
Modélisation sémantique des relations de dépendance entre les processus de l'analyse en cycle de vie

Benjamin Bertin^{*,}, Vasile-Marian Scuturici^{*,**}, Jean-Marie Pinon^{*,**}, Emmanuel Risler^{*,**}**

^{*} Université de Lyon, CNRS

^{**} INSA-Lyon, LIRIS, UMR5205, F-69621, Villeurbanne Cedex

{benjamin.bertin,marian.scuturici,jean-marie.pinon}@liris.cnrs.fr

^{***} INSA-Lyon, ICJ, UMR5208, My C-Sense, F-69621, Villeurbanne Cedex

emmanuel.risler@insa-lyon.fr

ABSTRACT. Environmental impact assessment of goods and services is nowadays a major challenge for both economic and ethical reasons. Life Cycle Assessment (LCA) provides a well-accepted methodology for modelling environmental impacts of human activities. One stage of the LCA methodology is Life Cycle Inventory (LCI), which consists of decomposing economic activities as elementary processes linked together through interdependency relations. A global analysis of economic activities requires considering a huge amount of elementary processes and interdependency links, making the model difficult to understand. In this paper, we propose a semantic approach for the modelling of LCI databases. The method has the advantage of offering a more comprehensible model. We explain our model and illustrate it with LCI data for the U.S. electricity production.

RÉSUMÉ. L'analyse des impacts environnementaux de la production de biens et de services est aujourd'hui devenue un enjeu majeur. L'Analyse en Cycle de Vie est la méthode consacrée pour modéliser les impacts environnementaux des activités humaines. L'une des étapes de cette méthode consiste à décomposer les activités économiques en processus élémentaires interdépendants. L'audit et la compréhension de ces inventaires nécessite de s'intéresser à un très grand nombre de processus élémentaires et de leur relations d'interdépendances. Dans cet article, nous proposons une approche sémantique de la modélisation des bases de données d'inventaires en cycle de vie. Notre proposition permet d'offrir une vue globale des inventaires et donc d'en simplifier la compréhension et l'audit. Nous illustrons cette approche sur l'inventaire en cycle de vie de la production d'électricité aux États-Unis.

KEYWORDS: Life Cycle Assessment, Ontology, Environment

MOTS-CLÉS : Analyse en cycle de vie, Ontologie, Environnement

1. Introduction

La réduction des impacts environnementaux des activités humaines, tels que les émissions de gaz à effets de serre, nécessite de modéliser et évaluer les effets sur l'environnement de ces activités. C'est l'objectif de la méthode dite de l'Analyse en Cycle de Vie (ACV)(ISO, 2006). Elle permet de déterminer les impacts liés à la production d'un produit, liés à un service ou, plus généralement, liés à toute activité humaine. Cette méthode peut prendre en compte toutes les étapes de la vie d'un produit comme la production ou le recyclage. L'ACV permet d'évaluer différents impacts tels que les émissions de gaz à effet de serre ou les rejets de produits chimiques.

Une étude basée sur la méthode ACV se décompose en quatre étapes(ISO, 2006). La première consiste à définir les objectifs et le champ de l'étude. Le système étudié est ensuite décomposé en processus élémentaires pour réaliser un *inventaire en cycle de vie*. Ces processus correspondent à des étapes spécifiques d'un cycle de vie ou à des activités humaines (comme la production d'énergie, l'épandage d'engrais, un déplacement en avion, etc.). Ils peuvent être composés les uns en fonction des autres, par exemple : la production d'une voiture dépend (la plupart du temps) de la production d'acier. L'analyse et l'interprétation des données de l'inventaire en cycle de vie sont réalisées dans la troisième et la quatrième étape de la méthode ACV. La méthode consacrée pour modéliser les interactions entre les processus d'un inventaire et entre les processus et l'environnement est basée sur l'utilisation d'une matrice d'Entrée/Sortie(Leontief, 1986). La deuxième section de cet article présente succinctement cette méthode.

Plusieurs organismes proposent des bases de données d'inventaires en cycle de vie(Frischknecht, 2005b)(GaBi, 2011)(NREL, 2011) qui servent de base à l'étude en cycle de vie d'un système. Mais ces bases de données peuvent contenir plusieurs milliers de processus, ce qui les rend difficile à comprendre à moins de réaliser une étude détaillée de toute la base. Il existe pourtant des similarités sémantiques entre les processus et leurs relations, telles que pour les processus de production d'électricité à partir du charbon (à partir de différents types de charbon : lignite, bitumineux, etc.) qui sont tous dépendants des processus de transports. Ces proximités sémantiques sont évidemment difficile à appréhender quand elles sont dispersées dans une matrice d'Entrée/Sortie.

Nous proposons donc d'ajouter des informations sémantiques à la modélisation des inventaires en cycle de vie et de regrouper sémantiquement les processus. Ces regroupements sont ensuite utilisés pour créer des relations de dépendances entre plusieurs processus afin d'offrir une vue d'ensemble des données. Notre approche est basée sur l'existence de deux graphes orientés pondérés. Le premier, que l'on nomme *graphe détaillé*, est une simple transposition de la matrice d'Entrée/Sortie. Le deuxième, que l'on nomme *macro-graphe*, contient des relations de dépendance entre des groupes de processus. La modélisation se fait alors en exprimant des relations de dépendances entre des groupes de processus. Ces relations doivent ensuite être traduites en relations de dépendances dans le graphe détaillé afin de calculer les

impacts des processus. La figure 1 illustre les deux niveaux de graphes. La troisième section de cet article présente notre modèle en détail. Puis, nous expliquons dans la quatrième section la méthode permettant de convertir des relations de dépendance entre des groupes de processus du macro-graphe en relations de dépendance entre processus élémentaires dans le graphe détaillé. Dans la dernière section, nous présentons une application de notre méthode sur des données extraites à partir de la base de données d'inventaire en cycle de vie du National Renewable Energy Laboratory (NREL, 2011). L'extrait que nous avons choisi concerne la production d'électricité aux États-Unis.

