

Livrable 2.4.2 - 27 mai 2016

## Modélisation, collecte et visualisations des traces

**Résumé :** Dans le cadre de cette première phase du projet Epistémè nous nous intéresserons à l'activité du segment sol de la mission Herschel dans le cadre du programme clé Gould Belt qui étudie la formation des étoiles dans l'univers froid. Dans le présent rapport, nous faisons état des sources potentielles de traces que nous avons repérées. Puis nous abordons le cycle de l'information (symbolisé pour une « image » de l'univers) tel que nous le percevons actuellement, a minima depuis sa naissance au format numérique à son image.

## Introduction

Mardi 3 Mai 2016, le LINA a accueilli dans le cadre du projet Epistémè un ingénieur de recherche, Florian Melki, afin de remplacer le post-doctorant IHM / Visualisation attendu initialement pour le second semestre de l'année 2015. Sa première mission a été d'intégrer le contexte du projet Herschel selon trois axes :

- Initiation au métier de l'astrophysique : loi de Planck, corps noir, photométrie, univers froid, etc, tout le bagage nécessaire à la bonne appréhension de la suite du projet ;
- Compréhension des grandes étapes de la mission Herschel : lancement du satellite embarquant les trois appareils HIFI, PACS et SPIRE pour photographier l'univers dans différentes longueurs d'ondes, puis analyses des données enregistrées. En amont et cours de la mission, des éléments d'organisation (*Key Program, Guaranteed Time / Open Time, préparation des Astronomical Observation Requests...*) et des outils techniques (HSPOT, HIPE, IDL, *getsources/getfilaments...*) s'articulent autour du segment sol ;
- Focalisation sur l'évolution de la mission au cours du temps: depuis 2009 jusqu'à nos jours, les pratiques autour du traitement de l'information ont évolué, et différentes techniques d'étalonnage et d'analyse se sont succédées. Qu'est-ce qui a changé et comment en rendre compte ?

L'objectif de ce travail est de produire des visualisations interactives permettant de rendre compte des pratiques des astrophysiciens et des changements associés en les rendant accessibles à différents publics (épistémologues, astrophysiciens, grand public).

Une réunion s'est tenue au CEA Irfu le mardi 17 Mai dans un objectif double :

- Partager la vision des membres de projet Episteme avec deux astrophysiciens : Philippe André, responsable scientifique du programme « Formation des étoiles », et Vera Könyges, astrophysicienne.
- Entamer une discussion sur les différents processus entrés en compte lors du projet Herschel :
  - Au niveau macro, depuis la préparation du satellite Herschel en 1982, la conception et fabrication des appareils embarqués HIF, PACS et SPIRE entre 1995 – 1997, jusqu'à nos jours
  - Au niveau micro, concernant la mise en place et l'évolution des pratiques ayant trait au traitement de l'image

## Des traces d'activités

Le segment sol est une « organisation humaine et matérielle<sup>1</sup> ». De ce fait, son activité se tourne à la fois sur l'objet manipulé, à savoir l'information récupérée par le satellite, que sur la communication entre les membres de la mission.

Ainsi, afin de rendre compte de ces deux aspects, nous viserons à modéliser et à collecter différents types de traces (Champin et al. 2013)<sup>2</sup> qui seront autant de marqueurs ou d'indices de l'activité des scientifiques et du cycle de l'image :

- traces d'outils spécifiques tels que HIPE (outil d'étalonnage propriétaire du CEA, lequel génère des fichiers *Flexible Image Transport System* ou FITS), outils d'analyse développés au CEA (l'IDL Scanamorphos; les logiciels d'extraction de sources *getsources* et *getfilaments*), télescope virtuel en ligne (CDS de Strasbourg dont Vizier, base de données de catalogues de sources, et Simbad, base de données d'objets), outils de recherche de ressources, etc.
  - Il est à noter que les fichiers FITS contiennent de nombreuses meta-données porteuses d'informations. Nous pourrions donc étudier aussi bien la vie du fichier FITS (création, modification, suppression) qu'exploiter son contenu
- traces d'outils liés au « travail de la connaissance » (*knowledge work*) comme le wiki utilisé par le CEA (eDoc), les archives du projet Herschel (HSA), les organisations de connaissances et les classifications (e.g. structures de dossiers plus ou moins partagées où sont stockées les FITS, mots-clés associés aux métadonnées), les échanges de fichiers par mail ou espaces partagés, etc.
- Traces via des communications extérieures : web, communications officielles, publications, etc.

