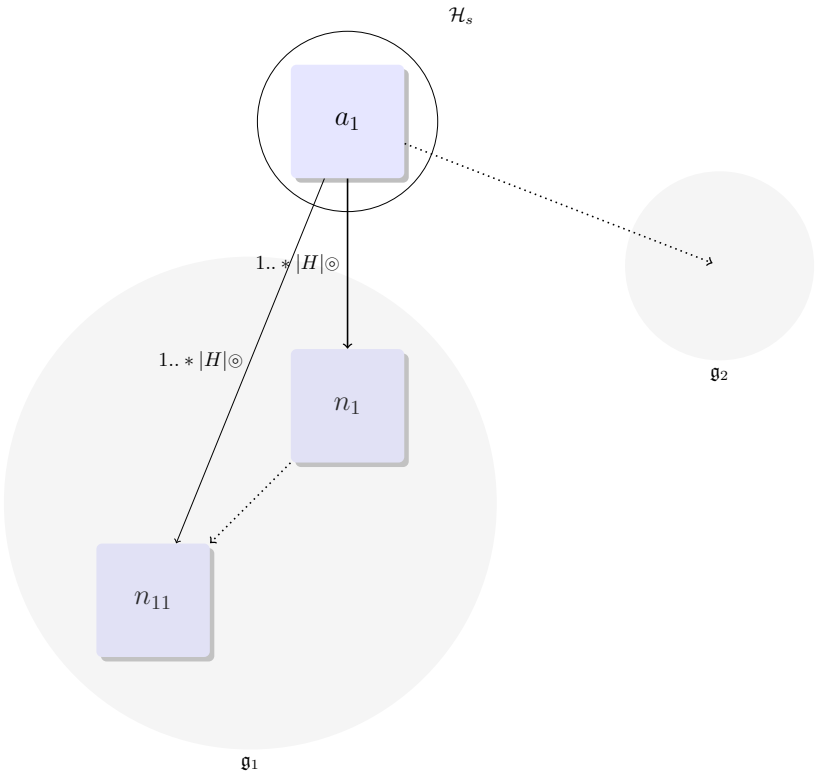

Figure 9. $D_1 = \{n_1, n_{11}\}$ *ensemble des nœuds entrants de* $\mathfrak{g}_1$ *non réduit à un singleton*

4.1.2. Les arcs : définition

La liaison entre le graphe $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) et cet ancêtre commun (noté ici a) peut être sur la base d'un arc, établissant une liaison Directe (D), ou via un chemin, établissant une liaison Indirecte (I). Le produit cartésien des possibilités nous définit donc DD, DI, ID, II dans $\mathcal{H}$. En présence de forêt la liaison directe est toujours privilégiée (notamment pour déterminer l'ancêtre commun). Soient $a \in A, n_1 \in D_1 et\ n_2 \in D_2$, la définition des liaisons possibles est la suivante :

DD : $lg(\mathcal{H}, a, n_1) = lg(\mathcal{H}, a, n_2) = 1$
DI : $lg(\mathcal{H}, a, n_1) = 1 \ \wedge \ lg(\mathcal{H}, a, n_2) > 1$
ID : $lg(\mathcal{H}, a, n_1) > 1 \ \wedge \ lg(\mathcal{H}, a, n_2) = 1$
II : $lg(\mathcal{H}, a, n_1) > 1 \ \wedge \ lg(\mathcal{H}, a, n_2) > 1$

La figure 8 présente le cas DD en illustrant les différents types d'étiquetage pour les arcs issus de a_1 : $e_1 = (a_1, n_1)|n_1 \in D_1$ et $e_2 = (a_1, n_2)|n_2 \in D_2$. La figure 10 illustre les cas II et ID (resp. DI), l'arc en pointillé représente un chemin (*i.e.* une succession d'arcs). Pour simplifier la présentation nous illustrons les différentes configurations à l'aide d'une liaison unique (la présence d'une forêt n'est

que l'application itérée des principes présentés).

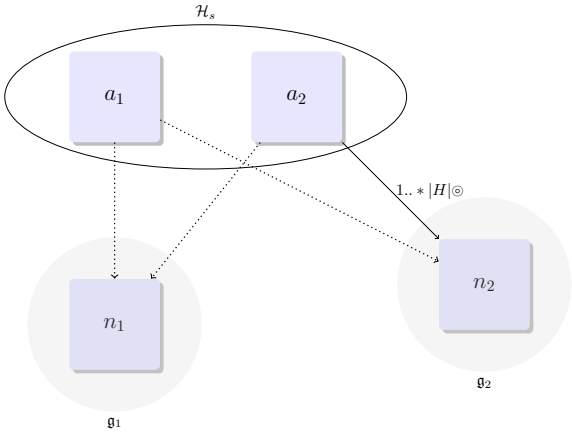

Figure 10. Illustration de liaisons II et ID

4.1.3. Les arcs : étiquetage

Quel que soit le sous-cas considéré (DD, DI, ID II) une expression régulière comme définie figure 6 représente l'étiquetage du chemin entre A et D_i. Les chemins (éventuellement réduits à un arc) sont traités indépendamment les uns des autres car $\mathfrak{g}_1$ et $\mathfrak{g}_2$ sont indépendants (une action sur $\mathfrak{g}_1$ ne peut pas avoir de répercussion sur $\mathfrak{g}_2$ et réciproquement car $\mathfrak{g}_1 \cap \mathfrak{g}_2 = \emptyset$, par définition). Le traitement de DI est donc similaire à l'application du D de DD pour la liaison directe et du I de II pour la liaison indirecte. Le principe de base de construction de $\mathfrak{h}$ est un graphe issu des graphes $\mathfrak{g}_1$ et $\mathfrak{g}_2$ auquel sont adjoints les nœuds de A avec une liaison définie à l'aide d'un arc (notion ensembliste de l'opérateur de *Collage*). Dans tous les cas de figure il convient de définir l'étiquetage de cet arc (a_i, n_j) où $a_i \in A$ et $n_j \in D_k, k \in \{1, 2\}$. La définition de l'étiquetage est trivial dans le cas d'un arc direct, l'arc est directement issu (origine, extrémité et étiquetage) du graphe de référence $\mathcal{H}$. Dans le cas où le lien n'est pas direct (chemin) une expression régulière correspondant aux étiquettes modélise le chemin. Les règles de calcul ($\mathbb{R}_{et}$) sont présentées table 1. Ces règles signifient que si l'expression régulière traitée de manière ensembliste et modélisant le chemin ne contient que des inclusions spatiales $\circledcirc$, éventuellement accompagnées de H alors l'étiquette calculée pour l'arc correspondant sera $\circledcirc$. En présence uniquement de H dans l'expression régulière, l'étiquette de l'arc résultat restera H. Enfin, dès que le symbole $1..*$ apparaît dans une expression régulière alors l'étiquette de l'arc résultat sera $1..*$.