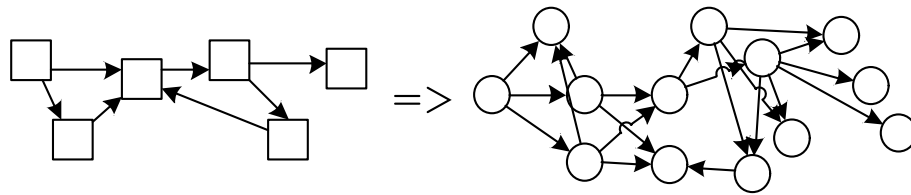


Figure 1. Les deux graphes orientés de notre approche. Le macro-graphe représenté dans la partie gauche, contient des relations entre groupes de processus. Ces relations sont traduites en un ensemble de relations entre processus élémentaires dans le graphe détaillé tel qu'illustré dans la partie droite du schéma.

2. Modèle de l'inventaire en cycle de vie

L'objectif de l'analyse en cycle de vie est d'évaluer les impacts environnementaux d'une activité humaine, telle que la production d'un produit ou l'activité d'une entreprise. Cette évaluation est déterminée en réalisant un inventaire des flux élémentaires entre le système étudié et l'environnement pour chaque étape de la production d'un produit ou pour chaque activité d'une entreprise (Guinée, 2002) (Heijungs, 2002). Dans la terminologie de l'analyse en cycle de vie, ces étapes sont appelées *processus*.

Ces processus sont interdépendants, comme la production d'électricité dans une centrale nucléaire qui nécessite d'extraire de l'uranium. La modélisation habituelle est basée sur une matrice d'Entrée/Sortie A (Leontief, 1986). Pour n processus, cette matrice est de taille $n \times n$, où a_{ij} correspond au degré de dépendance, nommée *coefficient de dépendance*, entre un processus i et un processus j . Cette matrice décrit les relations de dépendances entre les processus au sein de la technosphère¹, ces relations sont aussi appelées *flux élémentaires*. Chaque processus est associé à des impacts sur l'environnement, comme la quantité de gaz à effets de serre rejetés ou la pollution des eaux générée. Les m différents impacts produits par n processus sont modélisés au sein d'une matrice B de taille $n \times m$ où b_{ij} correspond au j^{me} impact du i^{me} processus. Cette matrice décrit des flux entre la technosphère et l'écosphère².

1. Le dictionnaire des termes environnementaux McGraw-Hill définit ce terme comme la part de l'environnement physique affecté par des constructions ou des modifications humaines.

2. L'agence américaine de l'environnement définit ce terme comme la "bulle biologique" qui convient la vie sur terre, dans les eaux et dans les airs.

La détermination des impacts d'un processus particulier nécessite de calculer récursivement les impacts de ses prédécesseurs. Comme expliqué dans (Peters, 2007), la matrice d'Entrée/sortie peut être considérée comme un simple système d'équation. Il est donc possible de calculer les impacts des processus en utilisant une méthode de résolution itérative³ ou directe telle que l'élimination de Gauss(Nicholson, 1990). Il est important de noter qu'il n'est possible de calculer les impacts des processus que si le système d'équation correspondant à la matrice des flux n'accepte qu'une seule solution. Par exemple, si pour produire un litre d'essence il était nécessaire de consommer plus d'un litre d'essence, un algorithme itératif ne pourrait converger pour le système d'équation induit par cette modélisation. La condition nécessaire pour qu'une méthode de résolution itérative appliquée sur un système d'équation converge est que le rayon spectral de la matrice des coefficients de ce système soit inférieur à un(Varga, 2010). Cette condition nous garantit qu'il existera une solution unique pour le système d'équation. L'existence, éventuelle, de cycles nous empêche de calculer directement les impacts des processus dans le macro-graphe, ce qui nous contraint à le convertir en un graphe détaillé et à calculer les impacts avec une méthode classique.

3. Modèle à deux niveaux de graphes

Notre approche est basée sur l'existence de deux niveaux de graphes orientés pondérés. Le premier niveau contient des relations entre des processus (le *graphe détaillé*) et le deuxième niveau contient des relations entre des groupes de processus (le *macro-graphe*). Le second niveau offre une vue simplifiée des données contenues dans le premier niveau et permet d'exprimer plus aisément des relations entre plusieurs processus sémantiquement proche. Chaque processus est décrit par un ensemble de mots-clefs qui sont stockés dans une ontologie(Gruber, 1993)(McGuinness, 2002). Le vocabulaire de cette ontologie est composé des mots-clefs et de prédicats servant à créer des relations binaires entre les mots-clefs. Cette ontologie nous permet de regrouper sémantiquement les processus et les coefficients de dépendances. Un groupe est assimilable à une matrice multi-dimensionnelle dans laquelle chaque dimension est un ensemble de mots-clefs. Ces ensembles de mots-clefs sont récupérés à l'aide de requêtes sur l'ontologie. Nous pouvons ensuite créer des relations de dépendance entre des groupes de processus en les pondérant avec un groupe de coefficients, de la même façon que nous créons des relations de dépendances entre des processus en les pondérant par des coefficients.

3. Les données d'un inventaire en cycle de vie sont entachée d'incertitude. Il est donc possible d'utiliser une méthode de résolution itérative et d'arrêter l'algorithme si la variation des impacts calculés est inférieure à l'incertitude. Ce qui peut permettre un gain de temps de traitement par rapport à une méthode de résolution directe. Mais une méthode de résolution itérative ne fournira qu'un résultat approximatif.

3.1. Le graphe détaillé

Un processus est associé à des impacts environnementaux et à d'autres processus. L'ensemble des impacts environnementaux d'un processus est la somme de ses impacts propre (*i.e.* : qui ne sont pas issus de ses prédécesseurs) et des impacts de ses prédécesseurs multipliés par des coefficients de dépendances. Les impacts environnementaux d'un processus peuvent donc être exprimés comme une composition linéaire des impacts d'autres processus. Soit P un processus et $I(P)$ l'ensemble de ses impacts. On note P_0, \dots, P_n les prédécesseurs de P (aussi appelé *processus amonts* de P), C_0, \dots, C_n les coefficients de dépendances entre les processus amont et P . On note $I(P_0), \dots, I(P_n)$ l'ensemble des impacts des processus amont et $I_{\text{propre}}(P)$ les impacts propres de P . On a alors :

$$I(P) = I_{\text{propre}}(P) + \sum_{i=0}^n (I(P_i) * C_i)$$

Soit $G(V, E)$ le graphe orienté pondéré représentant les relations d'interdépendance entre les processus, où l'ensemble des nœuds V correspond à l'ensemble des processus, l'ensemble des arcs E aux relations de dépendances entre les processus et l'ensemble des pondérations correspond à l'ensemble des coefficients. Soit P et P_0, \dots, P_n des nœuds de G , un arc entre P_i et P signifie que le processus P dépend de P_i (voir la Figure 2).