A court terme et suite à la réunion au CEA du 17 Mai, nous avons choisi d'explorer en premier lieu le travail effectué sur l'image, dans l'objectif de produire une visualisation pour le 20 septembre 2016, date du colloque Episteme à Nantes dans le cadre de la Digital Week 2016. A priori ce travail s'effectuera en traçant la génération des différentes versions d'un fichier FITS au cours de ses transformations successives. Cependant disposer de davantage de traces contextuelles au projet permet d'anticiper des besoins futurs, c'est-à-dire lorsque nous voudrions enrichir nos premières visualisations.

---

<sup>1</sup> Rapport Episteme L2.2.2

<sup>2</sup> Champin, P.-A., Mille, A., and Prié, Y. Vers des traces numériques comme objets informatiques de premier niveau : une approche par les traces modélisées. *Intellectica*, 59, 2013.

## La transduction de l'image dans le cadre du nuage Aquila

Nous choisissons de suivre la transduction d'une image via le suivi d'un fichier FITS et des traces produites lors de ses transformations successives. Ces transformations s'inscrivent dans le cadre d'un processus maîtrisable et a priori reproductible. Nous partons donc du postulat de cette reproductibilité pour viser la reproduction du scénario complet de transformations d'un fichier FITS. De ce fait nous serons capable de valider le processus et de collecter les traces dont nous avons besoin (logs, différentes versions du fichier FITS, etc).

Au sortir de cette phase, nous devons être capables de répondre aux questions suivantes :

- En quoi consistent exactement ces étapes d'étalonnage et d'analyse ? Quels logiciels rentrent en compte, et comment sont-ils utilisés (manuellement / avec un paramétrage automatique) ?
- Comment expliquer chaque étape dans un langage compréhensible par tous ?
- Comment ces étapes ont-elles été mises au point au cours du temps ? Il est certain que les pratiques d'aujourd'hui ne sont pas les pratiques d'hier. Qu'y avait-il en 2009, et comment en est-on arrivé aux processus actuels de travail de l'information ?

C'est au travers de l'étude qui a porté sur le nuage Aquila que nous comptons apporter des réponses. En effet, deux articles officiels ont été publiés en 2010<sup>3</sup> et 2015<sup>4</sup> par l'astrophysicien Vera Könyves qui travaille actuellement pour le CEA. Grâce à ceux-ci ainsi qu'à notre entrevue du 17 Mai, nous disposons de trois visions :