$\{\circledcirc\} \vee \{H, \circledcirc\}$	$\rightarrow$	$\circledcirc$
$\{H\}$	$\rightarrow$	H
$\{1..*\} \vee \{1..*, \circledcirc\} \vee \{1..*, H\} \vee \{1..*, H, \circledcirc\}$	$\rightarrow$	$1..*$

TABLE *1. Règles d'étiquetage ($\mathbb{R}_{et}$)*

4.1.4. Graphe résultat : définition

Le graphe résultat (*i.e.* $\mathfrak{h}$) est défini de la manière suivante :

– $N_{\mathfrak{h}} = N_1 \cup N_2 \cup A$

– $E_{\mathfrak{h}} = E_1 \cup E_2 \cup \{e_q = (a_i, n_j) | q \in \mathbb{N}, q > |E_1| + |E_2|, a_i \in A \; et \; n_j \in D_k \; avec \; k \in \{1, 2\}$

Dans le cas d'une liaison indirecte (I), $lg(\mathcal{H}, a_i, n_j) > 1$, ce qui correspond au fait que n_j n'appartient pas aux successeurs directs de a_i, $\epsilon(e_q)$ est définie à partir de l'expression régulière $exp = (\epsilon_1 \ldots \epsilon_p)$, $p \in \mathbb{N}$, selon les règles définies dans la table 1 par :

$\epsilon(e_q) = \circledcirc$ si $\forall \epsilon_l \in exp, \; 1 \leq l \leq p, \; \epsilon_l = (\circledcirc \vee H) \wedge \exists \, l | e_l = \circledcirc$

$\epsilon(e_q) = H$ si $\forall \epsilon_l \in exp, \; 1 \leq l \leq p, \; \epsilon_l = H$

$\epsilon(e_q) = 1..*$ si $\exists \epsilon_l \in exp, \; 1 \leq l \leq p \; | \epsilon_l = 1..*$

4.1.5. Graphe résultat : gestion des redondances potentielles

Deux types de redondances peuvent survenir : celles liées à la construction du graphe résultat et celles liées à l'application des règles de propagation.

Lors de la construction de l'ensemble d'arcs $E_{\mathfrak{h}}$, dès qu'il existe au moins deux chemins dans $\mathcal{H}$ tels que $lg(\mathcal{H}, a_i, n_j)$ sont identiques, autant d'arcs (a_i, n_j) sont générés (figure 11). Soit ces arcs possèdent la même étiquette et dans ce cas, $E_{\mathfrak{h}}$ supprime le doublon par propriété ensembliste, soit les étiquettes sont différentes et il faut appliquer les règles dans le tableau 2 afin de choisir l'étiquette à conserver pour l'arc ajouté à $E_{\mathfrak{h}}$. L'idée est de maintenir le maximum d'information à l'utilisateur, $1..*$ (resp. H) est moins précis que $\circledcirc$ qui impose l'inclusion spatiale alors que $1..*$ impose seulement une mise en relation (resp. l'information de sous-classe H, or nous privilégions la relation spatiale dans nos travaux).

Tableau 2. Choix de l'étiquette à conserver

arc (origine, destination, étiquette)	$(n_o, n_d, \circledcirc)$	$(n_o, n_d, 1..*)$	(n_o, n_d, H)
$(n_o, n_d, \circledcirc)$	-	$\circledcirc$	$\circledcirc$
$(n_o, n_d, 1..*)$		-	H
(n_o, n_d, H)			-

Lors de l'application des règles de propagation effectuée sur $\mathcal{H}$, il est possible que l'application de l'opérateur de $Collage$ génère des redondances au niveau du modèle puisque $\mathfrak{g}_1$ et $\mathfrak{g}_2$ sont indépendants. Dans ce cas, afin de les faire disparaitre, nous

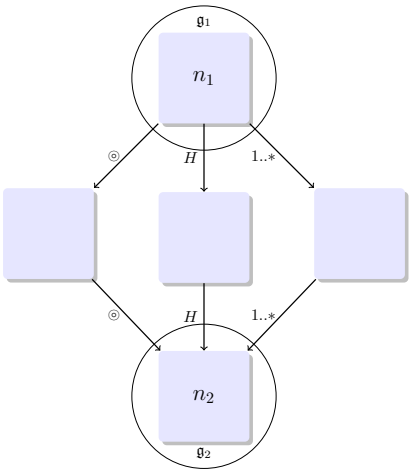

Figure 11. Cas des arcs multiples

appliquons l'algorithme 1 sur chacun des nœuds de A. Le principe de l'algorithme est : pour les nœuds de D_1 et D_2 reliés à un ancêtre dans A par un arc d'étiquette H, en présence d'un même attribut (dans chacun de ces nœuds) cet attribut est remonté dans l'ancêtre et supprimé dans les nœuds (de D_1 et D_2). Il n'est pas envisageable d'effectuer une telle manipulation pour une autre étiquette que H car il existe un risque de dégradation de la caractéristique de l'attribut (GLOBAL vs. SUBSET).

Soit $D'_1 = \{n \in D_1 | \epsilon_{\mathfrak{h}}(e(a,n)) = H\}$
Soit $D'_2 = \{n \in D_2 | \epsilon_{\mathfrak{h}}(e(a,n)) = H\}$
`/*`D'_1 `(resp.` D'_2`) : ensembles des nœuds de` D_1 `(resp.` D_2`) tels que l'étiquette de l'arc entre un ancêtre` $a \in A$ `et un nœud de cet ensemble soit` H `dans le graphe*/`
`/*`D'_2 `: inutile de le traiter car` a_{tt} `doit aussi appartenir à` D'_1`, d'où :*/`
pour chaque attribut a_{tt} **de** $\nu(n) | n \in D'_1$
  `/*si l'attribut est présent dans tous les nœuds de` D'_1 `(resp.` D'_2`)*/`
  si $(\forall n_1 \in D'_1, a_{tt} \in \nu(n_1) \wedge \exists e_1 = (a, n_1) \in E_{\mathfrak{h}}) \wedge (\forall n_2 \in D'_2, a_{tt} \in \nu(n_2) \wedge \exists e_2 = (a, n_2) \in E_{\mathfrak{h}})$ **alors**
    `/*on peut supprimer` a_{tt} `dans les nœuds fils, et le remonter dans l'ancêtre` a`*/`
    $\nu(n_1) = \nu(n_1) - a_{tt}$
    $\nu(n_2) = \nu(n_2) - a_{tt}$
    $\nu(a) = \nu(a) \cup a_{tt}$
  finsi
finpour

Algorithme 1 – Gestion des redondances : Nettoyage($\mathfrak{h}, a, D_1, D_2$)