Par exemple, la production d'électricité à partir de pétrole nécessite de le transporter de la raffinerie jusqu'à la centrale. Nous pouvons modéliser ces relations de dépendances dans un graphe à l'aide d'arcs (pondérés par des coefficients de dépendance) entre les nœuds correspondant au transport de marchandise selon le véhicule et le nœud correspondant au processus de production d'électricité à partir de pétrole (voir la Figure 3)⁴.

3.2. L'ontologie des mots-clefs

Dans notre approche, les processus et les coefficients sont indexés et identifiés par des mots-clefs. Nous introduisons une contrainte sur l'indexation : il n'y a qu'un seul processus ou coefficient associé à un ensemble précis de mots-clefs. Le processus de transport par camion pourrait être identifié par les mots-clefs : *Transport* et *Truck*. Nous pouvons donc définir la notion de processus indexé :

4. Les processus et les coefficients sont aussi associés à une unité. La valeur des coefficients de dépendances dépend donc de cette unité. Si les impacts du processus de production d'électricité sont exprimés par KWh produit, et les impacts des processus de transport par km parcouru. Les coefficients de dépendances entre les processus de transport et le processus de production d'électricité seront donc exprimés en KWh par km parcouru. Ils correspondront à la distance moyenne entre une raffinerie et une centrale et seront dépendant de la quantité de pétrole nécessaire pour produire un KWh d'électricité.

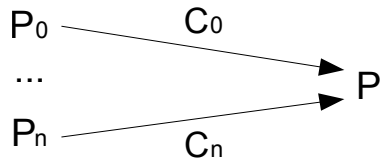


Figure 2. Graphe détaillé pour un processus P dépendant des processus P_0, \dots, P_n avec les coefficients de dépendances C_0, \dots, C_n .

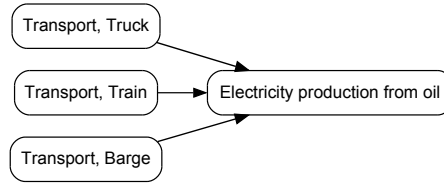


Figure 3. Graphe détaillé de la production d'électricité à partir de pétrole restreint aux processus amonts de transport de marchandise.

Définition 1. Soit k_1, \dots, k_n des mots-clefs. Un processus est un couple composé d'impacts et d'un ensemble de mots-clefs que l'on note : $p = (I(p), K_p)$ où $K_p = \{k_1, \dots, k_n\}$.

De façon analogue, nous pouvons définir un coefficient indexé :

Définition 2. Soit k_1, \dots, k_n des mots-clefs. Un coefficient est un couple composé d'une valeur scalaire et d'un ensemble de mots-clefs que l'on note : $c = (V(c), K_c)$ où $K_c = \{k_1, \dots, k_n\}$.

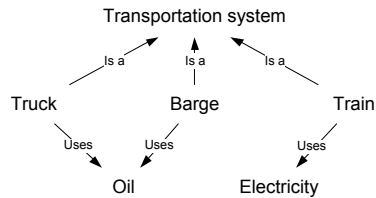
Les groupes de processus ou de coefficients sont des matrices multidimensionnelle dont les dimensions sont des ensembles distincts de mots-clefs. Il n'y a donc qu'un seul processus ou coefficient correspondant à une coordonnée d'un groupe.

L'ontologie des mots-clefs nous permet de définir dynamiquement les groupes. Un ensemble de mots-clefs (que nous nommons *dimension*) est le résultats d'une requête sur l'ontologie. A partir de l'ontologie présentée dans la Figure 4, nous pouvons créer une dimension contenant tous les modes de transport à l'aide d'une requête récupérant tous les mots-clefs reliés au mot-clef *Transportation system* en ne considérant que le prédicat *is a*. Nous pouvons aussi ne récupérer que les modes de transport qui consomment du pétrole en récupérant l'intersection entre l'ensemble des mots-clefs reliés au mot-clef *Transportation system* à l'aide du prédicat *is a* et l'ensemble des mots-clefs reliés au mot-clef *Oil* à l'aide du prédicat *uses*.

Le vocabulaire de notre ontologie s'exprime aisément en RDF(Klyne, 2004) avec les triplet suivants exprimés dans la syntaxe turtle(Beckett, 2011), en supposant que l'on dispose d'un espace de nommage xml *ex* pour notre ontologie :

```
ex:Keyword rdf:type rdfs:class;
ex:Predicate [
  rdf:type rdf:Property;
  rdfs:range ex:Keyword;
  rdfs:domain ex:Keyword
] .
```

L'ontologie présentée dans la Figure 4 correspond alors aux triplets RDF suivants :



```

ex:transportationSystem
  rdf:type ex:Keyword;
ex:is_a rdf:type ex:Predicate;
ex:truck ex:is_a
  ex:transportationSystem;
ex:barge ex:is_a
  ex:transportationSystem;
ex:train ex:is_a
  ex:transportationSystem;
ex:train ex:uses ex:oil;
ex:barge ex:uses ex:oil;
ex:train ex:uses
  ex:electricity;
  
```

Figure 4. Exemple d'ontologie pour stocker les mots-clefs décrivant des modes de transport. Les nœuds sont des mots-clefs et les étiquettes des arcs sont des prédicats.

