

Systeme d'information de l'OSU-Réunion : unification et centralisation des services

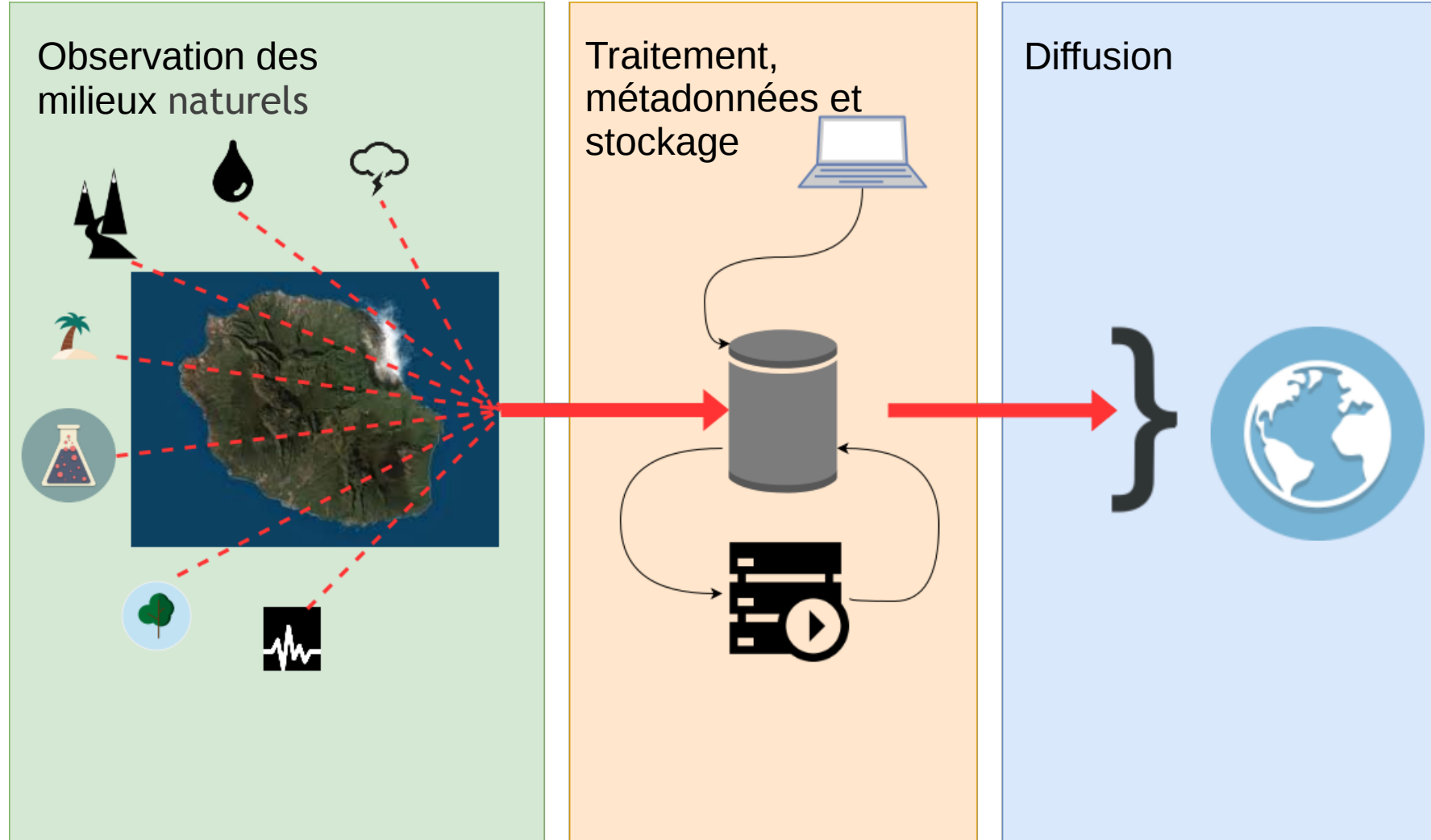
28-29 Juin 2018
SIST2018

Remy Decoupes



Introduction :

Travail du service informatique de l'OSU-Réunion



Problématique :

Thématiques scientifiques et traitements hétérogènes : Exemple

► SNO NDACC / IR ACTRIS



Acquisition permanente
des données



Nettoyage bruit



Envoi FTP à AERIS
~ < 30 jours

► IR OZCAR



Saisie manuelle des
données



Traitement complet



Format pivot de données
et métadonnées
~ en cours de discussion

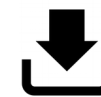
► SNO Dynalit / IR Ilico



Acquisition automatique
et manuelle



Traitement complet



hébergement des
données en local
~ < 1 an

Problématique :