4.2. Graphes indépendants : aucun lien conceptuel entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ **[C2]**

[C2] est construit sous les hypothèses définies dans la figure 12a). Ce cas illustre l'opérateur de $Collage$ entre deux graphes indépendants. Dans l'exemple présenté figure 12b), cet opérateur est appliqué entre le graphe $\mathfrak{g}_1$ (qui comprend les nœuds **Pays** et **Département**) et le graphe $\mathfrak{g}_2$ (qui comprend les nœuds **Action culturelle** et **Période historique**). Sous les hypothèses de la figure 12a), il n'existe pas dans le graphe du modèle de référence $\mathcal{H}$ un concept partagé entre au moins un nœud provenant du graphe $\mathfrak{g}_1$ et un nœud provenant du graphe $\mathfrak{g}_2$. Le traitement ne peut donc pas être complètement automatisé puisque nous ne disposons d'aucune information endogène permettant de lier un graphe $\mathfrak{g}_1$ à un graphe $\mathfrak{g}_2$. Conceptuellement et pris individuellement, l'opérateur de $Collage$ n'impose pas la connexité du graphe résultat. Nous recherchons la propriété de fermeture des opérateurs et l'opérateur de $ZoomIn$ nécessite un graphe connexe, nous allons donc imposer l'obtention d'un graphe connexe. L'opérateur de *Collage* a pour objectif de faire apparaître un lien entre ces deux graphes. Pour gérer ce cas, nous proposons une phase automatique permettant d'obtenir un graphe résultat cohérent, puis une phase d'interactivité avec l'utilisateur permettant de préciser des informations sémantiques supplémentaires.

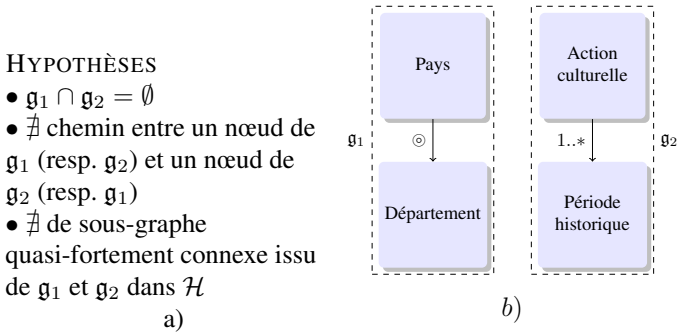

Figure 12. [C2] *: a) Hypothèses b) Exemple de sous-graphe de $\mathcal{H}$*

4.2.1. Le(s) nœud(s)

Notre proposition est d'ajouter un nœud jouant ce rôle de "concept partagé", soit n_o ce nœud, et $A = \{n_o\}$. Ainsi, $N_{\mathfrak{h}} = N_1 \cup N_2 \cup A$.

Selon les besoins de l'utilisateur (phase d'interactivité), ce nœud peut-être intégré de deux façons. Soit il permet de définir une étude au niveau macro (figure 13a), soit au niveau micro (figure 13b). Au niveau macro nous opérons via les racines du graphe $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$). Ces racines existent obligatoirement puisque le graphe est un graphe sans circuit, les ensembles de ces nœuds racines sont définis par $D_{r1} = \{n \in N_1 | \Gamma^-(n) = \emptyset\}$ dans $\mathfrak{g}_1$ et $D_{r2} = \{n \in N_2 | \Gamma^-(n) = \emptyset\}$ dans $\mathfrak{g}_2$. Dans cet exemple figure 13b) $D_{r1} = \{\text{Pays}\}$ et $D_{r2} = \{\text{Action culturelle}\}$. Un exemple de niveau macro ajouté au modèle de la figure 13 a) serait une entité représentant une manifestation d'ampleur

internationale afin de justifier le lien entre **Pays** et **Action culturelle** (e.g. l'exposition universelle).

Au niveau micro, nous opérons au niveau des feuilles du graphe $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$). Ces feuilles existent obligatoirement pour les mêmes raisons qu'au niveau macro et sont définies dans deux ensembles $D_{f1} = \{n \in N_1 | \Gamma^+(n) = \emptyset\}$ dans $\mathfrak{g}_1$ et $D_{f2} = \{n \in N_2 | \Gamma^+(n) = \emptyset\}$ dans $\mathfrak{g}_2$. Dans cet exemple figure 13a) $D_{f1} = \{$Département$\}$ et $D_{f2} = \{$Période historique$\}$. Un exemple de niveau micro ajouté au modèle de la figure 13 b) serait une entité représentant des évènements locaux comme par exemple les célébrations de l'anniversaire du débarquement (**Période historique** : seconde guerre mondiale) en Normandie (**Département** : Calvados). Quel que soit le niveau (macro ou exclusif micro), l'utilisateur précisera les attributs qu'il souhaite affecter au nœud, la visibilité de ces attributs (*i.e.* publique ou privée) et leurs caractéristiques (*i.e.* GLOBAL, SUBSET OU FUZZY) [1].

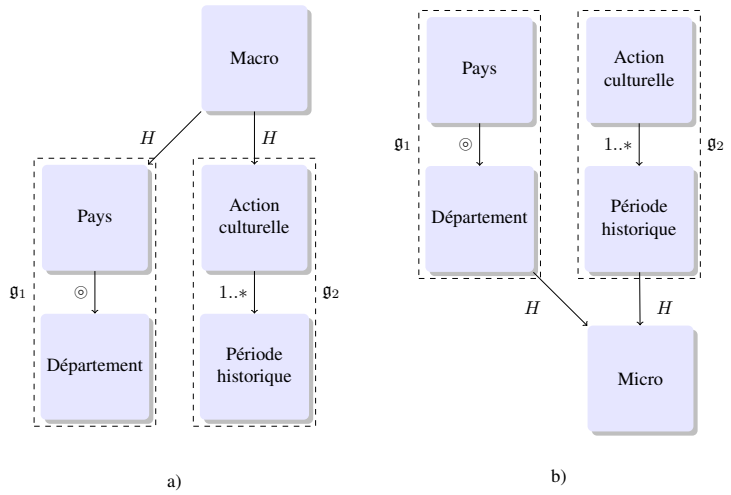

Figure 13. [C2] *: deux possibilités liés à ce collage a) Niveau macro b) Niveau micro*

4.2.2. Les arcs : définition

Afin de garantir la connexité du résultat, il faut établir les liaisons entre le graphe $\mathfrak{g}_1$ et le graphe $\mathfrak{g}_2$. Nous définissons des arcs d'étiquette H de n_o vers chacun des nœuds des ensembles D_{r1} et D_{r2} pour le niveau macro (resp. de D_{f1} et D_{f2} vers n_o pour le niveau micro).

