

Enquête Outils et logiciels : quels  
besoins ? quelles activités et  
expertises ?

# Pourquoi cette enquête ?

- Identifier les besoins informatiques en outils, services et logiciels nécessaires aux collaborations portées par l'institut.
  - En déduire les besoins et les items d'importants.
- Identifier nos engagements dans ces éléments, à la fois sur **les aspects opérationnels ( exploiter et utiliser) et sur les aspects développement.**
  - En déduire “notre impact” dans l'exploitation et l'évolution de ces outils/services/logiciel.
- Tout ça afin de permettre à l'institut de prendre des actions : priorisation/support/engagement/désengagement/...

# Construction de l'enquête?

- Le but n'était pas de faire la liste exhaustive des besoins des expériences.
- Le but n'était pas de faire la liste exhaustive des services à exploiter au sein de nos infrastructures.
- Le but n'était pas de lister l'ensemble des développements logiciels faits au sein de nos laboratoires.
- On se focalise sur un ensemble de composants identifiés.
  - Pas spécifiques à une collaboration.
  - Pas identifié à un besoin ponctuel.

# Éléments sur lesquels porte l'enquête

- VOMS (AAI)
- IAM
- Dcache ( storage)
- Xrootd
- EOS
- Irods
- HPSS
- POSIX File Systeme
- RUCIO
- DIRAC ( data management)
- AMI
- Local solution of metacatalogue
- FTS ( File Transfert Service)
- Globus ( data transfert)

- Slurm
- DIRAC ( job distribution)
- Panda
- GlideInWMS
- Geant
- Geant4 DNA GATE
- Pythia
- Corsika
- ROOT
- Fink
- HPC usage infra, GPU
- FPGA Usage infra
- Reconstruction + tps réel (ACTS, Allen, fitQun..)

- Heterogeneous Computing :  
Alpaka, kokkos, Allen core,  
Cuda
- AI tools :  
Pytorch/tensorflow/ONNX/JAX
- MLOps: MLflow, etc
- ML on FPGA: HLS4ML, VHDL,  
etc
- C++/Python
- Rust/Julia
- Gitlab/ GitlabCI
- Apptainer Docker
- Kubernetes

Services

Logiciels

Compétences  
/Usage

# Retour des collaborations

- On a simplement demandé aux collaborations ( composante française) l'importance pour elles de chacun des item et si elles participaient au développement et/ou aspect opérationnels.
  
- A partir des réponses et de divers échanges on a construit pour chaque item :
  - Une échelle de popularité au sein de l'institut.
  - Une échelle de l'intérêt de l'institut dans le développement
  - Une échelle de l'intérêt pour IN2P3 dans l'usage et le fait d'opérer.

# Retour des collaborations

Software Tools Services	Popularity	IN2P3 interest Developments	IN2P3 interest Operation/usage
VOMS (AAI)	Green		
IAM	Light Green	Orange	Red
Dcache ( storage)	Green		Red
Xrood	Green		Red
Irods	Light Green		Light Orange
HPSS	Green		Red
POSIX File Systeme	Light Green		Light Orange
RUCIO	Green	Orange	Red
DIRAC ( data management)	Light Green	Red	Light Orange
AMI	Light Green	Orange	Red
Local solution of metacatalogue	Light Green		
FTS ( File Transfert Service)	Green		Red
Globus ( data transfert)	Light Green		

Slurm	Green		Light Orange
DIRAC ( job distribution)	Green		Red
Panda	Green		
GlideInWMS	Light Green		
HPC usage infra, GPU	Green		Red
FPGA Usage infra	Green	Light Orange	Light Orange
Geant	Green	Light Orange	Light Orange
Geant4 DNA GATE	Green	Red	Light Orange
Pythia	Light Green	Light Orange	Light Orange
Corsika	Light Green	Light Orange	Light Orange
reconstruction + tps réel (ACTS, Allen, fitQun..)	Green	Red	Red
Heterogeneous Computing : Alpaka, kokkos, Allen core, Cuda	Green	Red	Red
Fink	Green	Light Orange	Red
AI tools : Pytorch/tensorflow/ONNX/JAX	Green	Light Orange	Red
MLops: MLflow, etc	Light Green	Light Orange	Light Orange
ML on FPGA: HLS4ML, VHDL, etc	Light Green	Light Orange	Light Orange

