



HAL
open science

Observatoire virtuel pour diffuser et réutiliser l'expertise en télédétection

Laurent Demagistri, Yuan Lin, Thérèse Libourel Rouge

► To cite this version:

Laurent Demagistri, Yuan Lin, Thérèse Libourel Rouge. Observatoire virtuel pour diffuser et réutiliser l'expertise en télédétection. Spatial Analysis and Geomatics - SAGEO 2012, Nov 2012, Liège, Belgique. hal-01830214

HAL Id: hal-01830214

<https://hal.science/hal-01830214>

Submitted on 4 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Observatoire virtuel pour diffuser et réutiliser l'expertise en télédétection

Laurent Demagistri* — Yuan Lin* — Thérèse Libourel*

* UMR ESPACE-DEV (IRD-UM2-UAG-ULR)
500 Avenue JF Breton
F-34000 Montpellier - nom.prenom@ird.fr

RESUME. L'observation de la Terre et de l'Espace à l'aide d'images satellitaires donne lieu à la mise en place de plusieurs organisations : "réseaux" de stations de réception. Leur principale vocation est de diffuser les produits bruts ou élaborés (les images). Cependant pour augmenter les potentialités offertes aux chercheurs, il serait souhaitable que ceux-ci puissent facilement décrire et exécuter leurs propres expérimentations en concentrant ainsi leur effort sur les traitements qu'ils souhaitent effectuer sur les données. L'objectif que nous nous sommes fixé est de faciliter l'expression et l'exécution de telles expérimentations par la mise en place d'observatoires virtuels conçus à partir des réseaux de stations de réception existants. Notre proposition sera détaillée sur un exemple traité à partir du réseau SEASnet.

ABSTRACT. Earth and Space observation by means of satellite images gives rise to the implementation of several organizations: "networks" of receiving stations. Their main vocation is to disseminate raw or elaborated products (the images) to users. However to increase the potentialities that they are given, researchers should also be offered the ability to easily describe and execute their own experiments on the data. They could thus concentrate their effort on the processings that they are designing. We settled the objective of facilitating the definition and the execution of such experiments by the implementation of virtual observatories designed from the existing networks of receiving stations. Our proposal will be detailed on an example treated from the SEAS network (SEASnet).

MOTS-CLÉS : Image satellitaire, Chaîne de traitements, Infrastructure de mutualisation, Environnement de Workflow

KEYWORDS: Satellite image, Processing line, Infrastructure, workflow environment

1. Introduction

L'imagerie satellitaire constitue une source d'information précieuse pour la connaissance et la gestion des observations de phénomènes terrestres. En effet, l'image satellite donne aux chercheurs la possibilité de croiser une vision globale d'une zone avec l'observation plus locale de terrain. Les données brutes recueillies nécessitent un ensemble de traitements avant d'être utilisables. De plus leur pérennisation demande des supports autorisant la sauvegarde et la diffusion aisée. De véritables infrastructures sont donc nécessaires et celles-ci doivent respecter les standards actuels à des fins d'interopérabilité. La mise en place de ce type d'infrastructure autorisant la mutualisation de ces données ainsi que celle des traitements d'analyse adéquats a constitué une des priorités que ce sont fixées l'IRD (Institut de Recherche pour le Développement), et plus précisément l'unité mixte de recherche Espace-Dev, pour répondre aux enjeux actuels liés à l'observation de la Terre et de l'Espace.

La profusion des images actuellement accessibles fait qu'au delà des traitements standards qui doivent être appliqués aux images brutes, les défis concernent les chaînes de traitements plus élaborés qui doivent être établies pour construire des informations utiles ou indicateurs. Les systèmes dédiés aux observations à l'aide d'images satellitaires (que nous désignons observatoires) se multiplient et sont disponibles sur des sites distribués, ce qui complique le travail du scientifique qui souhaite résoudre des problèmes mettant en œuvre divers traitements et données dispersés sur ces divers observatoires. Le challenge que nous souhaitons relever est de concevoir une réponse pragmatique, prenant en compte ces différents aspects. C'est ce que nous désignons par *observatoire virtuel*. Un observatoire virtuel doit permettre aux chercheurs d'accéder aux données et aux traitements (quelle que soit leur localisation), et leur offrir les outils pour mettre au point leurs diverses expérimentations en explicitant et modélisant leur raisonnement.

L'article déclinera, en premier lieu, dans la section 2 la description générale de ce que l'on désigne par infrastructure d'acquisition ou réseau de stations dédiées à l'imagerie satellitaire. La section 3 présentera la notion d'observatoire virtuel, celle de contexte de travail et d'architecture de chaîne de traitements afférents établis à partir d'un bref état de l'art autour de travaux qui ont inspiré cette proposition. La section 4 présentera la déclinaison de notre proposition sur un exemple. La section 5 précisera les perspectives ouvertes.

2. Infrastructure de données satellitaires et traitements (Observatoire)

Notre équipe a eu la chance de mener à bien la mise en place de plusieurs projets d'installation de systèmes d'acquisition d'images satellites durant les vingt dernières années : acquisition d'images des satellites météorologiques à large champ (basse résolution) mais également d'images "haute résolution" de satellites optiques ou radar.

2.1. Les principes généraux

Quel que soit le type de système d'acquisition d'images satellites, trois grands sous-systèmes sont généralement présents :

- le sous-système "Antenne" (l'infrastructure physique de télécommunication et son pilotage),
- le sous-système de production des données images "brutes",
- le sous-système d'exploitation de ces données brutes.

Ces différentes parties sont détaillées sur la figure 1. Nous les décrivons succinctement dans les paragraphes qui suivent.

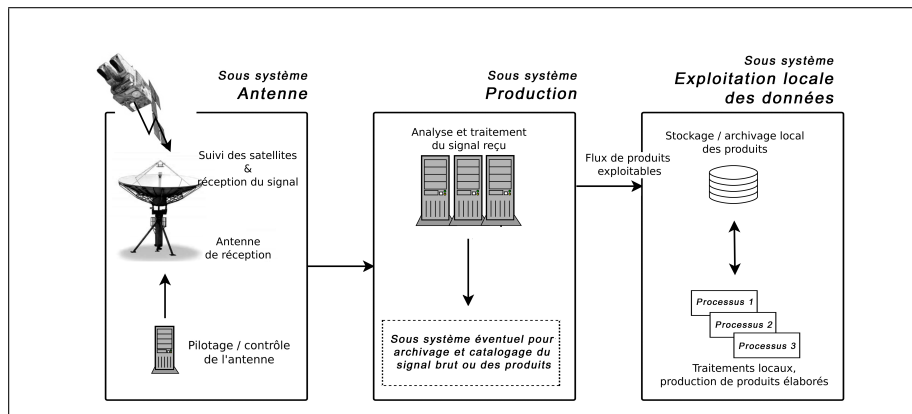


Figure 1 – Infrastructure locale d'acquisition de données satellites (pied d'antenne).