4.2.3. Graphe résultat : définition

Le graphe résultat (*i.e.* $\mathfrak{h}$) est défini de la manière suivante :

1. Seuls les attributs de visibilité publique pourront être propagés, la caractéristique attribuée par défaut est GLOBAL (la plus restrictive)

– $N_{\mathfrak{h}} = N_1 \cup N_2 \cup A$

– $E_{\mathfrak{h}} = E_1 \cup E_2 \cup \{e_q = (n_o, n_j) | q \in \mathbb{N} \wedge q > |E_1| + |E_2|,\ n_j \in D_{rk}\ avec\ k \in \{1, 2\}\}$ pour le niveau macro (resp. $e_q = (n_j, n_o) | q \in \mathbb{N} \wedge q > |E_1| + |E_2|,\ n_j \in D_{fk}\ avec\ k \in \{1, 2\}$ pour le niveau micro)

et $\epsilon(e_q) = H$

4.2.4. Graphe résultat : gestion des redondances potentielles

Compte tenu de l'ajout de ces informations qui au préalable n'existaient pas, il ne peut y avoir de redondances au niveau micro. Pour le niveau macro, des redondances peuvent apparaître puisque $\mathfrak{g}_1$ et $\mathfrak{g}_2$ sont indépendants. Nous sommes dans une configuration similaire à [C1]. A est défini par le nœud ajouté $A = \{n_o\}$. D_{r1} (resp. D_{r2}) est l'ensemble contenant les racines de $\mathfrak{g}_1$ (resp. de $\mathfrak{g}_2$). Dans cette configuration, nous pouvons appliquer l'algorithme 1, soit Nettoyage($\mathfrak{h}$, n_o, D_{r1}, D_{r_2}).

5. Graphes non indépendants

La perte d'indépendance peut-être matérialisée par deux configurations. La première configuration est le partage de nœud/arc entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$. La deuxième configuration est l'existence d'un chemin dans le graphe de référence $\mathcal{H}$, de $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) vers $\mathfrak{g}_2$ (resp. $\mathfrak{g}_1$).

5.1. Graphes non indépendants : partage de nœuds et/ou d'arcs [C3]

[C3] est construit sous les hypothèses de la figure 14a). Ce cas illustre l'opérateur de $Collage$ entre deux graphes $\mathfrak{g}_1$ et $\mathfrak{g}_2$ où il existe a minima un (des) nœud(s) (et/ou un (des) nœud(s) et un (des) arcs) qui appartiennent aux deux graphes. Dans l'exemple présenté figure 14b), l'opération de collage est appliquée entre le graphe $\mathfrak{g}_1$ (qui comprend les nœuds **Pays**, **Département** et **Ville**) et le graphe $\mathfrak{g}_2$ (qui comprend les nœuds **Ville**, **Musée** et **Section**). Ici seul le nœud **Ville** est en commun, d'autres contextes sont possibles comme celui proposé figure 14c) où un arc (*i.e.* e_1) et donc a minima deux nœuds (*i.e.* **Ville, Musée**) sont en commun entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$.

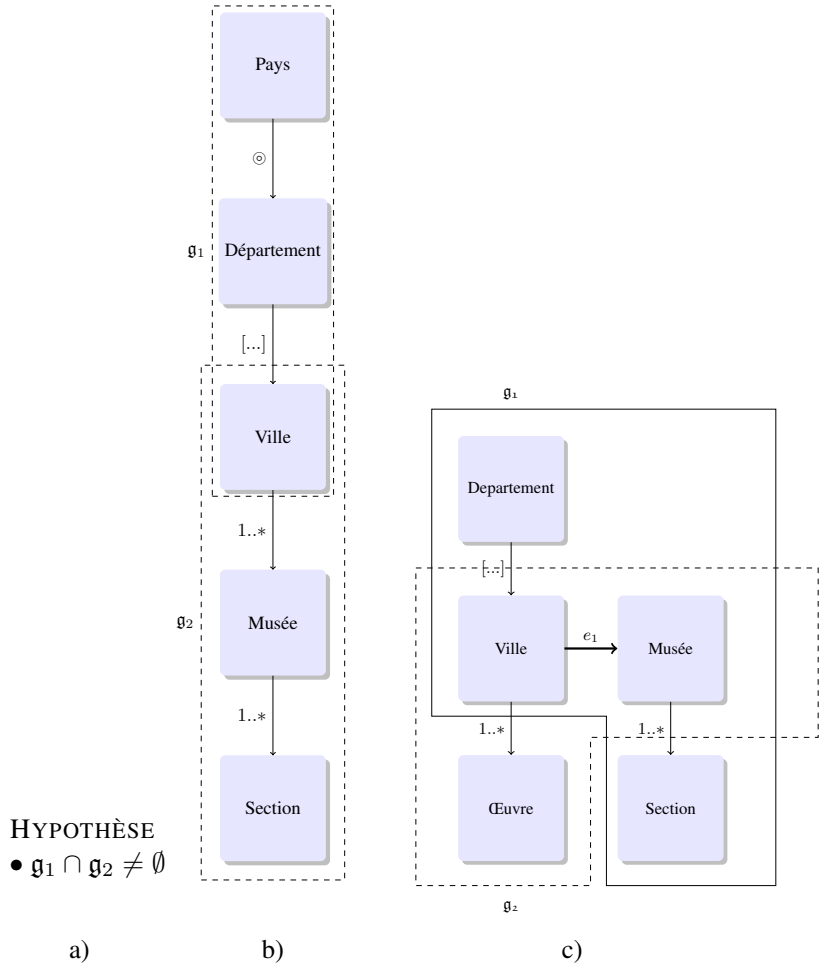

Figure 14. [C3] *: a) hypothèses b) nœud en commun c) arcs et nœuds en commun*

5.1.1. Le(s) nœud(s)

Selon les hypothèses de la figure 14a), il existe a minima un nœud appartenant aux deux ensembles N_1 et N_2. Les caractéristiques des attributs (de ces nœuds communs) peuvent être différentes (*i.e.* issues de l'application d'opérateurs conduisant à des résultats de propagation différents). Notre objectif est de maintenir le maximum d'information sur la spatialisation des attributs concernés (à l'image des choix dans le traitement des arcs multiples de [C1] tableau 2). Nous appliquons donc la règle explicitée dans le tableau 3 qui privilégie la caractéristique SUBSET par rapport aux deux autres (GLOBAL et FUZZY). La caractéristique GLOBAL est privilégiée par rapport à FUZZY.

Tableau 3. Choix de la caractéristique à privilégier en cas de redondance d'attribut dans une même entité

	SUBSET	GLOBAL	FUZZY
SUBSET	-	SUBSET	SUBSET
GLOBAL	SUBSET	-	GLOBAL
FUZZY	SUBSET	GLOBAL	-

5.1.2. Les arcs : définition

Selon les hypothèses de la figure 14a), les arcs issus ou arrivant sur le(s) nœud(s) commun(s) sont donc déjà présents, soit dans E_1, soit dans E_2, soit dans E_1 et E_2. Suivant la sémantique liée aux graphes $\mathfrak{g}_1$ et $\mathfrak{g}_2$, un arc, e_i, peut être présent dans les deux graphes mais avec un étiquetage différent. Nous sommes donc ramenés au même cas de gestion de la redondance de [C1], et nous le traitons de la même façon grâce à le tableau 2. A titre d'exemple, l'application des règles de la table 1 ou la composition d'opérations de type $ZoomIn$ peut conduire à une telle configuration.