ROOT	Green	Light Orange	Red
C++/Python	Green	Light Orange	Red
Gitlab/ GitlabCI	Green		Red
Aptainer Docker	Green		Red
Kubernetes	Light Green	Light Orange	Light Orange

# Retour des collaborations

- Les retours étant assez disparates, pas basés sur des métriques. Il y a donc une part d'interprétation/contextualisation dans les codes couleurs.
- Quelques tendances claires tout de même :
  - Pas de forte corrélation entre popularité d'un besoin et implication dans son développement.
    - Des exceptions pour des thèmes d'actualités ( HPC usage, hétérogenous computing) ou pour l'historique DIRAC.
  - Un absence notoire d'engagement dans le développement d'outils populaires et présentant un fort intérêt pour l'institut.
  - Une bonne corrélation entre popularité et compétence/utilisation.

# Retour des laboratoires

- Dans un second temps on a demandé aux laboratoires de nous faire remonter leurs points de vue sur les même items:
  - Opèrent t'il les services ?
  - Déploient ils un infrastructure correspondante ?
  - Font ils des développement ?
  - Utilisent ils et ont ils des compétences dans l'usage ?
  - Dans la mesure du possible quantifier les ressources engagées.

L'enquête envoyée aux responsables informatiques MAIS avec l'ambition de collecter les informations de l'ensemble du laboratoire ( service info, équipes scientifiques, autres)

- Objectif : collecter au plus près, les ressources et moyens que l'institut mets dans chaque items.

# Retour des laboratoires

Software/Tools/Services	Intérêt pour IN2P3 à déployer et opérer	Déployé/opéré dans nos laboratoires	Utilisation/Usage et Compétences	Intérêt pour IN2P3 à développer	Développé dans nos laboratoires
VOMS (AAI)		IJCLAB	IJCLAB, plus de façon historique dans certains laboratoires grille		NON
IAM		CC, IJCLAB	Dans les laboratoires grille		NON uniquement des adaptations pour un usage sur les services locaux
Dcache ( storage)		CC, CPPM, LAPP,IPHC,LPCA	Collaborations de grille et un grand nombre de collaborations IN2P3		NON
Xrootd		CC, LAPP	Collaborations de grille et une grande majorité des collaborations IN2P3		NON
EOS		IJCLAB,LLR	Collaboration de grille		NON
irods		CC,IPHC,GANIL	Un certain nombre de collaboration IN2P3		NON
HPSS		CC	La quasi totalité des collaborations		NON sauf plugin local (treqs au cc)
POSIX File Systeme		Tous les laboratoires	Utilisation par les utilisateur locaux		NON . A noter un panel de solutions ( NFS,CEPH-FS,appliance) divers
RUCIO		LAPP ?, en cours à IJCLAB	Utilisé par un nombre significatif de collaborations		NON
DIRAC ( data management)		CC,CPPM,LUPM	Utilisé par un nombre significatif de collaborations/projets		OUI : CC et CPPM . Intérêt du LUPM
AMI		CC	Par un nombre limité de collaboration		Oui LPSC, Développement majeur
Local solution of metacatalogue		GANIL	GANIL		GANIL pour eurolabs
FTS ( File Transfert Service)		NON	Utilisé par un nombre significatif de collaborations/projets		NON
Globus ( data transfert)		NON	NON		NON

