

EUCLID 2015

APC - Université Paris Diderot, Paris



SGS System Team

C. Dabin
EC SGS ST Coordinator
On behalf System Team members

Agenda

1. Activités System Team
2. Mission Data Base
3. Challenges
4. Data Access Policy

- Plate forme d'intégration continue (CODEEN), gestion des tickets utilisateurs, librairies externes : analyse, CCB, gestion en configuration, compatibilité des licences (LGPL), chaîne de production des exécutables, génération des packages
- EDEN : référentiel de l'environnement de développement : libs, compilateur, tools, chaîne de build
- LODEEN : Version locale de EDEN embarquée dans une VM
- Animation et diffusion de la Newsletter Euclid Software Developers, point de contact unique : commontools.support@lists.euclid-ec.org
- Préparation des workshop développeurs et data model : de 20 à 80 participants Euclid (OU / SDC / ST) : thèmes, appels à contributions, feedback

- Animation de 10 groupes de travail thématiques sur les Common Tools
- Définition, négociation avec les développeurs et implémentation des métriques qualité logicielle sur CODEEN
- Support aux équipes VIS/LE3 pour la mise en place des infrastructures de développement : SDC DEV
- Développement, maintenance des outils transverses autour du Data Model : bindings (C++/Python), wrappers (C++/Python), dictionnaire de types, support à l'utilisation des outils
- Définition et implémentation des interfaces avec les autres composants de l'architecture : Archive / Infrastructure Abstraction Layer



Difficultés à faire en sorte que les équipes de développement 'jouent le jeu'



- Favoriser et communiquer sur les initiatives 'locales' candidates à être partagées : Librairies scientifiques images, catalogues, spectres, ...
- Relecture croisée (et critique) et contribution à l'écriture de modèles de données communs, mise en cohérence et ... simplifications
- Ingénierie sur les techniques de déploiement CernVM-FS et/ou Dockers et virtualisation CernVM : pour faciliter et automatiser le déploiement des codes de processing sur les SDC's Prod
- Développement d'un outil d'optimisation de la distribution des observations sur les différents SDC's qui minimise les transferts de fichiers entre SDC's en :
 - Favorisant la co localisation des observations ayant un recouvrement spatial
 - Assurant la redondance (SDC primary / SDC secondary)
 - Respectant les investissements en infrastructures de chaque SDC Prod : SDC Prod XXL, XL, M, S

- Archive et composants transverses
 - Les jeux sont faits sur la techno Archive : RuG/ESAC ... Oracle, Astrowise
 - Vigilance à travers :
 - Spécifications sur la flexibilité (évolutions du data model), spécifications sur les performances : ingestion / query / retrieve
 - Revue de code, visibilité exigée sur le code et les choix d'implémentation, challenger la solution techno choisie avec des solutions 'plus modernes' , évaluer la scalabilité des solutions
 - Stress tests ... très tôt pour soulever les pbs de design au plus tôt
- ! Mettre les utilisateurs de l'Archive en interaction avec les services implémentés par RuG/ESAC !

Architecture : HPC versus HTC ?

- Le modèle de référence de l'infrastructure de production cible est de type : HTC (High Throughput Computing) qui convient à la majeure partie des PF's sauf ...aux 'gros' consommateurs

- 'petits' jobs : hour / day
- IO oriented
- 'petite' consommation mémoire
- 'faible' parallélisation

- INCOMPATIBLE de certains besoins LE3

- 'gros' jobs : week / month
- CPU oriented
- 'grosse' consommation mémoire
- 'forte' parallélisation



Architecture : LE3 et les autres ?