2.1.1. Réception du signal

Les satellites d'observation de la Terre les plus courants sont soit géostationnaires (apparemment immobiles par rapport au point au sol qu'ils survolent) soit à orbites polaires (ils survolent les pôles terrestres à chaque révolution, la rotation de la Terre entraînant un décalage de leur orbite par rapport aux zones survolées). Ils peuvent enregistrer à bord le signal correspondant aux mesures qu'ils effectuent en vol puis le restituer lorsqu'ils survolent une zone de réception ou bien émettre en permanence le signal mesuré. Une antenne motorisée programmée de manière adéquate pour suivre leur trajectoire peut acquérir ces données (fig. 1, sous-système antenne).

2.1.2. Production des images

Dans ce sous-système on trouve tous les équipements permettant de transformer en donnée exploitable pour les utilisateurs le signal brut issu du sous système "antenne". La plupart du temps il s'agit d'équipements "propriétaires" avec des logiciels spécifiques sur lesquels il est en principe impossible d'intervenir. Certains opérateurs

privés incluent à ce stade un système de catalogage des données acquises disponibles et des données produites (fig. 1, sous-système de production).

2.1.3. *L'exploitation des données*

Cette partie héberge l'ensemble des équipements mis en place pour exploiter, en fonction des besoins recensés, les données issues du système d'acquisition (fig. 1, sous-système d'exploitation des données).

Au minimum il s'agit d'un système de stockage suffisamment bien dimensionné : en effet, en fonction des caractéristiques des données acquises, le flux quotidien issu de l'antenne peut être conséquent.

Malgré une diffusion "grand public" de plus en plus répandue, les données satellitaires restent la plupart du temps largement sous-exploitées. Ce constat impose de mettre en place tous les outils permettant d'exploiter au mieux leur potentiel avec, en particulier, un effort particulier mis sur les capacités de diffusion : faciliter donc l'accès direct aux données (accès aux mesures "télé-détections") mais également publication sur internet via un catalogage normalisé de leurs métadonnées. Les utilisateurs potentiels ne disposant pas tous nécessairement d'outils de télé-détection ou de traitement d'images suffisamment évolués, un objectif serait également de faciliter le traitement de haut niveau des données, via la mise en place d'infrastructures de traitement des données à distance.

Dans ces conditions, tout utilisateur potentiel, pourrait avoir accès non seulement à des données de télé-détection mais également aux traitements standards de ce type d'images : prétraitements, corrections radiométriques, corrections géométriques, calcul d'indices, etc. De plus, il devrait aussi accéder aux produits dérivés mis en place par les producteurs à partir des données brutes. Ces produits dérivés, parfois complexes, constituent pour les utilisateurs des "boîtes noires" spécifiques et de ce fait ne sont pas facilement réutilisables. Pour aller au delà, il serait judicieux d'ouvrir ces "boîtes noires" afin d'offrir des traitements simples, réutilisables et combinables, à loisir, par les scientifiques. L'infrastructure d'exploitation doit donc comporter des serveurs de traitement puissants, afin d'offrir aux utilisateurs le loisir de concevoir leurs expérimentations, en connaissance de cause, à partir des différents traitements disponibles (simples / complexes).

2.2. *Expérience de notre équipe*

Dès 1991, dans le cadre de ses propres programmes de recherche et développement, l'Institut de Recherche pour le Développement (IRD) a développé et exploité plusieurs stations de réception d'images des satellites météorologiques à orbite polaire (NOAA/AVHRR et ORBview/SeaWiFS). Les capteurs à bord de ces satellites permettent d'acquérir en continu des données multispectrales. Ils ont une résolution spatiale à la verticale du satellite, de l'ordre du kilomètre et lors d'un passage au dessus d'une région, la fauchée (largeur de la zone au sol imagée) est de l'ordre de

2900 km. Ces caractéristiques font de ces systèmes d'acquisition de très bons outils de suivi à grande échelle de l'environnement. En 1999, les différentes stations se sont regroupées pour donner naissance au réseau SEASnet [PET 02]. La figure 2 présente la localisation des différentes antennes d'acquisition du réseau SEASnet avec une schématisation de leur couverture respective.

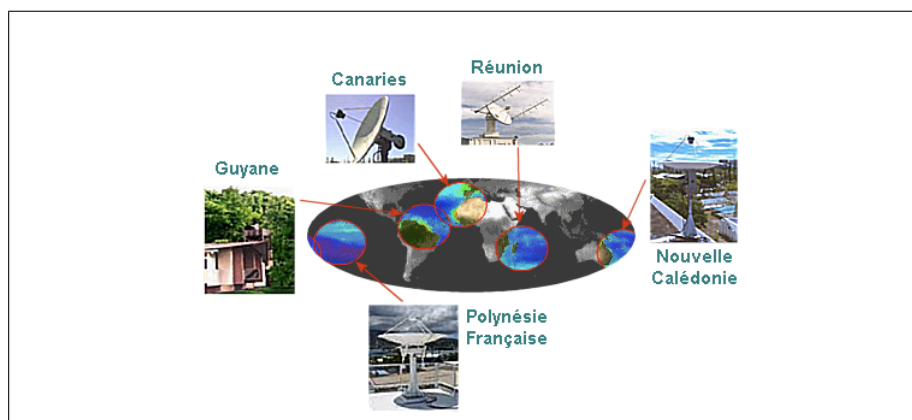


Figure 2 – Localisation des infrastructures d'acquisition du réseau SEASnet.

Cette initiative a été le point de départ de la réflexion que nous menons actuellement : comment mettre à la disposition des chercheurs un *observatoire virtuel* leur permettant la mise au point de diverses expérimentations à partir de ces infrastructures d'acquisition de données réparties. Un tel observatoire autorise la conception et la réalisation de chaînages entre divers services dans l'objectif de construction d'indicateurs pertinents pour l'observation terrestre.

Pour la diffusion de ces données et traitements, plusieurs choix étaient possibles :

- mettre en place un observatoire complet à chaque pied d'antenne,
- centraliser l'accès à ces divers observatoires sur un site dédié (situé à Montpellier)

Dans un premier temps, c'est le système centralisé qui a été préféré (sans créer de réels observatoires à chaque pied d'antenne) ; la figure 3 schématise cette configuration.

Les produits élaborés calculés en pied d'antenne de chacune des implantations locales sont transférés (via le protocole FTP) sur un serveur unique. Différents protocoles d'accès ont été mis en place sur ce dernier. Tout d'abord OPENDAP ¹, très répandu dans la communauté scientifique des océanographes et des météorologues mais également les protocoles de l'OGC ², tout particulièrement le WMS et le WCS (Web Map Service et Web Coverage Service).