Grâce à cette représentation, les dimensions peuvent être exprimées via des requêtes SPARQL (Prud'hommeaux, 2008) sur notre ontologie. Par exemple, la requête permettant de récupérer tous les mots-clefs décrivant des modes de transport est :

```

SELECT ?keyword
WHERE { ?keyword ex:is_a ex:transportationSystem. }
  
```

Tout changement apporté à cette ontologie des mots-clefs peut déclencher une mise à jour des dimensions précédemment définies, modifiant ainsi les groupes déjà existants. Par exemple, si un mode de transport est ajouté à l'ontologie de la Figure 4 (tel que *air-plane*), chaque groupe qui possède une dimension contenant les modes de transport sera mis à jour.

3.3. Le macro-graphe

Nous pouvons créer des relations entre des groupes de processus en utilisant des groupes de coefficients (ces relations sont nommées *macro-relations*). Les groupes et leurs dépendances sont représentés dans un graphe orienté pondéré $G_M(V, E)$ où l'ensemble des nœuds V correspond à l'ensemble des groupes de processus, l'ensemble des arcs E correspond aux relations de dépendance entre les groupes de processus et l'ensemble des pondérations des arcs correspond à l'ensemble des groupes de coefficients. La correspondance entre notre approche et la modélisation basée sur une matrice d'Entrée/Sortie est assez évidente : la matrice des flux est décomposée en sous-matrices qui sont utilisées comme pondération pour le macro-graphe. Les groupes de coefficients correspondent donc à des sous-ensembles de la matrice des flux.

Le macro-graphe a deux avantages : faciliter la compréhension de l'inventaire en cycle de vie et consolider la modélisation. Par exemple, si nous devons modéliser les

relations de dépendances entre tous les mode de transport et tous les modes de production d'électricité, le grand nombre de relations entre processus à créer pourrait induire des erreurs (ou des oublis). Alors que la modélisation des relations de dépendances entre des groupes, qui sont traduites en relations entre processus automatiquement, réduit le nombre d'informations sources dans le modèle et, donc, réduit le nombre potentiel d'erreurs de modélisation.

Comme la modélisation peut contenir des cycles entre les groupes de processus, il n'est pas possible de calculer directement les impacts depuis le macro-graphe. Il est nécessaire de traduire les relations entre des groupes de processus en relations entre des processus. En d'autres termes, nous devons convertir les arcs du macro-graphe en arcs du graphe détaillé. Puis, nous pouvons extraire la matrice des coefficients⁵ afin de calculer les impacts de la même manière que dans la méthodologie ACV. Nous expliquons en détail la procédure de conversion d'une macro-relation dans la quatrième section mais nous devons d'abord exposer quelques notions formelles de notre approche.

3.4. Formalisme des groupes

Soit S une ontologie. On note $\mathcal{P}(S)$ l'ensemble des sous-ensembles de S (privé de l'ensemble vide). On appelle **dimensions** les éléments de $\mathcal{P}(S)$ et on note $\mathcal{P}(\mathcal{P}(S))$ l'ensemble des sous-ensembles de $\mathcal{P}(S)$. Un groupe n'est valide que s'il a des dimensions distinctes, *i.e.* : si ses dimensions n'ont pas de mots-clefs en commun. Nous définissons la notion de consistance d'un ensemble de dimensions tel que :

Définition 3. Un élément $\mathcal{D} \in \mathcal{P}(\mathcal{P}(S))$, $\mathcal{D} = (D_0, \dots, D_n)$, est dit consistant si $\forall (i, j) \in \mathbb{N}^2, i \neq j, D_i \cap D_j = \emptyset$.

Pour qu'il n'y ait pas d'ambiguïté sur le choix des processus à associer quand une macro-relation est convertie en un ensemble de relations dans le graphe détaillé, les dimensions des groupes impliqués dans une relations ne doivent être en correspondance qu'avec, au plus, une seule dimension des autres groupes. Une dimension est en correspondance avec une autre dimension si leur intersection est non nulle. Nous définissons la notion de compatibilité d'ensemble de dimensions tel que :

Définition 4. Deux éléments \mathcal{D} et \mathcal{D}' de $\mathcal{P}(\mathcal{P}(S))$ sont compatible si les propriétés suivants sont vérifiées :

$$\forall D \in \mathcal{D}, \text{Card}\{D' \in \mathcal{D}' \mid D \cap D' \neq \emptyset\} \leq 1$$

$$\forall D' \in \mathcal{D}', \text{Card}\{D \in \mathcal{D} \mid D' \cap D \neq \emptyset\} \leq 1$$

5. La matrice des flux est construite en extrayant les pondération des arcs du graphe détaillé : un arc entre deux processus p_i et p_j pondéré par un coefficient c correspondra a un coefficient de la matrice des flux a_{ij} tel que $a_{ij} = c$.

Exemple : soit $\mathcal{D} = \{D_1, D_2\}$, $\mathcal{D}' = \{D_3\}$ et $\mathcal{D}'' = \{D_4\}$ trois ensembles de dimensions tel que $D_1 = \{A, B\}$, $D_2 = \{C, E\}$, $D_3 = \{A, F\}$ et $D_4 = \{B, C\}$. Nous pouvons dire que \mathcal{D} et \mathcal{D}' sont compatibles, alors que \mathcal{D} et \mathcal{D}'' sont incompatibles car $D_1 \cap D_4 = \{B\}$ et $D_2 \cap D_4 = \{C\}$.

Avec la notion de dimension et les restrictions que nous venons d'énoncer, nous pouvons définir ce qu'est un groupe de processus tel que :

Définition 5. Soit P l'ensemble des processus. On appelle groupe de processus le couple (\mathcal{D}, p) où $\mathcal{D} = (D_0, \dots, D_n)$ est un ensemble consistant de dimensions et p est une application $p : D_0 \times \dots \times D_n \rightarrow P$.

Une autre notation pour exprimer un groupe de processus est basée sur l'énumération de ses processus : $G_p = (\mathcal{D}, P_p) = (\{D_1, \dots, D_n\}, \{p_1, \dots, p_n\}) = (\{D_1, \dots, D_n\}, \{(I(p_1), K_{p_1}), \dots, (I(p_n), K_{p_n})\})$.