Systeme d'information complexe

- ▶ Risques / problèmes
 - ▶ Avoir autant de SI que flux/pipeline différents
 - Les données et métadonnées éparpillées sur plusieurs serveurs
 - Des contrôles d'accès différents (base d'authentification locale)
 - ▶ Difficultés de maintenance
 - Détection anomalie longue : Aucune visualisation d'ensemble de la production de données
 - Correction bug compliquée
 - ▶ Les utilisateurs sont perdus :
 - ils ne savent pas sur quels serveurs aller
 - Ils perdent leurs mots de passe

Solution :

Unification et centralisation

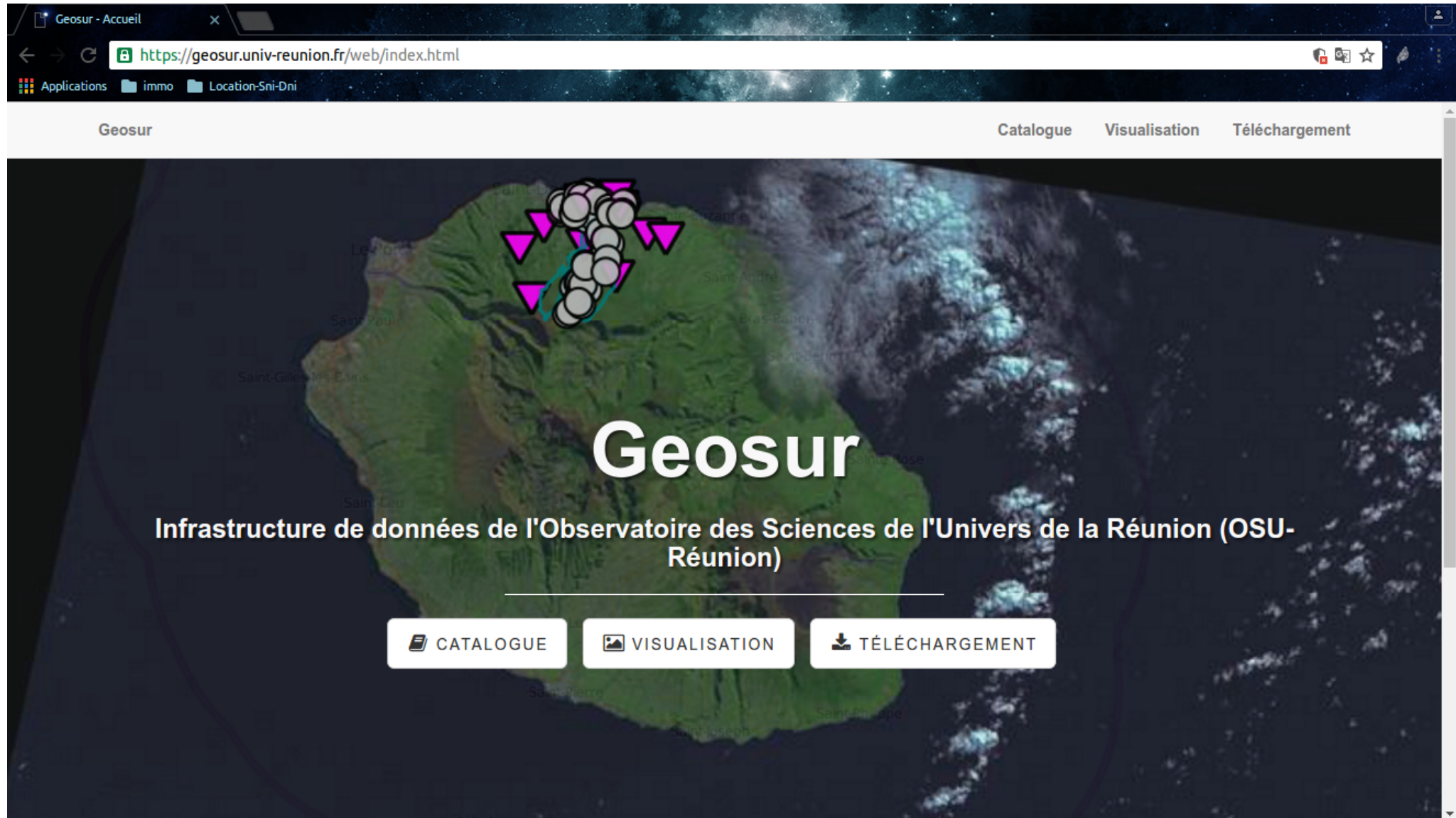
1/ Portail d'accès unique

2/ Une seule base d'authentification

3/ Gestion centralisée des flux

1/ Portail d'accès unique

<https://geosur.univ-reunion.fr>



1/ Portail d'accès unique Catalogue

► Geonetwork :