1. Open-source Project for a Network Data Access Protocol : <http://opendap.org/>

2. Open Geospatial Consortium : <http://www.opengeospatial.org/standards/is>

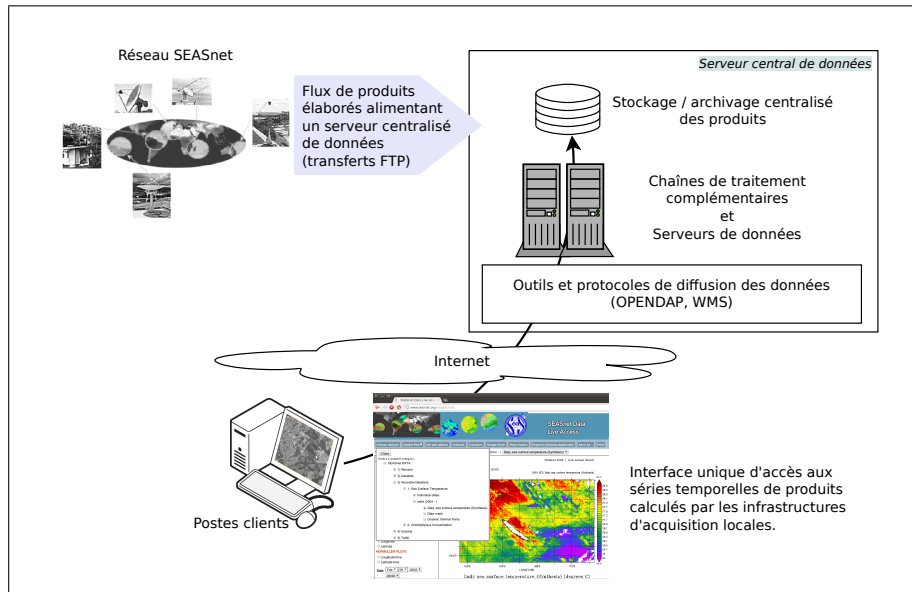


Figure 3 – Observatoire centralisé.

Des bases de métadonnées ont été élaborées afin d'annoter le flux de fichiers calculés quotidiennement. Elles respectent les normes courantes relatives à la gestion des métadonnées d'images d'observation de la Terre (normes ISO19115³ et ISO19115-2⁴). Les images sont indexées géographiquement et temporellement afin de constituer un outil parfaitement adapté à la gestion de séries temporelles de données cartographiques.

Par la suite, en 2006, une infrastructure d'acquisition de données à haute résolution spatiale a été implantée en Guyane, au centre IRD de Cayenne (réception de données optiques SPOT et radar ENVISAT). Le même type de plateforme a été installé à la Réunion en 2011 et, en 2013, le projet SEAS-Gabon⁵ permettra également la mise en place d'une telle infrastructure à Libreville au Gabon. Cette fois, une approche centralisée est difficilement envisageable en raison du poids des données qu'il faudrait transférer quotidiennement (plus de 500 Mo pour une image SPOT5 panchromatique et des performances de réseau pas toujours optimales).

Sur l'ensemble de ces installations l'objectif est de mutualiser les développements. Nous travaillons donc à l'harmonisation des systèmes d'exploitation et de diffusion

3. ISO 19115 :2003 : Information géographique – Métadonnées, http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

4. ISO 19115-2 :2009 : Information géographique – Métadonnées – Partie 2 : Extensions pour les images et les matrices, http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=39229

5. <http://seas-gabon.espace-dev.fr/seasgabon/> et <http://www.ageos.org/ageos>

des données grâce à l'utilisation d'infrastructures de données spatiales (SDI) normalisées.

La démarche se veut la plus générique possible afin de tendre vers l'interopérabilité des systèmes.

3. Observatoire virtuel

Pour rendre l'architecture décrite précédemment encore plus générique, et se mettre réellement dans une configuration d'*observatoire virtuel* tel que nous l'imaginons, l'infrastructure de la figure 3 devra évoluer, au fur et à mesure des démarches de standardisation que nous menons, vers une architecture entièrement normalisée schématisée par la figure 4. L'accès aux données et traitements se fera directement sur les infrastructures informatiques de chaque site de réception d'images ; et ce de manière complètement transparente et virtuelle pour l'utilisateur.

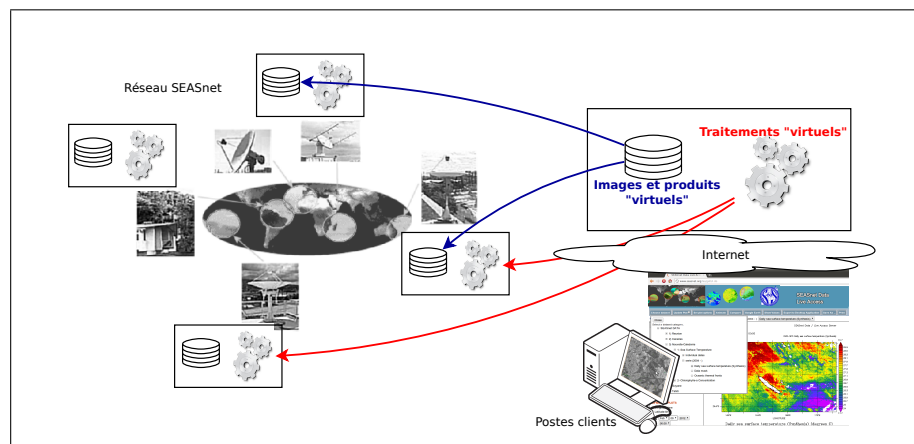


Figure 4 – Vers un observatoire virtuel SEASnet : accès direct et combinaison des données et traitements de chaque site.

Notre idée est que les chercheurs sont amenés à construire leurs propres expérimentations pour vérifier et valider leurs hypothèses. Pour construire ces expérimentations complexes, ils ont souvent besoin de rechercher les ressources adéquates et de les organiser ou réorganiser.

Un observatoire virtuel doit mettre à leur disposition ces données et les traitements existants (quelle que soit leur localisation) ; il doit en outre leur offrir un environnement pour concevoir et réaliser les expérimentations correspondant aux raisonnements qu'ils envisagent de mettre en œuvre.

Chaque plan d'expérimentation doit pouvoir être conçu, sauvegardé afin de pouvoir être réexécuté plusieurs fois soit dans diverses configurations, soit sur diverses données de test.

Ces deux observations nous amènent tout naturellement aux notions de workflows et de contextes de travail associés.

3.1. *État de l'art*

Nous établissons ici un rapide état de l'art autour des propositions existantes en termes de protocole d'accès aux données et de workflow scientifique.

3.1.1. *Diffusion de données*

Beaucoup de produits et images de télédétection sont actuellement accessibles sur internet. Plusieurs organisations mettent en ligne des sites de diffusion de données. Trois exemples nous paraissent particulièrement pertinents :

– Le site du "Physical Oceanography Distributed Active Archive Center" (PO.DAAC⁶) fournit gratuitement des produits d'océanographie physique (courants, température, vents, salinité, etc.).

– Le site du "Global Land Cover Facility" (GLCF⁷) de l'Université du Maryland met à disposition des utilisateurs des produits et des données pour l'étude de la Terre et de l'environnement, particulièrement les images des satellites LANDSAT⁸.

– Du côté européen, l'Organisation européenne pour l'exploitation des satellites météorologiques (EUMETSAT⁹) fournit sur son site des données, images et produits satellitaires dédiés à la météorologie et la climatologie.

La plupart du temps ces sites proposent de récupérer leurs fichiers via des protocoles standards afin que les utilisateurs puissent les exploiter localement sur leur poste. En fonction du type de données, ces fichiers peuvent être lourds et nécessitent un certain degré de performances en termes de connexion réseau ou de capacité de stockage. Le protocole le plus couramment utilisé est le transfert de fichiers "FTP", certes simple d'utilisation mais ne permettant pas de mettre en place des outils élaborés de recherche de données.