Similairement, nous pouvons définir la notion de groupe de coefficients tel que :

Définition 6. Soit C l'ensemble des coefficients. On appelle groupe de coefficients le couple (\mathcal{D}, c) où $\mathcal{D} = (D_0, \dots, D_n)$ est un ensemble consistant de dimensions et c est une application $c : D_0 \times \dots \times D_n \rightarrow C$.

Nous pouvons aussi exprimer un groupe de coefficients comme une énumération : $G_c = (\mathcal{D}, C) = (\{D_1, \dots, D_n\}, \{c_1, \dots, c_n\}) = (\{D_1, \dots, D_n\}, \{(V(c_1), K_{c_1}), \dots, (V(c_n), K_{c_n})\})$.

Par extension de la définition de la compatibilité de deux ensembles de dimensions, deux groupes de processus $G_{P_1} = (\mathcal{D}_1, p_1)$ et $G_{P_2} = (\mathcal{D}_2, p_2)$ sont compatibles si et seulement si \mathcal{D}_1 et \mathcal{D}_2 sont compatibles. Une macro-relation entre un groupe de processus amont G_{P_1} , un groupe de coefficients G_c et un groupe de processus aval G_{P_2} n'est valide que si les trois groupes sont compatibles entre eux (G_{P_1} doit être compatible avec G_c et G_{P_2} , G_c doit être compatible avec G_{P_2}).

Exemple : soit G_{P_1} un groupe de processus contenant les mode de transport tel que $G_{P_1} = (\{\{Truck, Barge\}\}, p_1)$. Soit G_C un groupe de coefficient contenant les coefficients de dépendances entre des modes de transport et des processus de production d'électricité tel que $G_C = (\{\{Truck, Barge\}, \{Electricity\}, \{Coal, Oil\}\}, c)$. Nous pouvons dire que G_{P_1} et G_C sont compatibles. La Figure 5 présente une représentation simplifiée de ces deux groupes. La notation p_{Truck} utilisée dans cette représentation signifie que ce processus, appartenant à G_{P_1} , est indexé par le mot-clef *Truck*. De la même façon, le coefficient $c_{Truck, Electricity, Coal}$, appartenant à G_C , est indexé par les mots-clefs *Truck*, *Electricity* et *Coal*.

4. Conversion des relations du macro-graphe en relations du graphe détaillé

Afin de calculer les impacts des différents processus intervenant dans les groupes de processus, il est nécessaire de convertir chaque macro-relation en un ensemble de

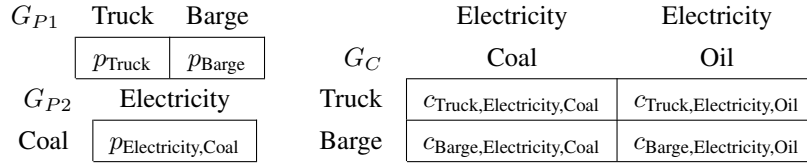


Figure 5. Représentation graphique d'un groupe de coefficients G_C et de deux groupes de processus G_{P1} et G_{P2} . Le groupe G_{P1} contient les processus correspondant à différents modes de transport et est composé d'une dimension. Le groupe G_{P2} contient les processus de production d'électricité à partir du charbon et est composé de deux dimensions. Le groupe de coefficients G_C contient les coefficients de dépendances entre les processus de G_{P1} et un groupe contenant des processus de production d'électricité (tel que G_{P2}) et est composé de trois dimensions.

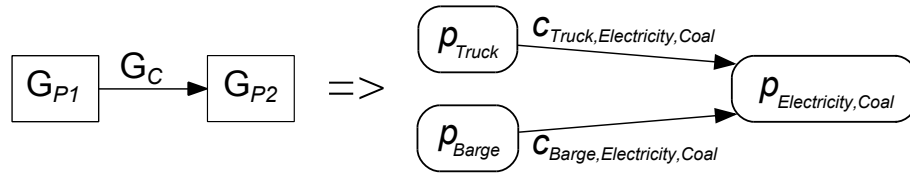


Figure 6. Le macro-graphe à gauche contient une relation entre le groupe G_{P1} (contenant les processus de transport de marchandise) et G_{P2} (contenant le processus de production d'électricité à partir de charbon) et est converti dans le graphe détaillé à droite (les arcs sont pondérés par un groupe de coefficients ou par un coefficient).

relations détaillées. Une macro-relation correspond à un arc dans le macro-graphe, pondéré par un groupe de coefficients. Plus formellement, on note une telle relation sous la forme : $((G_{P1}, G_{P2}), G_c)$ où G_{P1} et G_{P2} sont des groupes de processus et G_c est un groupe de coefficients. Cette relation doit être traduite en un ensemble de relations détaillées (*i.e.* : à un ensemble d'arcs dans le graphe détaillé) que l'on note : $\{(p_i, p_j), c_{ij}\}$ où p_i et p_j sont des processus et c_{ij} est un coefficient. Seuls les processus et les coefficients des groupes de processus et du groupe de coefficient qui ont une indexation commune seront reliés entre eux dans le graphe détaillé. La Figure 6 présente le macro-graphe correspondant à la macro-relation $((G_{P1}, G_{P2}), G_c)$ impliquant les groupes présentés dans la Figure 5 et sa conversion dans un graphe détaillé.

Pour réaliser cette traduction nous avons besoin d'introduire deux notions : l'union de deux ensembles de dimensions et le nombre d'appariement entre deux ensembles de dimensions.

L'union de deux ensembles de dimensions correspond à la réunion de trois ensembles : 1) l'intersection des dimensions des deux ensembles qui ont une intersection non nulle (qui ont donc des mots-clés en commun) ; 2) toutes les dimensions du premier

ensemble de dimensions qui n'ont pas d'intersection non nulle avec les dimensions du deuxième ensemble ; 3) toutes les dimensions du deuxième ensemble de dimensions qui n'ont pas d'intersection non nulle avec les dimensions du premier ensemble⁶.