 Actions en cours pour 'optimiser' les perfos ... et les besoins en amont avec les SWG's, la tendance est à la confirmation des estimations SRR

- Solliciter les SDC's leaders pour couvrir aussi les besoins 'énormes' en HPC's et anticiper
- Positionnement SDC Fr ?
- Activisme SDC Finlande mais peu d'enthousiasme de la part des autres SDC's

 Le consortium est attendu sur ce sujet (cf outcomes de la SRR)

- Les activités SPR (cf Jérôme) peuvent impacter cette problématique de dispo d'une infra spécifique

- Services et contenus ici :
 - <http://euclid.roe.ac.uk/projects/missiondatabase>
 - Et browser ici :
<http://euclid.esac.esa.int/epdb/>



Processus de mise à jour et d'utilisation par les Développeurs de code Euclid pas satisfaisant

Autorité 'scientifique' sur le contenu : depuis une demande émanant des développeurs de PF's jusqu'à la mise à jour et publicité sur la disponibilité

Quid du cas d'utilisation des challenges : quelle PSF, quels modèles : AsRequired, AsDesigned, CBE ?

- Challenges 'purement' IT (Information Technology) mettant en œuvre les composants transverses avec des versions de pipeline 'anciennes'
 - Développement incrémental de ces composants
 - Compléments fonctionnels, feedback sur les modes opératoires et les défauts de design
 - Validation de l'intégration technique de ces composants : IAL/EAS/DSS/M&C/COORS
 - Validation fonctionnelle et stress tests aux limites
 - Testbed pour les technos de déploiement / virtualisation / archivage

Challenges scientifiques (1)

- Challenges mettant en œuvre les Processing Functions sur l'infrastructure 'cible' des SDC Prod

-  Difficulté de cerner les objectifs scientifiques
-  Difficulté de rendre compatibles les objectifs 'croisés' et le calendrier du challenge (glissements)

Quels effets instrumentaux ou modèles inclure dans la simulation ?

Quelle maturité de caractérisation de ces effets ?

Quelle correction envisagée ou 'attendre' ?

- Davantage orientés intégration technique

Challenges scientifiques (2)

- Adjoindre une responsabilité scientifique au chef de projet Challenge (OU MER leader ?)
- Anticiper le challenge MER à venir (2017) qui est (très, trop ?) ambitieux
 - EXT's, SIM, VIS, NIR, SIR, MER
 - Quelles PSFs ?
 - Quels EXT's ? KiDS, DES, PanSTARS, LSST ?
 - Quelles fonctionnalités dans MER ?
 - Quels requirements scientifiques ?
 - Quels produits : VIS/NIR/SIR/SIM/MER avec quelle qualité et quelle complétude pour MER ?



- Who can use what and when and under what conditions ?
- Who : Euclidians Users (EC member) and non Euclidian
- Who : Euclidians Groups : SWG, OU/SDC, IOT/IDTs
- Conditions : Quota / Restrictions (take care of Distributed environment) / Quality Of Services (on line, near on line, off line) / Security

Data Product Access Policy : What ?

- What user services ?
 - Processing software development, build and deployment on SDC's
 - Metadata and data browsing : query/retrieve
 - Ordering : running piece of processing software on (any ?) SDC's, stop/retry/cancel
- What data products ?
 - Data Products categorization
 - LE1/VIS/SIM/EXT/NIR/SIR/MER/SHE/PHZ/LE3s through metadata access (DataBase) and file access (File System)



Consortium Products vs Legacy products (Data release)

- What data products should be public ?
 - e.g : Calibration files / Spectra2D / Covariance matrices /
- Who is responsible for ?

Data Product Access Policy : When ?

- Does the Data Access Policy vary over time ?
- When :
 - Development : challenge, E2E
 - System tests : MPR
 - Commissioning
 - Operations

- Quality of Services :
 - Any metadata/data shouldn't be accessible in Real Time !
 - Capacity / Quotas / Type (archive, disk, fast cache)
 - Requires prioritization and archiving strategy

Last DR ?

Last run ?



Purge policy : files deleted if not access since TBD months ?

The big matrix : user/services/data

<i>Users</i>	<i>Group / Profiles</i>	<i>Services / Quota</i>	<i>Data / Quota</i>	<i>Timeline</i>
Name/Surname	SWG	Developers	LE1 Calibration files	Development
	OU/SDC	Run	VIS	Validation
	IDT/IOT	Browse / Query/ Retrieve metadata	NIR	Commissioning
	Non EC member	Browse files / upload files Cutout files	SIR	Operations
			MER	Analysis
			Catalog (s)	

To be managed under configuration at EC level