5.1.3. Graphe résultat : définition

Le graphe résultat (*i.e.* $\mathfrak{h}$) est défini de la manière suivante :

- $N_{\mathfrak{h}} = N_1 \cup N_2$
- $E_{\mathfrak{h}} = E_1 \cup E_2$

5.1.4. Graphe résultat : gestion des redondances potentielles

Par construction du graphe résultat aucune redondance ne peut apparaître dans le modèle de données car l'union est un opérateur ensembliste et élimine donc automatiquement doubles (nœuds et arcs).

5.2. Graphes non indépendants : existence de chemin(s) entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$ **[C4]**

En présence de chemins entre le graphe $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) et $\mathfrak{g}_2$ (resp. $\mathfrak{g}_1$), et éventuellement réciproquement, une ambiguïté d'interprétation peut intervenir. En effet, deux chemins peuvent être définis à partir du même couple (classe origine, classe destination) sans qu'ils modélisent une sémantique conduisant aux mêmes instances. Les chemins entre **Ville** et **Œuvre** de la figure 4 sont un exemple de cette configuration. Une œuvre n'est pas forcément exposée dans la ville où elle a été découverte. Nous distinguons donc deux sous-cas : sans tenir compte des instances (niveau modèle) et en tenant compte des instances (niveau instance).

5.2.1. [C4]a : sans tenir compte des instances

[C4]a est construit sous les hypothèses de la figure 15a). Ce cas illustre l'opérateur de $Collage$ entre deux graphes $\mathfrak{g}_1$ et $\mathfrak{g}_2$ où il existe un chemin dans $\mathcal{H}$ entre eux, mais ils ne partagent pas de nœud et donc a fortiori d'arc. Dans l'exemple présenté figure

15b), l'opération de collage est appliquée entre le graphe $\mathfrak{g}_1$ (qui comprend les nœuds **Pays** et **Département**) et le graphe $\mathfrak{g}_2$ (qui comprend les nœuds **Musée** et **Section**). Dans cet exemple, il existe dans $\mathcal{H}$ un chemin entre le nœud **Département** de $\mathfrak{g}_1$ et le nœud **Musée** de $\mathfrak{g}_2$.

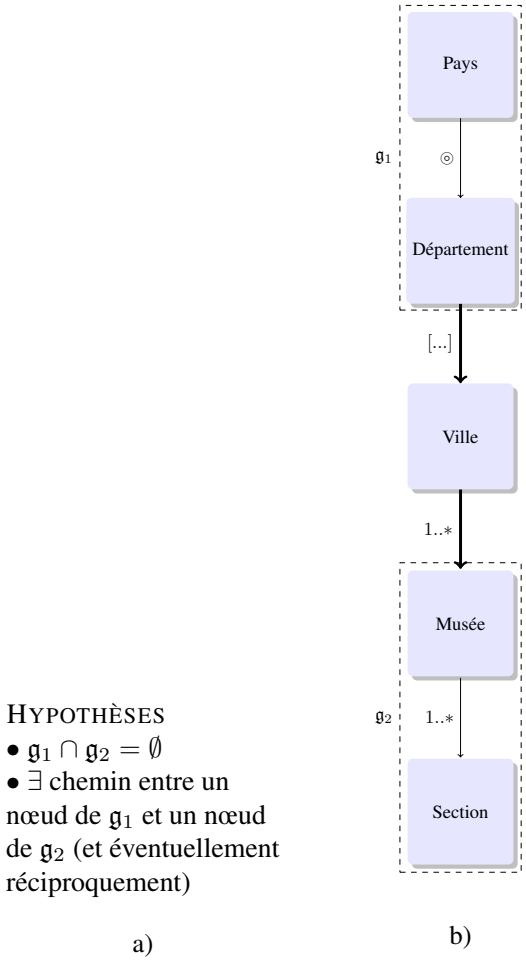

Figure 15. [C4]a : a)hypothèses b) exemple

5.2.1.1. Le(s) nœud(s) et les arcs

Sous les hypothèses de la figure 15a) le principe est de générer une abstraction du (des) chemin(s) minimaux (au sens de lg) entre les graphes $\mathfrak{g}_1$ et $\mathfrak{g}_2$ à l'aide d'un (ou plusieurs) arc(s). A la différence de [C1], nous n'avons pas d'ensemble d'ancêtre(s) A appartenant à $\mathcal{H}$ (qui contenait les origines des arcs à générer). Nous devons créer un ensemble $D_{g_{1-2}}$ (resp. $D_{g_{2-1}}$) d'arcs $(n_o, n_d)|n_o \in N_1$ et $n_d \in N_2$

(resp. $(n_o, n_d)|n_o \in N_2$ et $n_d \in N_1$).

Soient [] le constructeur de liste et $\{\}$ le constructeur d'ensemble.

Soit $Liaison$ le domaine modélisant un chemin minimal, $Liaison : [E]$. Un chemin minimal, l, est une liste d'arcs $l = [e_1 \ldots e_n]$.

Soient *nœudDepart* : $Liaison \rightarrow N$ et *nœudArrivee* : $Liaison \rightarrow N$ deux fonctions restituant les nœuds de départ (resp. d'arrivée) d'un chemin minimal.

Soient $C : \{Liaison\}$ le domaine modélisant l'ensemble des liaisons minimales entre deux graphes (*i.e.* $c_i \in C, lg(\mathcal{H}, \textit{nœudDepart}(c_i), \textit{nœudArrivee}(c_i)) > 0$ et minimale) et la fonction $chemins : \mathcal{H} \times G \times G \rightarrow C$ générant l'ensemble des chemins minimaux entre deux graphes.

Soit $c = chemins(\mathcal{H}, \mathfrak{g}_1, \mathfrak{g}_2)$ (resp. $chemins(\mathcal{H}, \mathfrak{g}_2, \mathfrak{g}_1)$).

Soit $nouveauxLiens : \{(n_o : N; n_d : N; \text{étiquette}\} = \{(\textit{nœudDepart}(c_i), \textit{nœudArrivee}(c_i), \epsilon)\}|c_i \in C, 1 \leq i \leq |C|$ l'ensemble qui contient tous les chemins auxquels ont été appliquées les règles de la table 1 pour n'obtenir qu'un ensemble d'arcs étiquetés. La figure 16 (a) illustre l'ensemble des chemins étiquetés minimaux entre $\mathfrak{g}_1$ et $\mathfrak{g}_2$. L'ensemble $nouveauxLiens$ est illustré figure 16 (b). Les arcs redondants et les arcs multiples ont été supprimés respectivement de par la notion ensembliste de $nouveauxLiens$ et via les règles de la table 2.