Le protocole **OPeNDAP**, cité précédemment, s'est beaucoup développé dans les communautés scientifiques de l'océanographie et de la météorologie, donnant naissance à différents outils, quasiment "clé en main", permettant une diffusion de données sur un serveur WEB à l'aide de formulaires de requête. Nous citons ici ceux qui ont été utilisés dans le cadre du réseau SEASnet : le *middleware* **THREDDS** (Thematic Realtime Environmental Distributed Data Services¹⁰) permettant d'activer un

6. Le PO.DAAC est un service du "Jet Propulsion Laboratory", un laboratoire dépendant de la

NASA - <http://podaac.jpl.nasa.gov/dataaccess>

7. <http://glcf.umiacs.umd.edu/data/landsat/>

8. <http://landsat.gsfc.nasa.gov/>

9. <http://www.eumetsat.int/Home/Main/DataAccess/index.htm>

10. <http://www.unidata.ucar.edu/projects/THREDDS/>

service OPeNDAP sur un serveur d'applications et le **LAS** (Live Access Server¹¹) encapsulant les opérations THREDDS de base et offrant des possibilités plus élaborées de recherche, d'exploitation voire de traitement en ligne des données.

Les protocoles de l'*OGC*, également cités précédemment, sont de plus en plus utilisés, particulièrement en lien avec l'effort de normalisation des données géographiques mené conjointement avec l'*ISO*.

3.1.2. Environnements de workflow scientifique

La notion de workflow scientifique existe depuis longtemps ; en traitement d'images, des outils comme Khoros [KON 94] ont montré le chemin en insistant sur la notion d'outil dédié à la conception de chaînes de traitements et à leur paramétrisation au travers d'expérimentations itératives.

Une rapide analyse, issue des travaux de [LIN 11] nous a permis de sélectionner un ensemble de projets et de démarches existants. Chacun des projets de workflow (scientifique ou généraliste) retenus propose un environnement de travail reposant sur une interface graphique et des formalismes divers, grâce auquel les utilisateurs construisent leurs plans de travail.

Kepler [LUD 06] est un environnement complet pour le workflow scientifique construit sur la plateforme Ptolemy II de l'Université de Berkeley. Dans cet environnement, les *acteurs* correspondent aux différents traitements et opérations envisageables, et ils sont dotés de *ports* qui représentent leurs paramètres d'entrée / sortie. Les compositions entre traitements sont faites interactivement par les scientifiques en reliant les *ports* d'*acteur* par des *chanel*s. Le contrôle et l'orchestration de modèle de workflow est confié à des *directeurs*. Les adaptations nécessaires sont réalisées par des programmes intermédiaires (*senders* et *receivers*), qui assurent la compatibilité des données transférées sur un *chanel*.

Taverna [HUL 06] est un projet de workflow initialisé par l'équipe *myGrid* en Angleterre, utilisé principalement dans les domaines biologiques. Les traitements dans cet environnement sont essentiellement des services web (auxquels peuvent s'adjoindre des bibliothèques locales, des scripts manuscrits, etc.). Lors de la composition de traitements, l'utilisateur apparie manuellement les paramètres d'entrée / sortie des services web, ou invoque des *shim services*, adaptateurs spécifiques conçus antérieurement à partir d'appariements établis dans des expérimentations déjà construits et testés.

NetBeans est un environnement IDE généraliste, dont un des modules intégrés permet de construire un workflow par composition de services web en utilisant le langage BPEL (Business Process Execution Language) [AND 03]. Une bonne connaissance du standard BPEL est un pré-requis. La composition se fait par appariement ou transformation manuelle entre éléments de messages échangés, ces règles d'apparie-

11. LAS (<http://ferret.pmel.noaa.gov/LAS>) est un logiciel Open-Source développé par un laboratoire de la NOAA (National Oceanic and Atmospheric Administration)

ment sont ensuite traduites à l'aide du langage XSLT (eXtensible Stylesheet Language Transformations) [KAY 07].

Weka [CUN 93] est une application du domaine apprentissage et fouille de données, réalisée par l'Université Waikato de Nouvelle Zélande. Elle comporte un composant **Weka KnowledgeFlow**, qui permet un chaînage de traitements relatifs aux expérimentations de fouille de données. Le modèle général de KnowledgeFlow obéit à la chaîne *Sélectionner les données* → *Filtrer* → *Classifier* → *évaluer* → *Visualiser*. Grâce à l'interface graphique, les scientifiques concrétisent interactivement leurs expérimentations, et choisissent les convertisseurs pré-établis pour assurer la compatibilité de leurs workflows. L'environnement repose sur des catégories de données et d'algorithmes relatifs aux différents traitements précédemment construits.

Tous ces projets disposent d'un environnement plus ou moins élaboré, et nous proposons de formaliser celui-ci sous la forme de ce que nous désignons par *contexte de travail*.

3.2. Contexte de travail

Le contexte de travail¹² (fig.5) proposé est composé de trois grandes organisations de descriptions de ressources, dénommées :

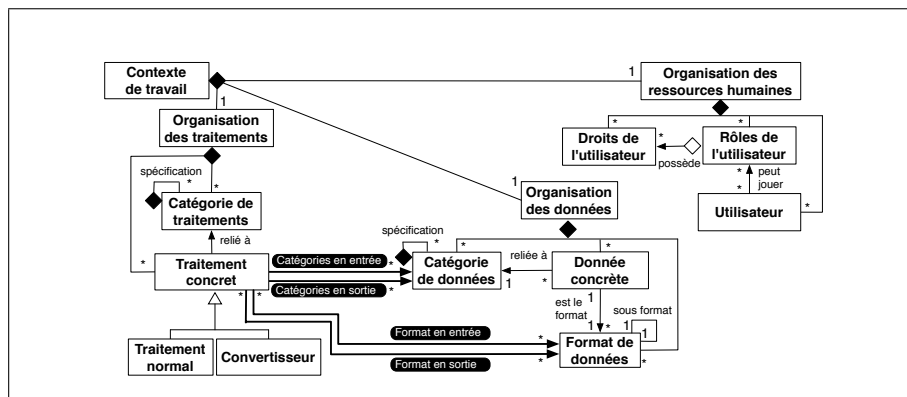


Figure 5 – Contexte de travail

– *Organisation des ressources humaines*, qui gère la description des utilisateurs de la plateforme ainsi que celles de leurs différents rôles et droits d'accès associés,

– *Organisation des données*, qui gère la description des catégories de données, des données concrètes et des divers formats de données associés¹³.

12. Pour simplifier la figure, on a noté traitements, données et ressources humaines, alors qu'il s'agit de descriptions de ces ressources.

13. Remarque : il est à noter que plusieurs catégories de données peuvent partager le même format.

– *Organisation des traitements*, qui gère la description des catégories de traitements et des traitements concrets.

La notion de *Convertisseur* introduite dans la figure 5 correspond à la notion de traitement spécifique ayant en charge la réalisation d’adaptation entre différents formats relevant d’une même catégorie de données.