Définition 7. On note \cup_D l'opérateur d'union de deux ensembles de dimensions tel que, pour deux ensembles de dimensions \mathcal{D}_1 et \mathcal{D}_2 :

$$\begin{aligned}\mathcal{D}_1 \cup_D \mathcal{D}_2 = & \{D_1 \cap D_2 \mid D_1 \in \mathcal{D}_1 \wedge D_2 \in \mathcal{D}_2 \wedge D_1 \cap D_2 \neq \emptyset\} \\ & \cup \{D_1 \mid D_1 \in \mathcal{D}_1 \wedge \forall D_2 \in \mathcal{D}_2, D_1 \cap D_2 = \emptyset\} \\ & \cup \{D_2 \mid D_2 \in \mathcal{D}_2 \wedge \forall D_1 \in \mathcal{D}_1, D_2 \cap D_1 = \emptyset\}\end{aligned}$$

Le nombre d'appariement entre deux ensembles de dimensions correspond au nombre de couples de dimensions de chaque ensemble qui ont une intersection non nulle.

Définition 8. On note $\alpha(\mathcal{D}_1, \mathcal{D}_2)$ le nombre d'appariement entre deux ensembles de dimensions \mathcal{D}_1 et \mathcal{D}_2 tel que :

$$\alpha(\mathcal{D}_1, \mathcal{D}_2) = \text{card}(\{(D_1, D_2) \mid D_1 \in \mathcal{D}_1 \wedge D_2 \in \mathcal{D}_2 \wedge D_1 \cap D_2 \neq \emptyset\})$$

A partir de ces deux notions, nous pouvons définir la règle de conversion d'une macro-relation noté $((G_{P_1}, G_{P_2}), G_c)$. Soit G_{P_1} et G_{P_2} deux groupes de processus, soit G_c un groupe de coefficients et soit $((G_{P_1}, G_{P_2}), G_c)$ un arc du macro-graphe. La règle de conversion de cet arc dans le graphe détaillé est :

$$\begin{aligned}((G_{P_1}, G_{P_2}), G_c) \rightarrow & \{(p_1, p_2), c \mid p_1 \in P_1 \wedge p_2 \in P_2 \wedge c \in C \\ & \wedge \text{card}(K_{p_1} \cap K_c) = \alpha(\mathcal{D}_1, \mathcal{D}_c) \\ & \wedge \text{card}((K_{p_1} \cup K_c) \cap K_{p_2}) = \alpha(\mathcal{D}_1 \cup_D \mathcal{D}_c, \mathcal{D}_2)\end{aligned}$$

La conversion d'un arc du macro-graphe nécessite donc d'essayer d'associer chaque processus du groupe amont G_{P_1} avec un coefficient du groupe de coefficient G_c et un processus du groupe aval G_{P_2} . Étant donné que les trois groupes doivent être compatibles entre eux, si le cardinal de l'intersection des mots-clefs d'un processus amont avec les mots-clefs d'un coefficient est égal au nombre d'appariement $\alpha(\mathcal{D}_1, \mathcal{D}_c)$, cela signifie que les deux éléments partagent une indexation commune et peuvent donc potentiellement faire partie d'une relation dans le graphe détaillé. Le résultat de la combinaison de ces deux éléments sera indexé par l'union de leurs mots-clefs. La condition $\text{card}((K_{p_1} \cup K_c) \cap K_{p_2}) = \alpha(\mathcal{D}_1 \cup_D \mathcal{D}_c, \mathcal{D}_2)$ jouera donc le même rôle pour l'association de cette combinaison avec un processus aval que dans le cas

6. Cette opération n'est pas identique à l'union de deux ensembles ou à l'union de deux familles d'ensembles. Par exemple, soit $\mathcal{D}_1 = \{\{A, B, C\}\}$ et $\mathcal{D}_2 = \{\{A, B, D\}, \{E, F\}\}$, on a $\mathcal{D}_1 \cup_D \mathcal{D}_2 = \{\{A, B\}, \{E, F\}\}$ car $\{A, B, C\} \cap \{A, B, D\} = \{A, B\}$ et $\{E, F\}$ n'a pas d'intersection non nulle avec une dimension de \mathcal{D}_1 .

d'un processus amont et d'un coefficient. L'union des deux ensembles de dimension \mathcal{D}_1 et \mathcal{D}_c avec l'opérateur \cup_D est le pendant de l'union des mots-clefs d'un processus amont avec les mots-clefs d'un coefficient.

Le calcul du nombre d'appariement entre deux ensembles de dimensions nécessite de compter le nombre de couples de dimensions dont l'intersection est non nulle. Comme les deux ensembles sont consistants, nous pouvons stocker tous les mots-clefs du deuxième groupe dans une table de hachage et tester si l'un des mots-clefs de chaque dimensions du premier ensemble se retrouve dans cette table. De façon similaire, nous pouvons calculer l'union de deux ensembles de dimensions \mathcal{D}_1 et \mathcal{D}_2 en utilisant une table de hachage pour déterminer si un mot-clef de \mathcal{D}_1 se trouve dans une dimension de \mathcal{D}_2 . Nous pouvons donc simplement parcourir les dimensions de \mathcal{D}_1 et, si elle a une intersection non nulle avec une dimension de \mathcal{D}_2 , stocker cette intersection tout en marquant la dimension de \mathcal{D}_2 correspondante comme déjà utilisée. Il nous suffit ensuite de rajouter toutes les dimensions de \mathcal{D}_2 non utilisées dans le résultat. La table de hachage contenant les mots-clefs de \mathcal{D}_2 indique quelle est la dimension d'appartenance des mots-clefs. Cet algorithme est présenté dans le listing 1, nous considérons que nous disposons d'une structure de données permettant de stocker des ensembles de dimensions et une autre pour stocker des dimensions. Ces deux structures possèdent un opérateur += permettant d'ajouter un élément.