Soient Π opérateur de projection et σ l'opérateur de sélection.

L'ensemble $D_{g_{1-2}}$ (resp. $D_{g_{2-1}}$) est défini par :

$D_{g_{1-2}}$ (resp. $D_{g_{2-1}}$) $= \{\Pi_{N \times N}(nouveauxLiens)\}$. $D_{g_{1-2}}$ (resp. $D_{g_{2-1}}$) est l'ensemble d'arcs (*i.e.* des couples de nœuds) : $(n_o, n_d)|n_o \in N_1$ et $n_d \in N_2$ (resp. $n_o \in N_2$ et $n_d \in N_1$).

Dans l'exemple figure 15b), $D_{g_{1-2}} = \{(\text{Département}, \text{Musée})\}$ et $D_{g_{2-1}} = \emptyset$.

5.2.1.2. Graphe résultat : définition

Le graphe résultat (*i.e.* $\mathfrak{h}$) est défini de la manière suivante :

- $N_{\mathfrak{h}} = N_1 \cup N_2$
- $E_{\mathfrak{h}} = E_1 \cup E_2 \cup D_{g_{1-2}} \cup D_{g_{2-1}}$

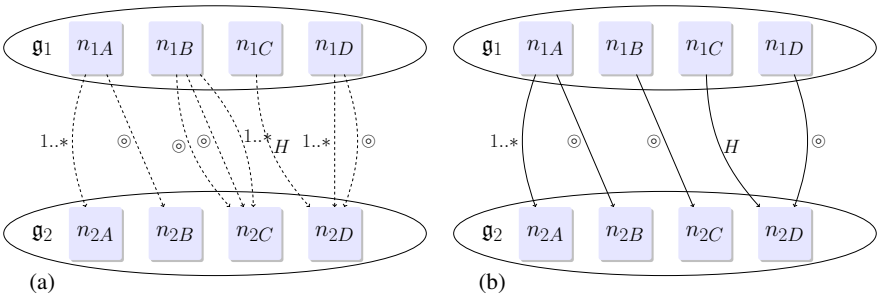

Figure 16. Application de la méthode de calcul de $D_{g_{1-2}}$

5.2.1.3. Graphe résultat : gestion des redondances potentielles

Par construction des ensembles $D_{g_{1-2}}$ et $D_{g_{2-1}}$ il ne peut pas y avoir de redondances d'arcs ni d'arcs multiples.

Les redondances qui apparaîtraient au niveau des attributs sont supprimées par application de l'algorithme 2 dont la philosophie est similaire à celle de l'algorithme 1. Les nœuds à traiter sont définis par $D' = \Pi_{n_o}(nouveauxLiens)$. Pour tous les nœuds a de D' l'algorithme 2 $Nettoyage(\mathfrak{h}, a, \sigma_{n_o=a}(nouveauxLiens))$ est appliqué.

si $(\forall(\text{ancêtre}, n_d, \epsilon) \in \text{arcsIssusAncêtre} \Rightarrow \epsilon = H)$ **alors**
  si $(\exists\ att | \forall(\text{ancêtre}, n_d, \epsilon) \in \text{arcsIssusAncêtre}, a_{tt} \in \nu(n_d)$ **alors**
    `/*on peut supprimer` a_{tt} `dans les nœuds fils, et le remonter dans l'ancêtre` *`ancêtre`*`*/`
    $\nu(n_d) = \nu(n_d) - a_{tt}$
    $\nu(\text{ancêtre}) = \nu(\text{ancêtre}) \cup a_{tt}$
  finsi
finsi

Algorithme 2 – Gestion des redondances : $Nettoyage(\mathfrak{h}, \text{ancêtre}, \text{arcsIssusAncêtre})$

5.2.2. [C4]b : en tenant compte des instances

La problématique générale est illustrée figure 17. Figure 17a), un nœud n_1 de $\mathfrak{g}_1$ est relié à un nœud n_2 de $\mathfrak{g}_1$ par deux chemins différents (e.g. c_1 et c_2). L'un de ces chemins, c_2, fait intervenir des nœuds/arcs de $\mathfrak{g}_2$ suite à une opération de collage. Si la sémantique conduit à relier les mêmes instances par c_1 et c_2 nous nous retrouvons dans [C4]a figure 17b). Par contre figure 17c), si la sémantique conduit à relier n_1 et n_2 à partir de deux instances de n_1, il faut être en mesure de gérer les informations alphanumériques liées à ces deux instances (*i.e.* i_1, i_2).

Dans l'exemple présenté figure 18, l'opération de collage est appliquée entre le graphe $\mathfrak{g}_1$ (**Ville**, **Œuvre**) et $\mathfrak{g}_2$ (**Musée**, **Section**)). A titre d'exemple, une ville (i_1) possède un musée et dans ce musée sont exposées des œuvres. Par ailleurs une œuvre (i_3) a été découverte dans une ville (i_2 par c_1), différente de celle où l'œuvre est exposée (i_1 par c_2). Ce collage conduit donc à deux chemins entre **Ville** et **Œuvre**. Dans [C4]b, nous étudions donc la propagation des attributs au niveau instance, car

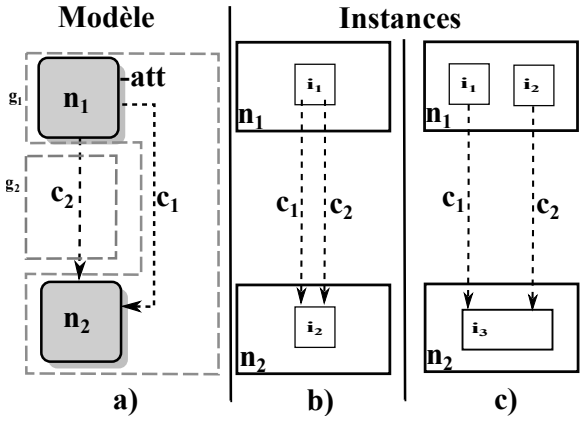

Figure 17. Problématique générale

deux instances différentes de **Ville** (*i.e.* n_1) sont reliées à une même instance de **Œuvre** (*i.e.* n_2).

5.2.2.1. Le(s) nœud(s)

Conceptuellement il existe alors deux types de nœud (illustration figure 19) : (1) les nœuds pour lesquels aucune duplication d'instances n'apparaît, le nœud initial de $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) garde son rôle. (2) les nœuds pour lesquels cette duplication entre en ligne de compte, il apparaît alors le besoin de différencier les informations venant de la première instance (prises en charge par le nœud initial) et les informations venant de la deuxième instance. Les attributs de n_1 devant être propagés voient leurs noms préfixés par le rôle du chemin c_2 apparu suite au collage.