Pour la construction de cet environnement, un formalisme simple, dédié aux descriptions de ressources a été proposé et formalisé en utilisant XML schéma¹⁴. La figure 6 montre les structures schématiques des descriptions de catégories de données et de traitements, de formats de données, et de données et traitements eux-mêmes. Les propriétés de ces ressources sont notées entre parenthèses, les différents niveaux de la structure représentent les sous-éléments de celle-ci, les chiffres indiqués devant chaque élément représentent leurs cardinalités. Les relations entre ces descriptions sont garanties par les références enregistrées dans les différents éléments, par exemple l’élément *DataCategoryRef* dans l’élément *Input* d’une description de traitement pointe vers la catégorie de données que ce paramètre utilise.

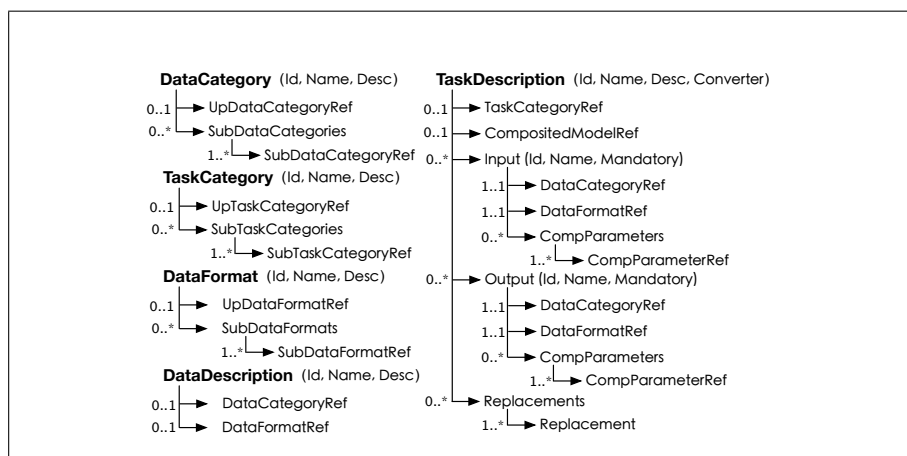


Figure 6 – Représentation schématique du formalisme simple dédié aux descriptions de ressources

Pour illustrer la notion de contexte de travail, nous prenons un exemple concret relatif au domaine (fig.7). La partie supérieure de chaque hiérarchie (traitements et données) représente un ensemble de catégories (représentées par des ovales), ordonnées selon la relation de généralisation / spécialisation. Les descriptions des ressources concrètes (données ou traitements) sont ensuite associées à leur catégorie.

14. Ces fichiers XML Schémas sont mis en ligne, et accessibles sur la page <http://www.lirmm.fr/~lin/project>. Nous n’avons cependant pas détaillé l’organisation de ressources humaines.

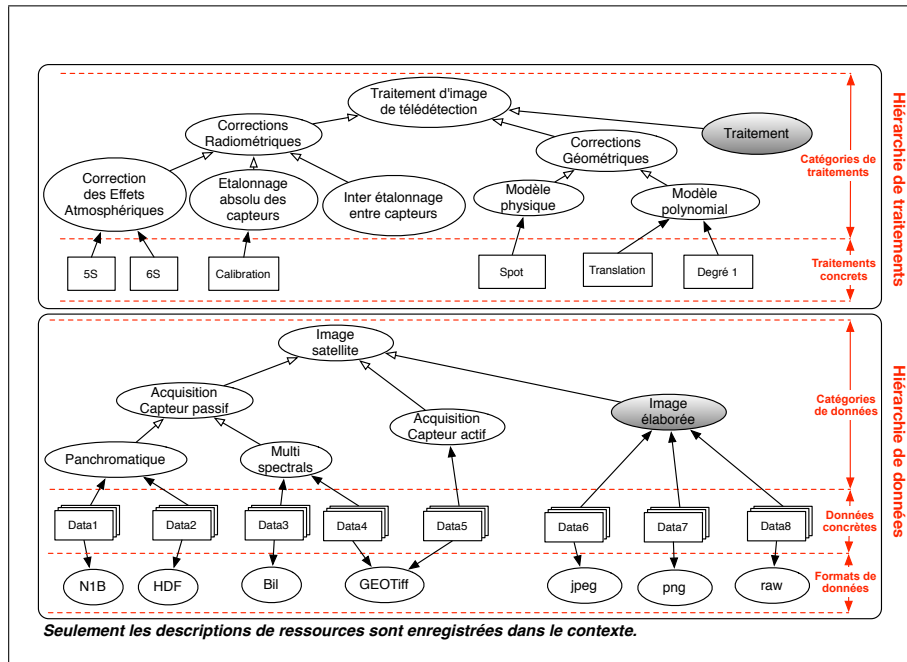


Figure 7 – Un exemple de contexte de travail

Pour une meilleure compréhension de la figure 7, nous explicitons ci-dessous les éléments qui la constituent.

Dans le domaine de l'imagerie satellitale [CAL 01], il est d'usage de distinguer les capteurs passifs des capteurs actifs. Les capteurs actifs, émettent un signal électromagnétique en direction de la cible (le sol Terrestre) et enregistrent la quantité de signal rétrodiffusé par celle-ci. Les capteurs passifs quant à eux, enregistrent le signal électromagnétique émis (cas des capteurs thermiques) ou réfléchi par le sol (réflexion du rayonnement solaire). En raison des processus physiques impliqués, les deux technologies induisent des types d'images différents tant par leur contenu que par leur analyse et les traitements pour les exploiter. Dans la catégorie des capteurs passifs, on peut imaginer faire la distinction entre les capteurs *panchromatiques* et les *multispectraux* :

- Les capteurs panchromatiques mesurent le signal émis par chaque portion élémentaire du terrain, les pixels, sur une large fenêtre spectrale (par exemple tout ou partie du spectre de la lumière visible) ; ils produisent une image comportant une seule "bande".
- Les capteurs multispectraux font cette mesure sur plusieurs fenêtres spectrales de largeurs variables (par exemple les bandes spectrales du rouge, du vert, du proche infrarouge, etc.) ; ils produisent une image comportant plusieurs canaux.

La résolution des capteurs étant dépendante des performances spectrales, les capteurs panchromatiques présentent un meilleur pouvoir séparateur (la capacité à distinguer deux éléments distincts du sol espacés d'une certaine distance). Quoi qu'il en soit, ce paramètre sera bien évidemment un attribut de la classe "Image satellite".

De la même manière que pour les données, on peut essayer de bâtir une classification des traitements habituellement réalisés à partir d'une image satellite. Quel que soit le mode d'acquisition, on se retrouve confronté à des problèmes de corrections géométrique et radiométrique des images : les prétraitements [CAL 01].

Les corrections géométriques visent à corriger les déformations liées à l'angle d'acquisition, à la courbure de la Terre, etc., afin de mettre l'image en conformité géographique (projection de l'image) ou pour la rendre superposable à une source de référence (par exemple une autre image). On distingue les modèles physiques qui modélisent précisément la géométrie d'acquisition du capteur et les modèles polynomiaux plus génériques.

Les corrections radiométriques ont pour objectif la calibration des valeurs numériques contenues dans les images afin d'homogénéiser les données acquises à différentes dates ou par différents capteurs [BRO 85]. On range dans cette catégorie la correction des effets atmosphériques, la calibration "absolue" permettant de convertir les "comptes" numériques (ou niveaux de gris) d'une image en valeurs géophysiques (albedo, températures de brillance, coefficient de rétrodiffusion, etc.), l'égalisation radiométrique entre plusieurs images, etc.