Algorithme 1: Calcul de l'union de deux ensembles de dimensions

```

Entrées :  $\mathcal{D}_1, \mathcal{D}_2$ 
tableHachage  $\leftarrow$  table de hachage des mots-clefs de  $\mathcal{D}_2$ 
union  $\leftarrow$   $\emptyset$  // union est un ensemble de dimensions
pour chaque dimension  $\in \mathcal{D}_1$  faire
    nouvelleDimension  $\leftarrow$   $\emptyset$ 
    pour chaque motClef  $\in$  dimension faire
        si motClef  $\in$  tableHachage alors
            nouvelleDimension += motClef
            marquer dimension comme utilisée
        si nouvelleDimension =  $\emptyset$  alors
            union += dimension
        sinon
            union += nouvelleDimension
    pour chaque dimension  $\in \mathcal{D}_2$  non utilisée faire
        union += dimension
retourner union

```

L'algorithme présenté dans le listing 2 présente la fonction permettant de créer les relations détaillées correspondant à une macro-relation $((G_{P1}, G_{P2}), G_c)$. Nous considérons que nous avons une structure de données pour stocker des processus et des coefficients qui permet d'accéder aux mots-clefs de leurs indexations via l'appel à *element.motClef*. Nous avons aussi une structure de données permettant de stocker les relations détaillées qui possède un opérateur += permettant d'ajouter des éléments.

Algorithme 2: Conversion d'une macro-relation en un ensemble de relations détaillées

Entrées : G_{P1}, G_{P2}, G_c
relationsDétaillées $\leftarrow \emptyset$
pour chaque $p_1 \in G_{P1}$.processus **faire**
 pour chaque $c \in G_c$.coefficients **faire**
 si p_1 .motClef \cap c .motClef = $\alpha(\mathcal{D}_1, \mathcal{D}_c)$ **alors**
 pour chaque $p_2 \in G_{P2}$.processus **faire**
 si $(p_1$.motClef \cup_D c .motClef) \cap p_2 .motClef =
 $\alpha(\mathcal{D}_1, \mathcal{D}_c)$ **alors**
 relationsDétaillées += $((p_1, p_2), c)$
retourner relationsDétaillées

Nous avons réalisé des tests de passage à l'échelle de cet algorithme pour nous assurer qu'il était possible de convertir un grand nombre de macro-relations dans un temps raisonnable. Cette expérimentation a consisté à tester la conversion d'une macro-relation impliquant un nombre de processus croissant. Nous avons donc généré un groupe de processus amont avec une seule dimension et un nombre de mots-clefs croissant. Les coordonnées des cellules de ce groupe ne comportant qu'un seul mot-clef, nous avons autant de processus que de mots-clefs dans la dimension. Par conséquent, ce groupe contient un nombre croissant de processus. Puis nous avons généré un groupe de coefficients et un groupe de processus aval à partir de ce groupe amont. Ce tests a été réalisé sur une machine équipé d'un Core i5-750 avec 4Go de mémoire. Les résultats sont présentés dans la Figure 7. La perte de continuité dans le temps de traitement aux alentours de 5 000 processus est dû à une ré-allocation de la mémoire des structures de données permettant de stocker les relations détaillées. Nous nous sommes limité à 10 000 processus maximum car la base de données d'inventaire la plus conséquente ne compte que 4 000 processus(Frischknecht, 2005b).

5. Etude de cas : production d'électricité aux Etats-Unis

Le National Renewable Energy Laboratory (NREL) fournit une base de données d'inventaires en cycle de vie contenant, entre autres, les données environnementales de la production d'électricité pour les 27 sous-régions énergétiques du territoire américain(NREL, 2011). Ces sous-regions sont définies par l'agence de protection de l'environnement américaine (U.S. EPA) et la Generation Resource Integrated Database (eGRID)(Pechan, 2010). Chaque sous-région de l'eGRID représente une portion du territoire américain qui ont des mix énergétiques similaires et qui sont potentiellement isolés en raison de contraintes de transmission. Les données d'inventaire en cycle de vie du NREL peuvent être exportées au format ecospold(Frischknecht, 2005a). Ce format d'échange est largement utilisé dans la communauté liée à l'ACV. En plus d'inclure des méta-données pour chaque processus (source, commentaires,

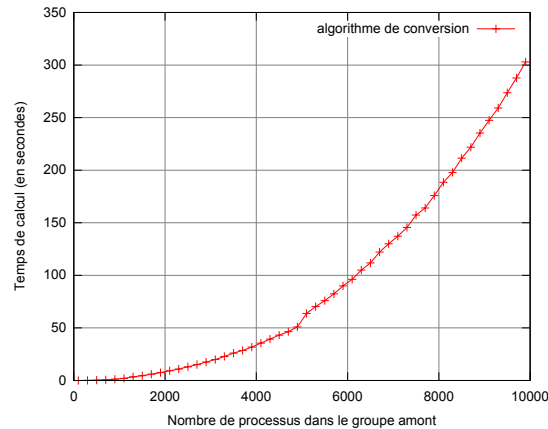


Figure 7. Temps de calcul de l'algorithme de conversion des macro-relations en fonction du nombre de processus dans le groupe amont.

etc.), ce format inclus les relations de dépendance entre les processus ainsi que les coefficients de dépendance. Certains processus de cette base de données ne sont pas détaillés (et marqués comme "factice"), nous n'avons donc pas un jeu de données complet en terme de relations de dépendances. Mais même après l'avoir élagué de ces processus, il reste suffisamment complexe pour illustrer notre proposition.

Les figures 8 et 9 présentent, respectivement, le graphe détaillé et le macro-graphe correspondant à ces données limitées à 7 sous-régions. Nous obtenons 27 processus et 72 relations de dépendances dans le graphe détaillé. Alors que dans le macro-graphe nous n'avons besoin que de 13 groupes de processus et 17 macro-relations (nous avons donc 17 groupes de coefficients). Il est aussi possible de réduire le nombre de groupes de coefficients nécessaire à 12. Par exemple, le groupe de processus *Transports* peut être relié à chaque groupe de processus de production d'électricité à l'aide d'un groupe de coefficients regroupant tous les coefficients de dépendance entre les différents modes de transport et les processus de production d'électricité (d'une façon similaire à l'illustration de la méthode de conversion tel que présentée dans la Figure 6). Nous pouvons aussi utiliser seulement 8 groupes de coefficients si nous stockons dans un seul groupe tous les coefficients de dépendances entre les différents processus de production d'électricité et le groupe contenant les processus de production d'électricité dans les différentes régions de l'eGRID.