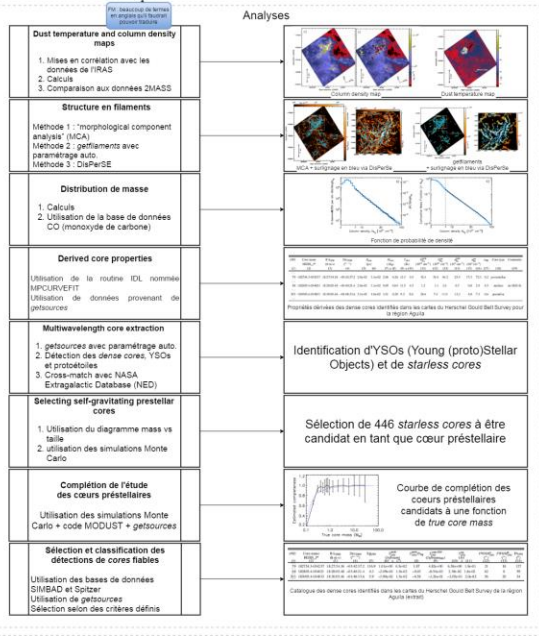
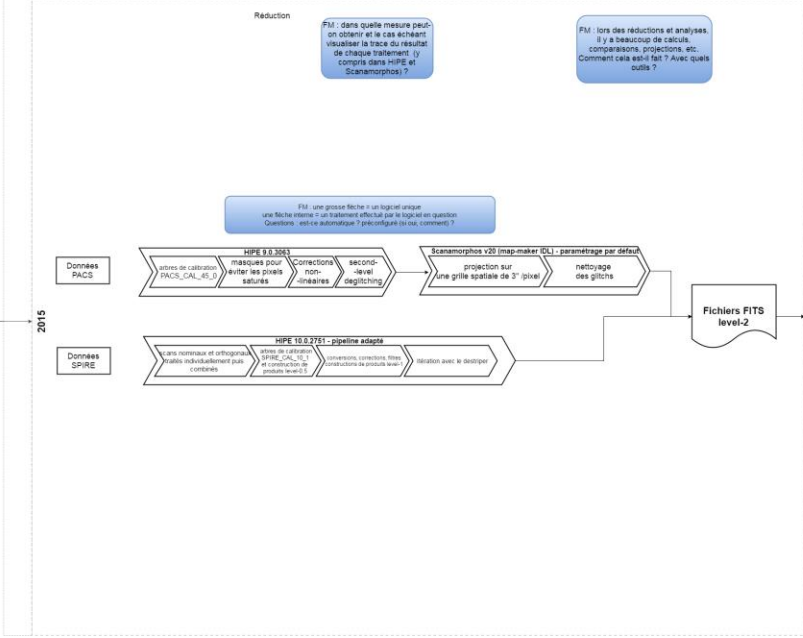
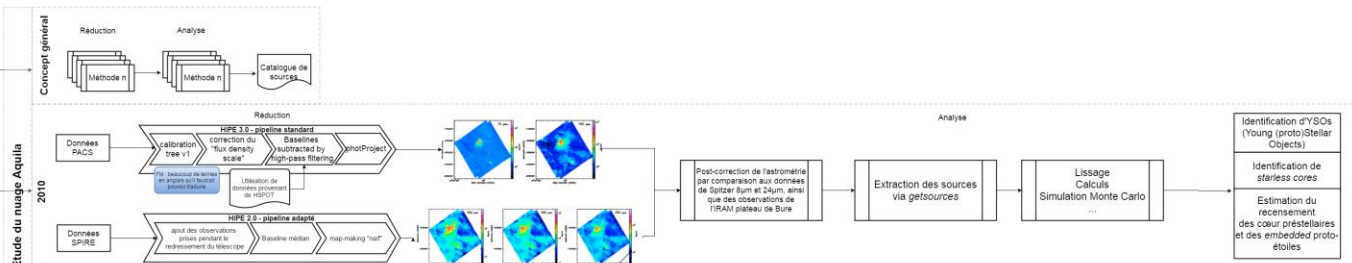
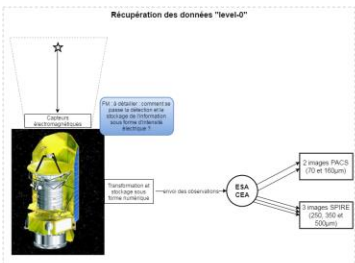
- En 2009, il n'y avait aucun processus établi, seulement des propositions de méthodes d'étalonnage et d'analyse provenant des différents groupes de scientifique ;
- En 2010, un premier processus est établi et explicité au travers de l'article écrit par Vera Könyves ;
- En 2015, un second processus est proposé au travers du second article de Vera Könyves.

Le détail des différents processus mis en œuvre se trouve ci-dessous.