Les images pré-traitées subissent des "Traitements" pour constituer des "Images élaborées" (en grisé sur la fig.7).

Un ensemble de formats de données (*NIB, HDF, Bil, GEOTiff, jpeg, png, raw*) est aussi présenté. Lors de l'enchaînement des traitements, prétraitements radiométriques et géométriques d'une part puis traitements élaborés d'autre part, des images intermédiaires sont produites dans ces formats d'images standards.

3.3. Chaîne de traitements (workflow)

La réflexion menée a été guidée par le point de vue "métier" (expérimentateurs). Réaliser un protocole expérimental correspond au modèle général comportant trois étapes : 1) **Définition** : définition abstraite d'une chaîne de traitements correspondant à une expérimentation (planification d'expériences), 2) **Instanciation** : définition plus spécifique après identification des divers éléments de la chaîne (données / traitements), 3) **Exécution** : exécution personnalisée (selon des stratégies correspondant aux besoins).

3.4. Architecture

C'est à partir de ce cycle de vie expérimental et inspirés par les styles architecturaux proposés par l'Object Management Group [OMG 06], que nous avons proposé pour le composant workflow la vision architecturale à 3-niveaux (fig.8) :

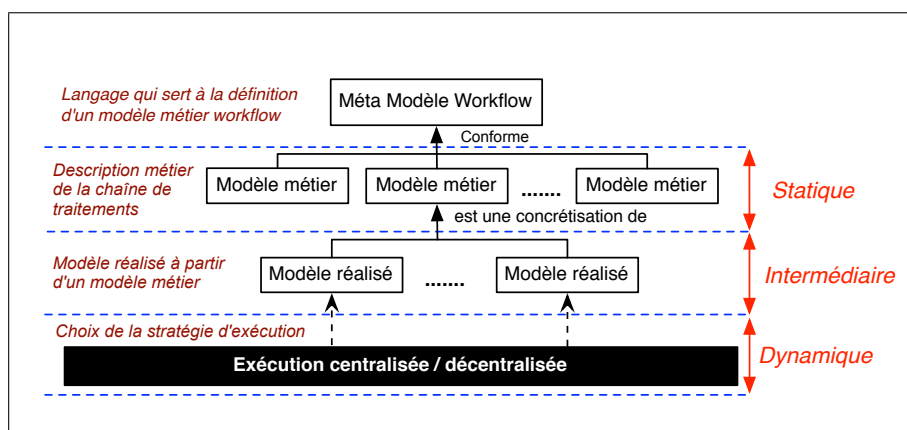


Figure 8 – Architecture du composant workflow

1) Le niveau **statique** concerne la phase de conception. Il s'agit de construire des modèles de traitement métier (abstrait) à partir d'un langage simple.

2) Le niveau **intermédiaire** traduit une phase d'instanciation et de pré-contrôle. à partir du modèle du traitement métier, l'utilisateur construira la chaîne de traitements réelle en déterminant et localisant les traitements et les données les plus adéquates à l'expérimentation visée. Cette phase sera réalisée à l'aide de moteur de recherche à partir des descriptions de ressources nécessaires (fig.6).

3) Le niveau **dynamique** concerne la phase d'exécution proprement dite. Celle-ci peut alors se dérouler à partir de différentes stratégies définies à la fois par l'expérimentateur et par les configurations opérationnelles.

Le niveau *statique* a été approfondi et étudié dans [LIN 09, LIB 10].

Il se base sur un langage défini par un méta modèle dont les éléments abstraits *tâches* ou *traitements* sont reliés par les liens unidirectionnels, et par l'intermédiaire de *port*. Pour faciliter la manipulation des chaînes de traitements abstraits, un langage graphique correspondant a été proposé (fig.9) ; il sera utilisé dans les exemples ultérieurs.

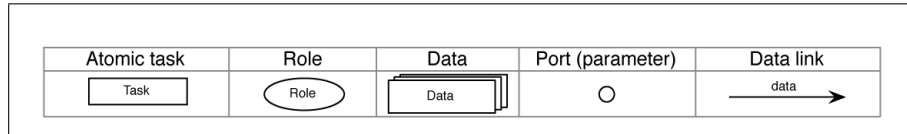


Figure 9 – Extrait des éléments essentiels de notre langage graphique

4. Illustration

4.1. Données et traitements de départ

Pour illustrer l'intérêt de l'architecture proposée ici, nous détaillons en exemple la chaîne de traitements aboutissant à l'obtention de cartes de température de surface de la mer à partir des données issues de notre réseau de stations d'acquisition d'images basse-résolution. Cette chaîne de traitement a été déployée dans chacune de nos implantations "pied d'antenne" basse-résolution.

Les données sont acquises via la réception des satellites américains NOAA¹⁵ (capteurs AVHRR-2 et AVHRR-3). Trois sont actuellement considérés comme opérationnels. Les mesures satellitaires sont réalisées par des radiomètres multicanaux capables d'enregistrer le rayonnement électromagnétique de la Terre selon six longueurs d'onde.

Canal	Résolution au Nadir (km)	Longueur d'onde (μm)
1	1.09	0.58 - 0.68
2	1.09	0.725 - 1.00
3A	1.09	1.58 - 1.64
3B	1.09	3.55 - 3.93
4	1.09	10.30 - 11.30
5	1.09	11.50 - 12.50

Tableau 1 – Résolution et longueurs d'onde des canaux AVHRR-3.

Ces satellites étant *héliosynchrones à orbites polaires*, ils permettent un survol plusieurs fois par jour d'une même région. Ils ont la particularité d'émettre leur signal enregistré à bord en continu. De fait, lorsque la zone d'émission d'un satellite intersecte le cercle de réception d'une antenne d'acquisition, une image correspondant à la trace au sol du satellite est enregistrée par les infrastructures en pied d'antenne. Dans des conditions optimales, il est possible d'acquérir plus de 10 images NOAA quotidiennement (tous satellites confondus). Les infrastructures d'acquisition du réseau

15. NOAA KLM USER'S GUIDE : <http://www2.ncdc.noaa.gov/docs/klm/>
 NOAA Satellite Information System : <http://noaasis.noaa.gov/NOAASIS/>

SEASnet ont permis la réception des données de la série de satellites météorologiques américains NOAA : de 1991 avec NOAA-12, jusqu'à maintenant avec NOAA-15, 18 et 19, les satellites actuellement encore exploitables.

Cette multiplicité d'images quotidiennes est intéressante dans un contexte inter-tropical à fort couvert nuageux car des processus de synthèse de données peuvent être envisagés. C'est cette chaîne de traitements précisément qui est utilisée en guise d'exemple dans ce qui suit et qui pourrait donner lieu, dans le cadre d'un observatoire virtuel tel que nous l'envisageons, à diverses adaptations ou expérimentations.