6. Conclusion et travaux futurs

Nous avons proposé une nouvelle approche pour modéliser les inventaires en cycle de vie basée sur l'utilisation d'une ontologie et des regroupements sémantiques des processus. L'intérêt principal de cette proposition est d'offrir un modèle plus compréhensible des bases de données d'inventaires. Cette approche permet aussi d'avoir

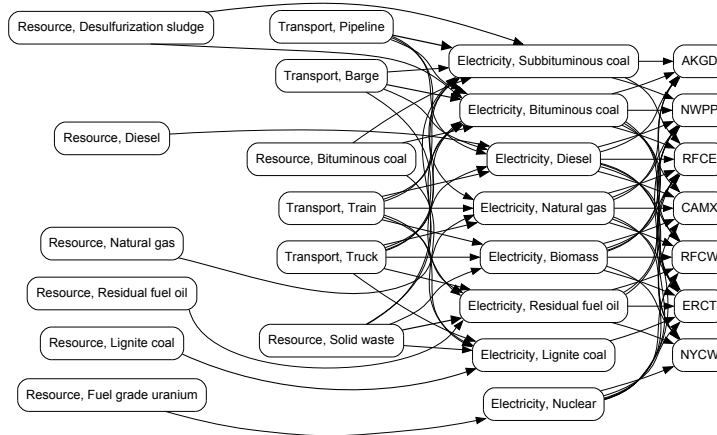


Figure 8. Le graphe détaillé correspondant au jeu de données extrait de la base de données d'inventaire en cycle de vie du NREL.

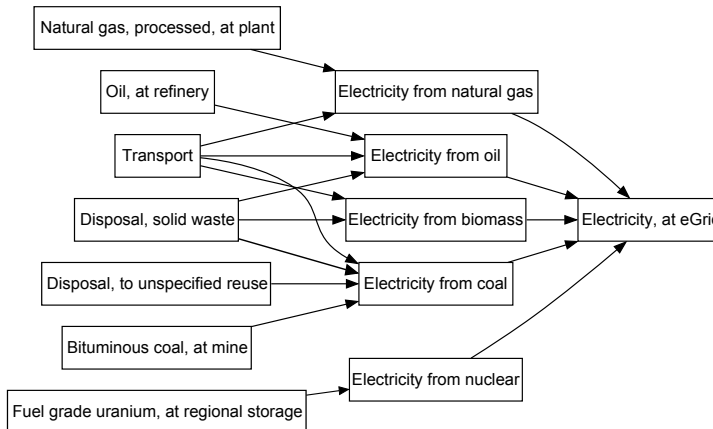


Figure 9. Le macro-graphe correspondant au jeu de données extrait de la base de données d'inventaire en cycle de vie du NREL.

une modélisation basée sur la manipulation d'une ontologie pour créer de nouvelles relations entre des processus.

Nous prévoyons d'implémenter notre méthode entièrement en OWL (McGuinness, 2009) à l'aide de règles exprimées en logique de description et d'un raisonneur. Nous pourrions alors réaliser une étude comparative des performances de notre implémentation et d'une version entièrement en OWL. L'indexation des processus à l'aide de mots-clés stockés dans une ontologie pourrait aussi permettre de répondre à des re-

quêtes du type : quel est l'impact des processus indexés avec un mot-clef (ou un ensemble de mots-clefs) sur un processus spécifique. Nous pourrions, par exemple, déterminer quel est l'impact des processus de transport sur la production d'électricité pour une sous-région spécifique de l'eGRID. Cette requête peut être traitée en restreignant le calcul aux processus amonts indexés avec un mot-clef spécifique.

7. References

- Beckett D., Turtle Terse RDF Triple Language, W3C recommendation, 2011.
- Frischknecht R., Jungbluth N., Althaus H., et al., « Introduction The ecoinvent Database : Overview and Methodological Framework », *The International Journal of Life Cycle Assessment*, vol. 10, n° 1, p. 3-9, 2005a.
- Frischknecht R., Rebitzer G., « The ecoinvent database system: a comprehensive web-based LCA database », *Journal of Cleaner Production 2005*, vol. 4, n° January, p. 1337-1343.
- GaBi, « GaBi Life Cycle Inventory Databases », <http://www.gabi-software.com>. Date de dernière consultation : 16/12/2011.
- Gruber T. R., « A Translation Approach to Portable Ontology Specifications », *Knowledge Acquisition*, vol. 5, n° 2, p. 199-220, 1993.
- Guinée J. B., *Handbook on Life Cycle Assessment: Operational Guide to the ISO Standards*, Springer, New York, 2002.
- Heijungs R., Suh S., *The computational structure of life cycle assessment*, Kluwer Academic Publishers, Dordrecht, 2002.
- ISO, ISO 14044 (2006): Environmental management – Life cycle assessment – Requirements and guidelines, International Organisation for Standardisation (ISO), 2006.
- Klyne G., Carroll J. J., editors, Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C recommendation, 2004.
- Leontief W., *Input-Output Analysis*, University of British Columbia Press, p. 53-83, 1986.
- McGuinness D. L., van Harmelen F., editors, OWL 2 Web Ontology Language Document Overview, W3C recommendation, 2009.
- McGuinness D. L., *Ontologies come of age*, vol. 0, MIT Press, p. 171–192, 2002.
- Nicholson W., *Elementary linear algebra with applications*, PWS-Kent Publishing Company, 1990.
- NREL, « U.S. Life Cycle Inventory Database », <http://www.nrel.gov/lci/>. Date de dernière consultation : 01/09/2011.
- Pechan & Associates Inc., The Emissions & Generation Resource Integrated Database for 2010 (eGRID2010) Technical Support Document, 2010.
- Peters G. P., « Efficient Algorithms for Life Cycle Assessment, Input-Output Analysis, and Monte-Carlo Analysis », *The International Journal of Life Cycle Assessment*, vol. 12, n° 6, p. 373-380, 2007.
- Prud'hommeaux E., Seaborne A., editors, SPARQL Query Language for RDF, W3C recommendation, 2008.
- Varga R., *Matrix Iterative Analysis*, Springer Series in Computational Mathematics, Springer Verlag, 2010.