5.2.2.2. Les arcs : définition

Il apparaît alors une spécialisation du nœud initial (soit n_{dif} ce nœud). Une liaison entre les deux nœuds est modélisée par une fonction d'étiquetage de valeur H permettant d'obtenir pour le nouveau nœud n_{dif}, par héritage, les informations de n_2 (figure 19). L'étiquetage des ensembles d'arcs $D_{g_{1-2}}$ et $D_{g_{2-1}}$ est réalisé de la même manière que dans [C4]a.

5.2.2.3. Graphe résultat : définition

- $N_{\mathfrak{h}} = N_1 \cup N_2 \cup \{n_{dif}\}$
- $E_{\mathfrak{h}} = E_1 \cup E_2 \cup D_{g_{1-2}} \cup D_{g_{2-1}} \cup \{(n_2, n_{dif})\}$
  $\epsilon(n_2, n_{dif}) = H$

5.2.2.4. Graphe résultat : gestion des redondances potentielles

La gestion de la redondance est similaire à [C4]a.

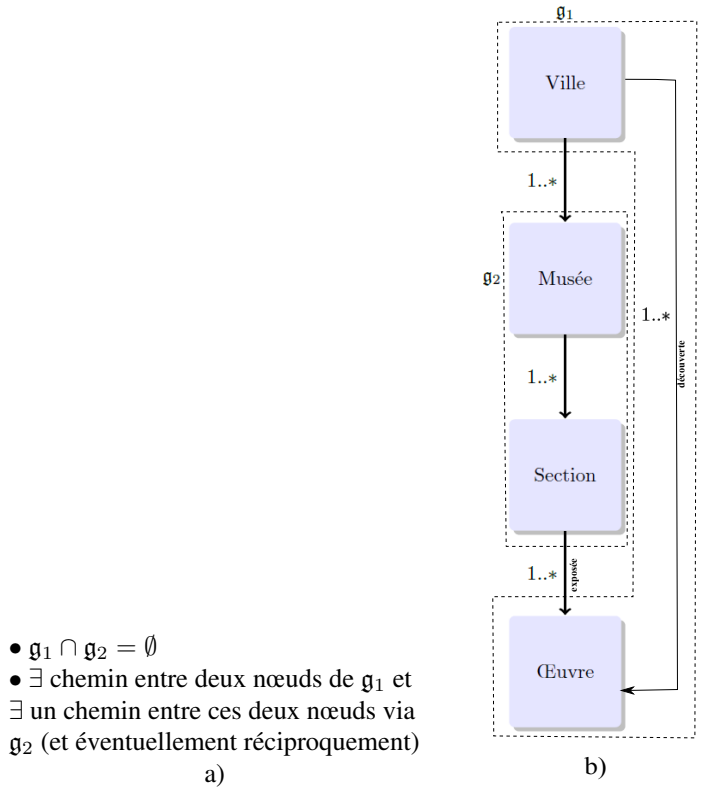

Figure 18. [C4]b : a) hypothèses b)sémantique induisant potentiellement deux instances

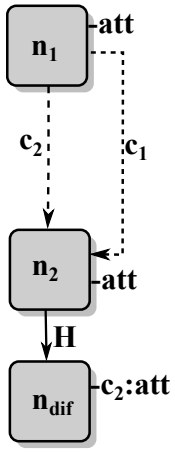

Figure 19. Mise en oeuvre

6. Conclusion

Devant l'explosion des services d'aide à la navigation sur différents supports (e.g. dans les véhicules, sur les téléphones portables), la mise à disposition d'informations sémantiques venant compléter les représentations cartographiques devient un enjeu important. Si aujourd'hui les approches commerciales se focalisent sur les présentations cartographiques de la navigation, la mise à disposition d'informations factuelles (alphanumériques conventionnelles) reste un problème ouvert. Dans cet article nous proposons donc d'aborder la gestion des informations alphanumériques, en particulier celles qui sont liées à la dimension spatiale. Nous proposons un opérateur de *Collage* définissant un ensemble de règles qui permettent de conserver une sémantique spatiale consistante dans ces données.

L'opérateur de *Collage* proposé permet d'offrir un complément d'information en intégrant au sein d'une même vision utilisateur des informations provenant de différents points de vue. Ces points de vue peuvent être définis comme appartenant au même système d'information ou comme provenant de sources multiples (interopérabilité). Notre pré-supposé est que dans le cas de sources multiples un alignement ontologique a déjà été effectué afin d'éviter tout problème d'interprétation lors des manipulations de modèles. Dans cet article nous partons d'un modèle de référence, défini par un graphe, et définissons les règles à mettre en oeuvre pour permettre le rapprochement de ces points de vue. Le tableau 4 présente une synthèse de nos propositions où sont précisés pour les différents cas (propriétés vérifiées par les graphes): les hypothèses d'application, le graphe résultat du collage et la gestion des redondances potentielles suite à cette opération. L'opérateur de *Collage* est envisagé sous deux hypothèses principales, graphes dépendants ou non, qui sont déclinées en sous-cas (en lien avec l'existence ou non de chemins entre/dans les graphes à rapprocher).

Il nous reste à intégrer les opérateurs de $ZoomIn$ (proposé dans un article antérieur [notreRef]) et de $Collage$ au sein d'une même application, et de terminer ce travail par la définition d'un opérateur de généralisation sur le plan sémantique, pendant du $ZoomIn$, le $ZoomOut$.