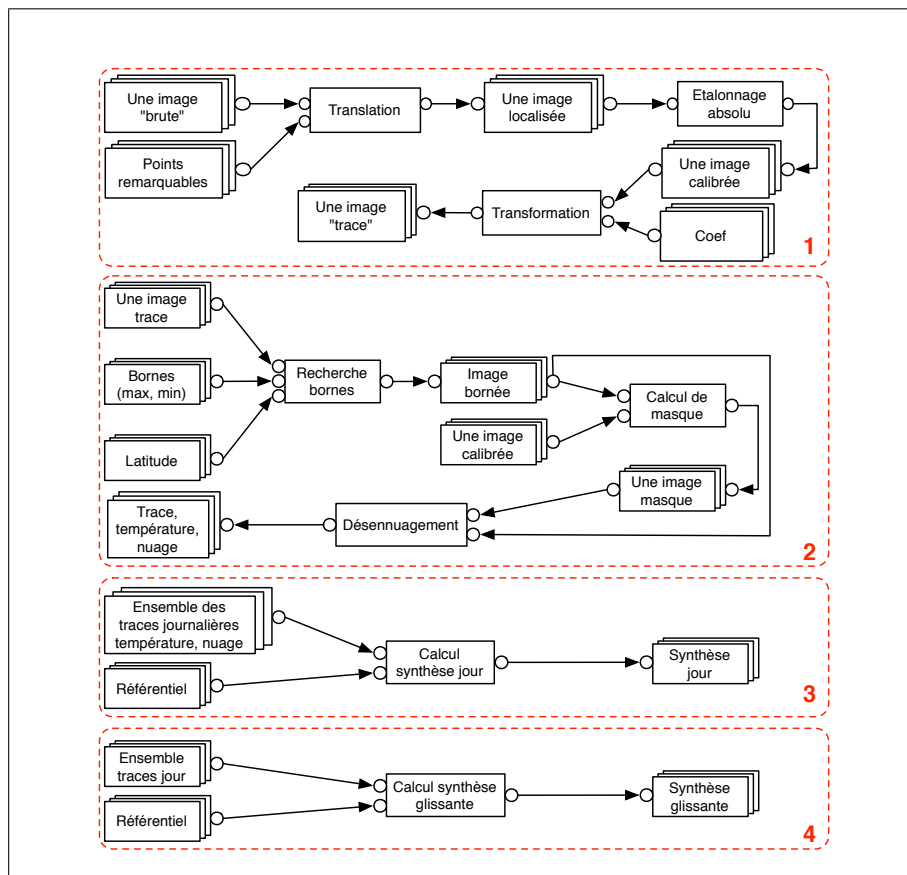


Figure 10 – Chaînes de traitements abstraites

La description complète de l'ensemble des traitements des quatre blocs (fig.10) est la suivante (seul le bloc 1 sera détaillé aux deux niveaux abstrait et concret conformément à la figure 11¹⁶).

16. Certains maillons de la chaîne peuvent être court-circuités comme le suggère la partie ombrée de la figure.

– Bloc 1 : Avant tout, il faut préciser que la localisation géographique des données NOAA n'est pas précise car il n'y a pas de GPS à bord du satellite pour affiner le positionnement orbitographique des images. Ces dernières sont donc générées à partir du modèle orbitographique prévisionnel utilisé pour le suivi du satellite par l'antenne d'acquisition. Une image NOAA est constituée d'un entête contenant toutes les informations relatives aux conditions d'acquisitions et aux paramètres de calibration du capteur. Viennent ensuite l'ensemble des enregistrements associés à chaque ligne de l'image. Pour chaque ligne, une information de localisation des pixels qui la compose est disponible en début d'enregistrement. Cette donnée n'est fournie que pour un sous échantillonnage de pixels (tous les 40 points). Une première lecture complète de l'image permet de récupérer ces informations afin de constituer une grille de localisation des pixels. La localisation de cette grille peut être ajustée par simple translation. Une interface utilisateur a été développée afin de visualiser l'image et d'afficher suivant cette grille de localisation la position du trait de côte. L'ajustement de ce dernier sur l'image, à condition bien évidemment qu'il soit visible dans l'image, permet d'ajuster le positionnement global de l'image. La correction utilisée pour ajuster le trait de côte est répercutée à l'ensemble des points de la grille de localisation et sauvegardée dans le fichier. Nous modélisons cette étape concrète de *localisation* de l'image par le traitement abstrait *Translation*. Le traitement prend en entrée une image NOAA au format N1B et un ou plusieurs points de localisation. Nous obtenons en sortie une image au format N1B dont le géoréférencement a été ajusté.

Le calcul d'une valeur de température sur un pixel donné de l'image est réalisé selon une procédure standard bien documentée dans la littérature : la "Multi Channel Sea Surface Temperature" (MCSST). Il s'agit d'appliquer un algorithme de *split window* prenant en compte les températures de brillance des canaux thermiques 4 et 5 des images NOAA ainsi que l'angle de prise de vue du capteur par rapport au sol [MCC 85]. Pour pouvoir travailler avec des températures de brillance des canaux 4 et 5, il est nécessaire de passer par une étape de calibration des comptes numériques de l'image (les niveaux de gris). Cette étape de traitement, notée *Etalonnage absolu* au niveau abstrait (bloc 1 de la fig.10 et fig.11) prend au niveau concret (bas de la fig.11) une image N1B et donne en sortie une image dans un format RAW. Le calcul de *MCSST* sur l'image calibrée avec un jeu de coefficients donne une image de température en sortie au format PNG (l'image "trace").

– Bloc 2 : Dans le contexte intertropical de notre réseau de stations de réception, la couverture nuageuse importante est souvent problématique pour les applications de cartographie. Dans les algorithmes de calcul de cartes de température de l'eau, les nuages parasitent les résultats car ils apparaissent beaucoup plus froids que la mer. Le calcul d'un masque des nuages est nécessaire si l'on veut pouvoir réaliser des synthèses entre différentes images. L'étape 1 est modélisée par le traitement *Recherche de bornes* dans l'image (fig.10). Les entrées sont : l'image de température et les paramètres permettant de filtrer ces valeurs en fonction de la latitude. On obtient en sortie une image bornée. L'étape 2 est modélisée par le traitement *Calcul de masque*. Il prend en entrée une image de température et une image NOAA calibrée (afin de pouvoir accéder aux températures de brillance). On obtient en sortie une image de

température et un masque que l'on croise avec le traitement désenuagement. En sortie de ce bloc on obtient des images de températures dans lesquels ont été masqués les pixels correspondant à des zones nuageuses.

– Bloc 3 : Nous l'avons vu précédemment, les satellites NOAA survolent plusieurs fois par jour la même région. Théoriquement 2 passages pendant la journée et 2 passages pendant la nuit. Durant la journée, l'ensoleillement provoque un réchauffement de la surface de la mer. Ainsi, pour calculer à partir des images, des valeurs de température plus proches de la réalité, il conviendrait de ne considérer que les images acquises durant la nuit et donc ne pas exploiter 50% des données acquises en pied d'antenne. Pour utiliser la totalité des images, nous avons mis au point une procédure de correction des images acquises durant la journée en prenant comme référence les températures calculées sur les images de nuit. Le croisement de toutes ces données réalise une carte quotidienne des températures : les images sont projetées dans un référentiel cartographique et la valeur de température effectivement retenue est la valeur maximale sur l'ensemble des images de la journée.