- Moteur de recherche parmi les dataset OSU-Réunion
- Explication et lien pour télécharger les données

The screenshot shows the Geosur catalog search results for the keyword 'dynamit'. The page features a search bar with the term 'dynamit' and a search button. Below the search bar, there are three search results displayed in a grid. Each result includes a title, a thumbnail image, and a brief description. The first two results are titled 'National Observation Service Dynamit, focuses on Coastline by BD ORTHO@,...' and 'National Observation Service Dynamit, focuses on Coastline by satellite image'. The third result is titled 'DGPS - Topographic beach profiles, Reunion island'. The page also includes a sidebar with filters for 'TYPE DE RESSOURCE', 'FOURNI PAR', 'ANNÉES', 'TYPES DE REPRÉSENTATION', 'FRÉQUENCES DE MISE À JOUR', 'ÉTAT', and 'ÉCHELLE'. The top navigation bar includes 'Geosur', 'Catalogue', 'Visualisation', and 'Téléchargement'.

The screenshot shows the detail page for the resource 'South-western Indian Ocean wave modelling reanalysis 2013-ongoing'. The page includes a search bar with the text 'Retour à la recherche'. Below the search bar, there is a title and a description: 'WW3 model reanalysis on SWIO (south-western indian ocean) area at 0.5 degree of resolution'. A blue banner with the text 'On going' is overlaid on the description. Below the description, there is a section for 'Téléchargements et liens' with a download button and a link to the catalog. The page also includes a section for 'À propos de cette ressource' with a 'Catégories' dropdown menu set to 'Océans' and a 'Mots-clés' section listing 'physical oceanography', 'South Western Indian Ocean', 'Reanalysis 2013-present', 'WAVEWATCH III', and 'Wave Modelling'. On the right side, there is a 'Aperçu' section with a thumbnail image of a wave model and an 'Etendue spatiale' section with a map of the South Western Indian Ocean region.

1/ Portail d'accès unique

Accès aux données

- ▶ Thredds:
 - Open (sans authentification)
 - WMS
 - OpenDAP
 - Web download
- ▶ NextCloud
 - Sous authentification
 - Web Download
- ▶ Serveur FTP
 - Sous authentification
 - FTP

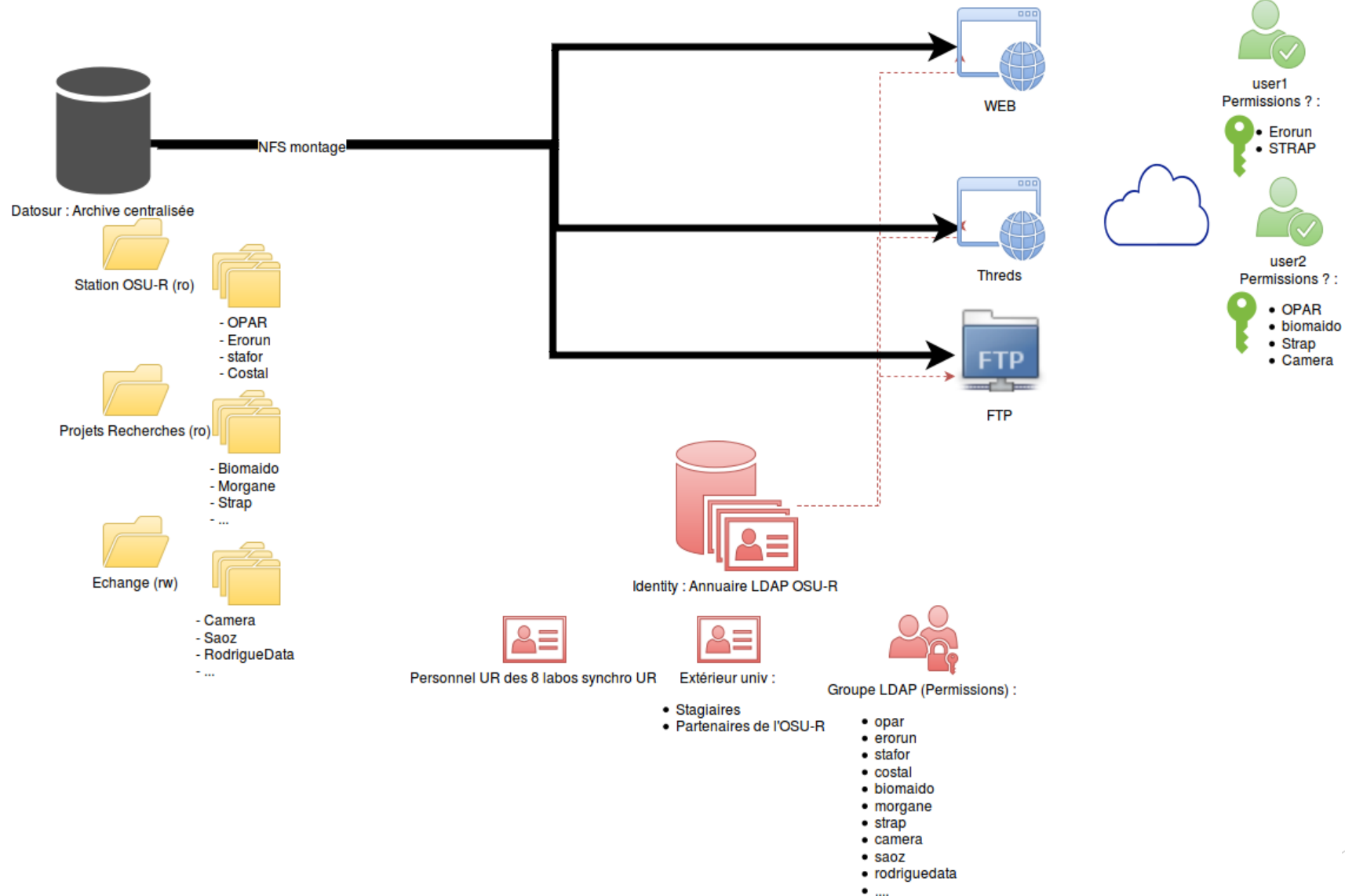
The image displays three screenshots illustrating data access methods for Geosur:

- Thredds Catalog:** A browser window showing a dataset list for 2017. The table includes columns for Dataset, Size, and Last Modified.
- NextCloud File Browser:** A browser window showing a file browser interface for 'coastalstation-data' with folders for 1997 and 2003.
- FileZilla FTP Client:** A screenshot of the FileZilla interface showing a local site 'ents/alicia-fin-aout2017/' and a remote site '/data-access' with various files and folders.

Dataset	Size	Last Modified
2017		--
ww3_swio02_201709.nc	790.3 Mbytes	2018-04-05T10:20:30Z
ww3_swio02_201708.nc	816.7 Mbytes	2018-04-05T10:20:13Z
ww3_swio02_201707.nc	816.7 Mbytes	2018-04-05T10:19:37Z
ww3_swio02_201706.nc	790.3 Mbytes	2018-04-05T10:19:18Z
ww3_swio02_201705.nc	816.7 Mbytes	2018-04-05T10:19:26Z
ww3_swio02_201704.nc	790.3 Mbytes	2018-04-05T10:20:21Z
ww3_swio02_201703.nc	816.7 Mbytes	2018-04-05T10:20:38Z
ww3_swio02_201702.nc	737.7 Mbytes	2018-04-05T10:20:04Z

1/ Portail d'accès unique

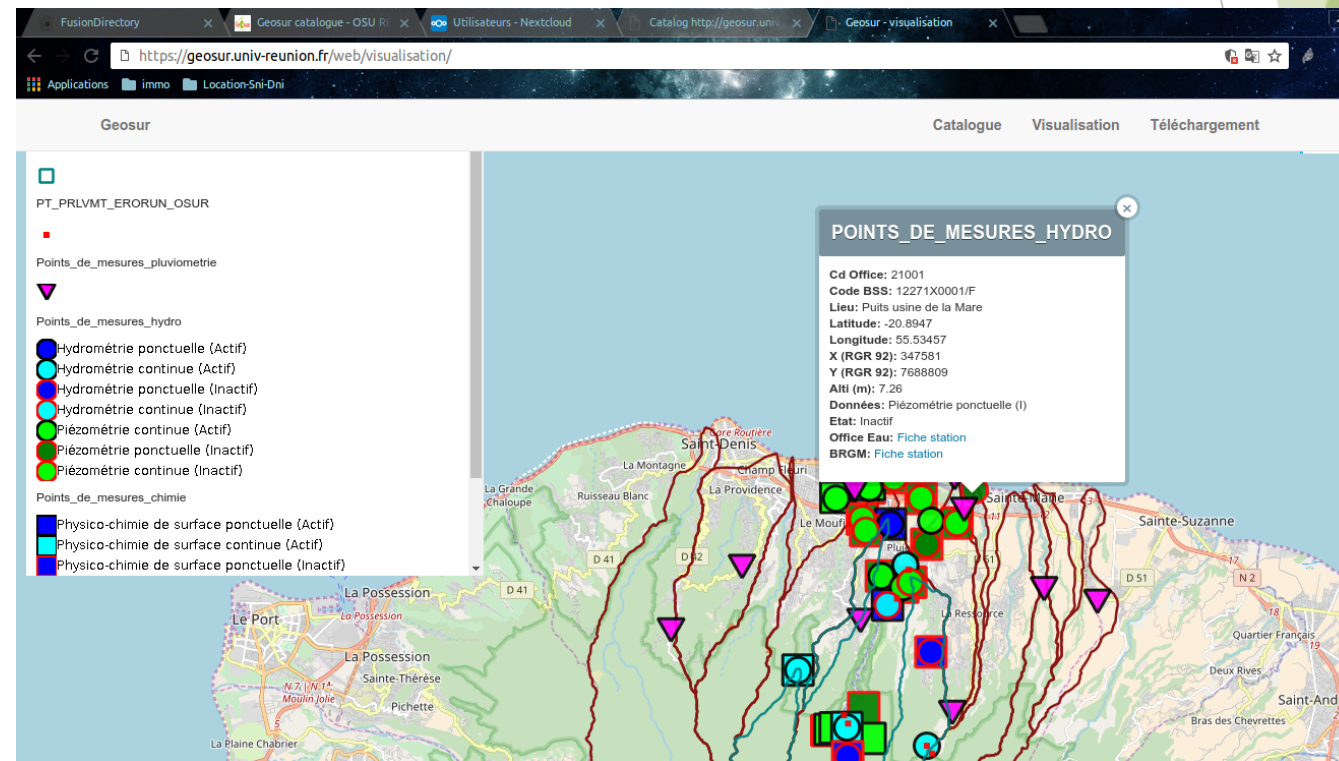
Zarlor : base de fichiers et serveurs NFS



1/ Portail d'accès unique

Visualisation données spatialisées

- ▶ Geocms (front end) :
 - Découvrir les jeux de données in-situ
 - Renvoi vers la fiche geonetwork
- ▶ Geoserver (back end) :
 - Shapefile avec table attributaire descriptif



1/ Portail d'accès unique

e-obs : gestion des métadonnées spécifiques

The screenshot shows the e-obs web portal interface. The browser address bar displays <https://opar.univ-reunion.fr/workspace/opfiles/#>. The page title is "e-obs".

Problèmes techniques