Tableau 4. Synthèse des cas d'application de l'opérateur de Collage

Indépendance des graphes	hypothèses	graphe résultat	gestion redondance
• $\mathfrak{g}_1 \cap \mathfrak{g}_2 = \emptyset$ $\nexists$ chemin entre un nœud de $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) et un nœud de $\mathfrak{g}_2$ (resp. $\mathfrak{g}_1$) dans $\mathcal{H}$	**[C1]** • sous-ensemble de $\mathcal{H}$ issu de $\mathfrak{g}_1$ et $\mathfrak{g}_2$ quasi-fortement connexe	$N_\mathfrak{h} = N_1 \cup N_2 \cup A$ $E_\mathfrak{h} = E_1 \cup E_2 \cup \{e_q = (a_i, n_j) \mid q \in \mathbb{N}, q > \lvert E_1\rvert + \lvert E_2\rvert, a_i \in A \; et \; n_j \in D_k \; avec \; k \in \{1,2\}$ et $lg(\mathcal{H}, a_i, n_j) \neq 0\}$ Etiquette cas d'une liaison indirecte : table 1	`arcs multiples` : table 2 `attributs` : algorithme 1 Nettoyage($\mathfrak{h}, a, D_1, D_2$)
	[C2] • $\nexists$ de sous-graphe quasi-fortement connexe issu de $\mathfrak{g}_1$ et $\mathfrak{g}_2$ dans $\mathcal{H}$	$N_\mathfrak{h} = N_1 \cup N_2 \cup A$ $E_\mathfrak{h} = E_1 \cup E_2 \cup \{e_q = (n_o, n_j) \mid q \in \mathbb{N} \wedge q > \lvert E_1\rvert + \lvert E_2\rvert, \; n_j \in D_{rk} \; avec \; k \in \{1,2\}\}$ pour le niveau macro (resp. $e_q = (n_j, n_o) \mid q \in \mathbb{N} \wedge q > \lvert E_1\rvert + \lvert E_2\rvert, \; n_j \in D_{fk} \; avec \; k \in \{1,2\}$ pour le niveau micro) et $\epsilon(e_q) = H$	niveau macro seulement `attributs` : algorithme 1 Nettoyage($\mathfrak{h}$, n_o, D_{r1}, D_{r_2})
• $\mathfrak{g}_1 \cap \mathfrak{g}_2 \neq \emptyset$	**[C3]**	$N_\mathfrak{h} = N_1 \cup N_2$; $E_\mathfrak{h} = E_1 \cup E_2$	pas de redondance possible par construction
• $\mathfrak{g}_1 \cap \mathfrak{g}_2 = \emptyset$	**cas4a** (modèle) • $\exists$ un chemin entre un nœud de $\mathfrak{g}_1$ et un nœud de $\mathfrak{g}_2$ (év. réciproquement)	$N_\mathfrak{h} = N_1 \cup N_2$ $E_\mathfrak{h} = E_1 \cup E_2 \cup D_{g_{1-2}} \cup D_{g_{2-1}}$	`attributs` : algorithme 2 $Nettoyage($ $\mathfrak{h}, a, \sigma_{n_o=a}(nouveauxLiens))$
	cas4b (instance) • $\exists$ un chemin entre deux nœuds de $\mathfrak{g}_1$ et $\exists$ un chemin entre ces deux nœuds via $\mathfrak{g}_2$ (év. réciproquement)	$N_\mathfrak{h} = N_1 \cup N_2 \cup \{n_{dif}\}$ $E_\mathfrak{h} = E_1 \cup E_2 \cup D_{g_{1-2}} \cup D_{g_{2-1}} \cup \{(n_2, n_{dif})\}$ $\epsilon(n_2, n_{dif}) = H$	

Bibliographie

Aiken A., Chen J., Stonebraker M., Woodruff A. (1996). Tioga-2: A direct manipulation database visualization environment. In *Proceedings of the twelfth international conference on data engineering*, p. 208-217.

Cruz I. F., Sunna W., Chaudhry A. (2004). Semi-automatic ontology alignment for geospatial data integration. In M. J. Egenhofer, C. Freksa, H. J. Miller (Eds.), *Geographic information science: Third international conference, giscience 2004, adelphi, md, usa, october 20-23, 2004. proceedings*, p. 51–66. Berlin, Heidelberg, Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-30231-5_4

Cruz I. F., Xiao H. (2008). Data integration for querying geospatial sources. In *Geospatial services and applications for the internet*, p. 110–134. http://dx.doi.org/10.1007/978-0-387-74674-6_6

El Adnani M., Yétongnon K., Benslimane D. (2001). A multiple layered functional data model to support multiple representations and interoperability of gis: Application to urban management systems. In *Proceedings of the 9th acm international symposium on advances in geographic information systems*, p. 70–75. New York, NY, USA, ACM. http://doi.acm.org/10.1145/512161.512177

Estublier J., Ionita A. D., Nguyen T. (2008). Code generation for a bi-dimensional composition mechanism. In *Software engineering techniques - third IFIP TC 2 central and east european conference, CEE-SET 2008, brno, czech republic, october 13-15, 2008, revised selected papers*, p. 171–185. http://dx.doi.org/10.1007/978-3-642-22386-0_13

Fonseca F. T., Egenhofer M. J., Davis C. A., Câmara G. (2002). Semantic granularity in ontology-driven geographic information systems. *Ann. Math. Artif. Intell.*, vol. 36, n° 1-2, p. 121–151. http://dx.doi.org/10.1023/A:1015808104769

Harel D., Rumpe B. (2004, octobre). Meaningful modeling: What's the semantics of "semantics"? *Computer*, vol. 37, n° 10, p. 64–72. http://dx.doi.org/10.1109/MC.2004.172

Herrmann C., Krahn H., Rumpe B., Schindler M., Völkel S. (2007). An Algebraic View on the Semantics of Model Composition. In *Proceedings of the third european conference on model driven architecture - foundations and applications*, p. 99–113. http://dx.doi.org/10.1007/978-3-540-72901-3_8

Hoel E. G., Bakalov P., Kim S., Brown T. (2015). Moving beyond transportation: utility network management. In J. Bao *et al*. (Eds.), *Sigspatial'15*. ACM. http://dblp.uni-trier.de/db/conf/gis/gis2015.html#HoelBKB15

Lieberman J., Singh R., Goad C. (2007). *W3c geospatial ontologies*. Report. W3C. (http://www.w3.org/2005/Incubator/geo/XGR-geo-ont/)

Longo J. S. C., Medeiros C. B. (2013). Providing multi-scale consistency for multi-scale geospatial data. In *Conference on scientific and statistical database management, SSDBM '13, baltimore, md, usa, july 29 - 31, 2013*, p. 1-12. http://doi.acm.org/10.1145/2484838.2484867

Mustière S., Devogele T. (2008). Matching networks with different levels of detail. *GeoInformatica*, vol. 12, n° 4, p. 435–453. http://dx.doi.org/10.1007/s10707-007-0040-1

OGC-WM. (2006). (http://www.opengeospatial.org/standards/wms)

Partyka J., Alipanah N., Khan L., Thuraisingham B., Shekhar S. (2008). Content-based ontology matching for gis datasets. In *Proceedings of the 16th acm sigspatial international conference on advances in geographic information systems*, p. 1–4. New York, NY, USA, ACM. http://doi.acm.org/10.1145/1463434.1463496

Pavel S., Euzenat J. (2013, janvier). Ontology matching: State of the art and future challenges. *IEEE Trans. on Knowl. and Data Eng.*, vol. 25, n° 1, p. 158–176. http://dx.doi.org/10.1109/TKDE.2011.253

Raimond A. O., Mustière S. (2008). Data matching - a matter of belief. In *Headway in spatial data handling, 13th international symposium on spatial data handling, montpellier, france, 23-25 july 2008*, p. 501–519. http://dx.doi.org/10.1007/978-3-540-68566-1_29

Stell J., Worboys M. (1998). Stratified map spaces: A formal basis for multi-resolution spatial databases. In *Sdh'98 proceedings 8th international symposium on spatial data handling, international geographical*, p. 180-189.

Zhao T., Zhang C., Wei M., Peng Z.-R. (2008). Ontology-based geospatial data query and integration. In T. J. Cova, H. J. Miller, K. Beard, A. U. Frank, M. F. Goodchild (Eds.), *Geographic information science: 5th international conference, giscience 2008, park city, ut, usa, september 23-26, 2008. proceedings*, p. 370–392. Berlin, Heidelberg, Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-87473-7_24