# Retour des laboratoires

Slurm		Déployé et opéré dans un nombre important de clusters de calcul de laboratoires	Oui dans la plus part des laboratoires. Des compétences importantes dans l'opération et le support d'usage de ce service	NON
DIRAC ( job distribution)		CC, CPPM, LUPM	Utilisé par un nombre significatif de collaborations/projets	OUI : CC et CPPM
Panda		NON	Utilisé par un nombre significatif de collaborations/projets	NON
GlideInWMS		NON	Utilisé par CMS	NON
HPC usage infra, GPU		La quasi totalité des laboratoires déploient et proposent des solutions GPU	Utilisé par divers groupes de physique	Not relevant
FPGA Usage infra		CPPM, LAPP, LLR et GANIL	CPPM, LAPP, LLR et GANIL	Des développements sont faits dans ces labos ( acquisition)
Geant		Pile logicielle installée dans la plupart des laboratoires ( cluster local)	Compétences présentent dans les laboratoire à la fois dans le déploiement du soft et dans son usage	IJCLAB
Geant4 DNA GATE		Proposé dans quelque laboratoires	Dans quelques laboratoires	NON
Pythia		Proposé dans quelque laboratoires	Inconnu	NON
Corsika		IP2I	?	Le LUPM participe au développement d'un sous ensemble avec une personne
reconstruction + tps réel (ACTS, Allen, fitQun..)		LAPP ?, Subatec, CPPM ? déploient cette technologie	Oui dans ces mêmes laboratoires	Un nombre important de laboratoire développent sur ces technologies, IJCLAB, IP2I, Subatech, LPHNE, LAPP et surement CPPM .
Heterogeneous Computing : Alpaka, kokkos, Allen core, Cuda		Des solutions de GPU sont proposées dans la grande majorité des laboratoires	Oui dans ces mêmes laboratoires	De nombreux laboratoires travaillent sur le développement autour des GPU (CC, L2IT, SUBATECH, CPPM, LPC Caen, LAPP, IJCLAB, LLR, LP2IB, IPHC)

# Retour des laboratoires

<b>Fink</b>		Déployé au CC et à IJCLAB pour des collaborations	Oui pour LSST	Développement essentiellement à IJClab avec un peu L2IT
<b>AI tools : Pytorch/tensorflow/ONNX/JAX</b>		Un nombre important de laboratoire proposent des outils de AI/ML ( APC,CC,CPPM,IP2I,LPCCaen, LLR, LP2IB,IPHC )	Ces laboratoires utilisent ces tools de AI/ML ( APC,CC,CPPM,IP2I,LPCCaen, LLR, LP2IB,IPHC ) à divers niveaux: Acquisition, analyses,....	Nombre importants de développements utilisent l'AI
<b>MLops: MLflow, etc</b>		Seul le LPC Caen propose cette pile logicielle à ces utilisateur	L'intérêt est pour l'instant très modeste dans la mis en œuvre de cette solution technologique L2IT ?	?
<b>ML on FPGA: HLS4ML, VHDL, etc</b>		Seul le CPPM,LLR,LP2IB et IPHC utilisent ( à travers quartus ?) cette approche	Oui dans ces même laboratoires	Des développements s'appuient sur ces technologies (acquisition)
<b>ROOT</b>		CC,Subatec,CPPM,IP2I,LPC Caen, LPNHE,GANL,IJCLAB,LLR,LP2IB,IPHC	Utilisé dans ces même laboratoires. Les compétences d'usage ?	NON
<b>C++/Python</b>		Disponible dans tous les laboratoires , travers parfois de "interface"jupiter,...	Utilisé a de nombreux niveaux, des compétences et connaissances importantes dans le savoir mettre en œuvre	?
<b>Rust/Julia</b>		Proposés par CC, IJCLAB,LAPP	?	?
<b>Gitlab/ GitlabCI</b>		Proposé par le CC, LPNHE. Certains laboratoires LPCCaen, LAPP proposent des runner complémentaires au gitlabCI du CC. Déploiement à venir au GANIL	Oui, utilisation importante à tous les niveaux, collaboration, services, projets,...	NON
<b>Apptainer Docker</b>		APC,CC,CPPM,IP2I,LPC Caen LPNHE et le LAPP,LLR et LP2IB,IPHC, IJCLAB propose des solution de containerisations	Utilisation importantes, Compétences ?	NON
<b>Kubernetes</b>		Le CC, IJCLAB et le LAPP en font une utilisation interne. APC et IPHC le propose au niveau collaboration/projet	Probablement	NON

# Retour des laboratoires

- Dans réponses très très disparates, d'où un besoin de resynthétiser et de clarifier certains retours.
- Il apparait que des retours sont pas complets au sens que nous n'avons pas eu une réponse représentative de l'ensemble des activités du laboratoire.
- Des outils/services qui n'ont pas été inclus dans l'enquête ont été signalés.
- Une seconde passe de cette enquête auprès des laboratoires va être faite afin de couvrir ces manques.