– Bloc 4 : Au terme de la synthèse décrite ci-dessus, calculée quotidiennement (au jour J), la couverture nuageuse reste importante. Pour réduire son étendue, un second processus de synthèse est envisagé et prend en compte 5 jours de données. Cette étape est modélisée par le traitement *Calcul synthèse glissante* (synthèse des 5 images quotidiennes en se centrant sur le jour J et en recherchant une valeur désenuagée à J, J-1, J+1, J-2 ou J+2 successivement).

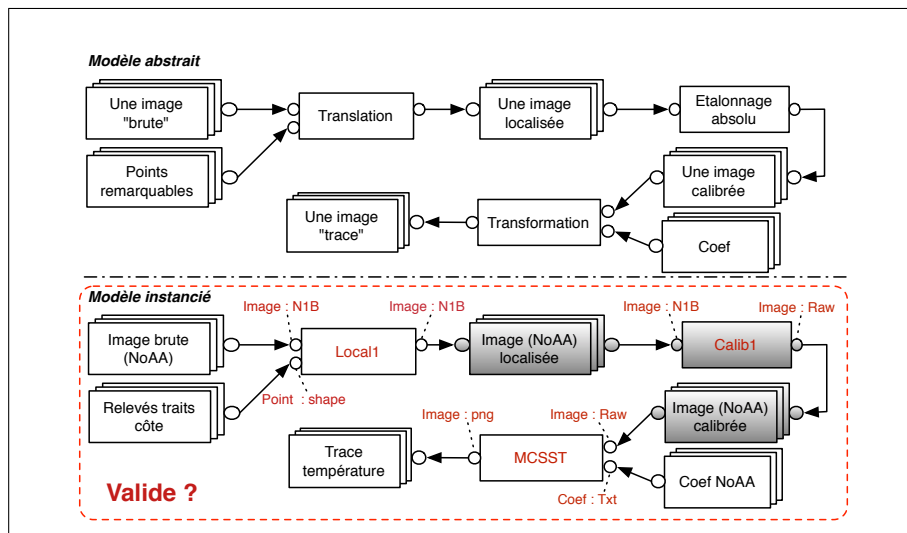


Figure 11 – Instanciation

Après la description des traitements, se pose la question de la validité de cette chaîne concrète (bas de la fig.11). Pour prendre en compte celle-ci, nous transcrivons la chaîne concrète formellement comme une suite de compositions de signatures de traitements. Une signature sera exprimée par :

Nom (liste de Param_E) : (liste de Param_S), avec

- **Nom** correspond à la désignation du traitement ou de l'opération concrète.
- **Param_E, Param_S** correspondent respectivement à un paramètre d'entrée et à un paramètre de sortie. Chaque paramètre **p** est formalisé (**dc :fo**), avec **dc** et **fo** relatifs aux catégories et formats de données utilisés.

Une composition **p1** → **p2** est le lien reliant un paramètre **p1** de sortie d'un traitement **T1** à un paramètre **p2** d'entrée d'un traitement **T2**.

Signature et composition nous permettent d'introduire la notion de compatibilité, sur laquelle reposera la validation de la chaîne concrète. L'objectif est d'assurer l'exécution sans faute de la chaîne concrète.

5. Conclusion et perspectives

La mise en place d'observatoire virtuel reposant sur des infrastructures distribuées devient, à nos yeux, incontournable. De nombreux domaines scientifiques, adoptent déjà cette démarche (astronomie, biologie, etc.). Les travaux déjà effectués, tant sur le plan de la réalisation des infrastructures, de l'accessibilité aux données et aux traitements (via des protocoles standardisés) que de la réflexion menée autour du contexte de travail, nous semble autant de maillons nécessaires à cette mise en place.

L'architecture workflow proposée est simple mais nous semble judicieuse car elle a pu être confrontée au domaine des images satellitaires mais aussi à d'autres domaines scientifiques (hydrologie, bioinformatique, etc. [LIN 11]). Les recherches ultérieures envisagées portent sur la validation des chaînes concrètes, c'est-à-dire sur les définitions relatives aux divers cas de *compatibilités*. L'étape ultime consistera à exécuter les chaînes concrètes valides au sein de l'observatoire virtuel (prises en compte de la virtualité et de la distribution).

6. Bibliographie

- [AND 03] ANDREWS T., CURBERA F., DHOLAKIA H., GOLAND Y., KLEIN J., LEYMAN F., LIU K., ROLLER D., SMITH D., THATTE S., TRICKOVIC I., WEERAWARANA S., Business Process Execution Language for Web Services, Version 1.1 , , 5 May 2003.
- [BRO 85] BROWN O., BROWN J., EVANS R., Calibration of Advanced Very High Resolution Radiometer Infrared Observations , *Journal of Geophysical Research*, , 1985.
- [CAL 01] CALOZ R., COLLET C., *Précis de Télédétection, Traitements numériques d'images de télédétection*, vol. 3, Presses de l'Université de Québec, Agence Universitaire de la Francophonie, 2001.
- [CUN 93] CUNNINGHAM S., DENIZE P., A Tool for Model Generation and Knowledge Acquisition , *Proc International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Florida, USA, 1993, p. 213-222.

- [HUL 06] HULL D., WOLSTENCROFT K., STEVENS R., GOBLE C., POCOCK M., LI P., OINN T., Taverna : a tool for building and running workflows of services , , 2006.
- [KAY 07] KAY M., XSL Transformations (XSLT) Version 2.0 , , 23 January 2007.
- [KON 94] KONSTANTINIDES K., RASURE J., The Khoros software development environment for image and signal processing , *Image Processing, IEEE Transactions on*, vol. 3, n° 3, 1994.
- [LIB 10] LIBOUREL T., LIN Y., MOUGENOT I., PIERKOT C., DESCONNETS J.-C., A Platform Dedicated to Share and Mutualize Environmental Applications , *ICEIS (1)*, 2010, p. 50-57.
- [LIN 09] LIN Y., LIBOUREL T., MOUGENOT I., A Workflow Language for the Experimental Sciences , *ICEIS (3)*, 2009, p. 372-375.
- [LIN 11] LIN Y., Méthodologie et composants pour la mise en œuvre de workflows scientifiques , PhD thesis, Université Montpellier II. École doctorale I2S, 2011.
- [LUD 06] LUDÄSCHER B., ALTINTAS I., BERKLEY C., HIGGINS D., JAEGER E., JONES M. B., LEE E. A., TAO J., ZHAO Y., Scientific workflow management and the Kepler system , *Concurrency and Computation : Practice and Experience*, vol. 18, n° 10, 2006, p. 1039-1065.
- [MCC 85] McCLAIN E., PICHEL W., WALTON C., Comparative performance of AVHRR-based multichannel sea surface temperatures , *Journal of Geophysical Research*, , 1985.
- [OMG 06] OMG, *Meta Object Facility (MOF) Core Specification, Version 2.0, formal/06-01-01*, January 2006.
- [PET 02] PETIT M., RAMOS A., From receiving stations to concept : Survey of the Environment Assisted by Satellite (SEASnet), a net for research and development , *Proceedings of ESA Workshop GODAE GHRSS, ESRIN, 02-04, 2002*.