---

<sup>3</sup> <http://www.aanda.org/articles/aa/pdf/2010/10/aa14689-10.pdf>

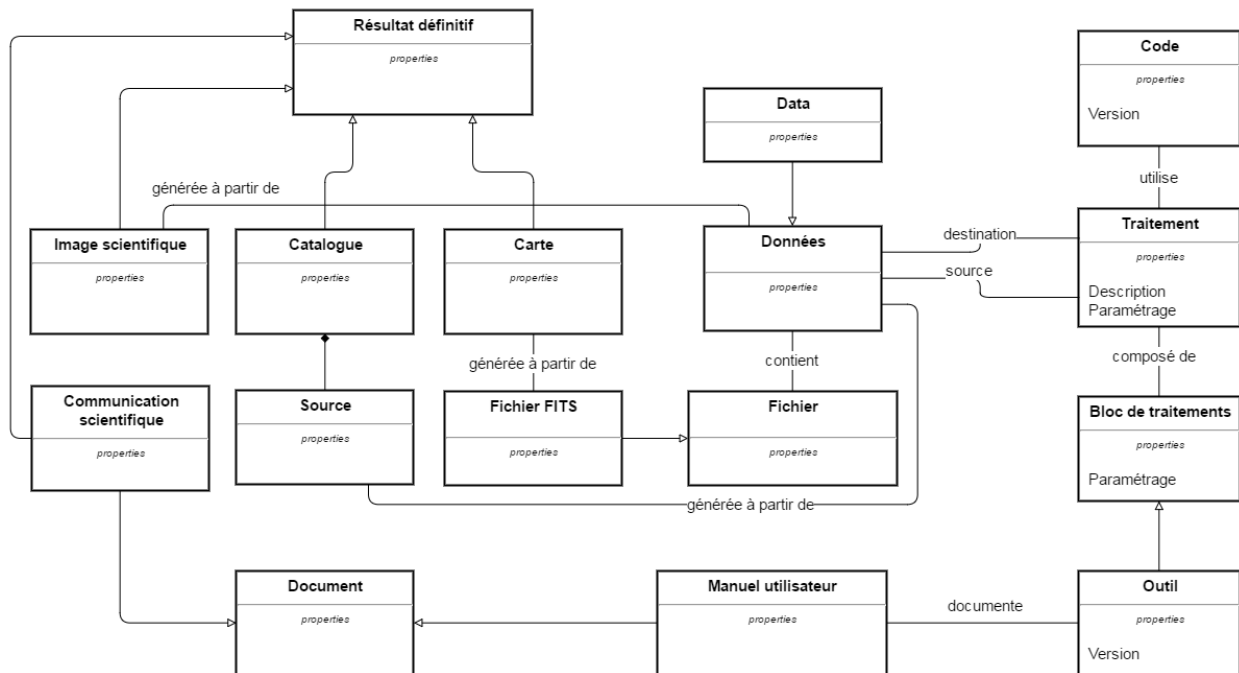
<sup>4</sup> <http://www.aanda.org/articles/aa/pdf/2015/12/aa25861-15.pdf>



## Poursuite du travail

Nous allons continuer l'analyse de l'évolution des pratiques dans le cadre de l'étude du nuage Aquila. Florian Melki se rendra à courte échéance au CEA afin de travailler avec Vera Könyves. Ensemble, ils s'attacheront à reproduire les mêmes étapes qu'en 2010 et 2015 afin de valider ou compléter le modèle, récupérer toutes les traces qui seront produites au sortir des différentes étapes tout en évaluant la possibilité de récupérer les traces des différents processus qui ont été mis en œuvre entre 2009 et 2016, et produire des descriptifs compréhensibles pour les éléments impliqués dans chacune des étapes, toujours dans un objectif de médiation.

Nous avons établi un modèle préliminaire de ce que nous percevons être l'instrument numérique utilisé par les scientifiques (ci-dessous). Celui-ci devra également être complété durant le travail avec Vera Könyves.



Une fois ce travail effectué commencera celui de la construction itérative de visualisations interactives de ces informations, en vue d'une présentation publique en septembre 2016.