Laser Laser YAG HAUT Refroidisseur YAG HAUT Laser YAG BAS Refroidisseur YAG BAS

Alignement Couplage des faisceaux Emission Reception

Optiques Emission Réception

Electronique d'acquisition Haute tension Carte d'acquisition Détecteur PC d'acquisition

Synchronisation Boîtier électronique Synchro autre laser

Pannes générales Trappe Panne électrique Autres

Commentaires

Conditions atmosphériques

Qualité du Ciel clair voilé cirrus nuages passagés ciel très nuageux pluie

The screenshot shows the e-obs web portal interface for data validation. The browser address bar displays osur-servicesweb.univ-reunion.fr/workspace/PI/. The page title is "Data Validation".

Data Validation

14 10 2014 / instrument: Stratospheric ozone lidar / product: Stratospheric_Ozone_Profile_level2a / version: 2

Data Files :

- renameDstFile_2014_10_14_sru201410141855.ps
- renameDstFile_2014_10_14_sru201410141855.cmt
- renameDstFile_2014_10_14_sru201410141855.mat
- renameDstFile_2014_10_14_sru201410141855.txt
- renameDstFile_2014_10_14_sru201410141855.c11
- renameDstFile_2014_10_14_sru201410141855.ps
- renameDstFile_2014_10_14_sru201410141855.cmt
- renameDstFile_2014_10_14_sru201410141855.mat
- renameDstFile_2014_10_14_sru201410141855.txt
- renameDstFile_2014_10_14_sru201410141855.c11

Validation Form :

1. Please fill in the form: 2. Confirm validation data:

Selected choice validate reprocess reject

Comment Justify choice

Quality index 1 (excellent) 2 (average) 3 (mediocre) 4 (bad)

Validity domain (in km) 4.2 - 6.4

Decision Support : Operator file No file found

Download data set archive : [Download](#)

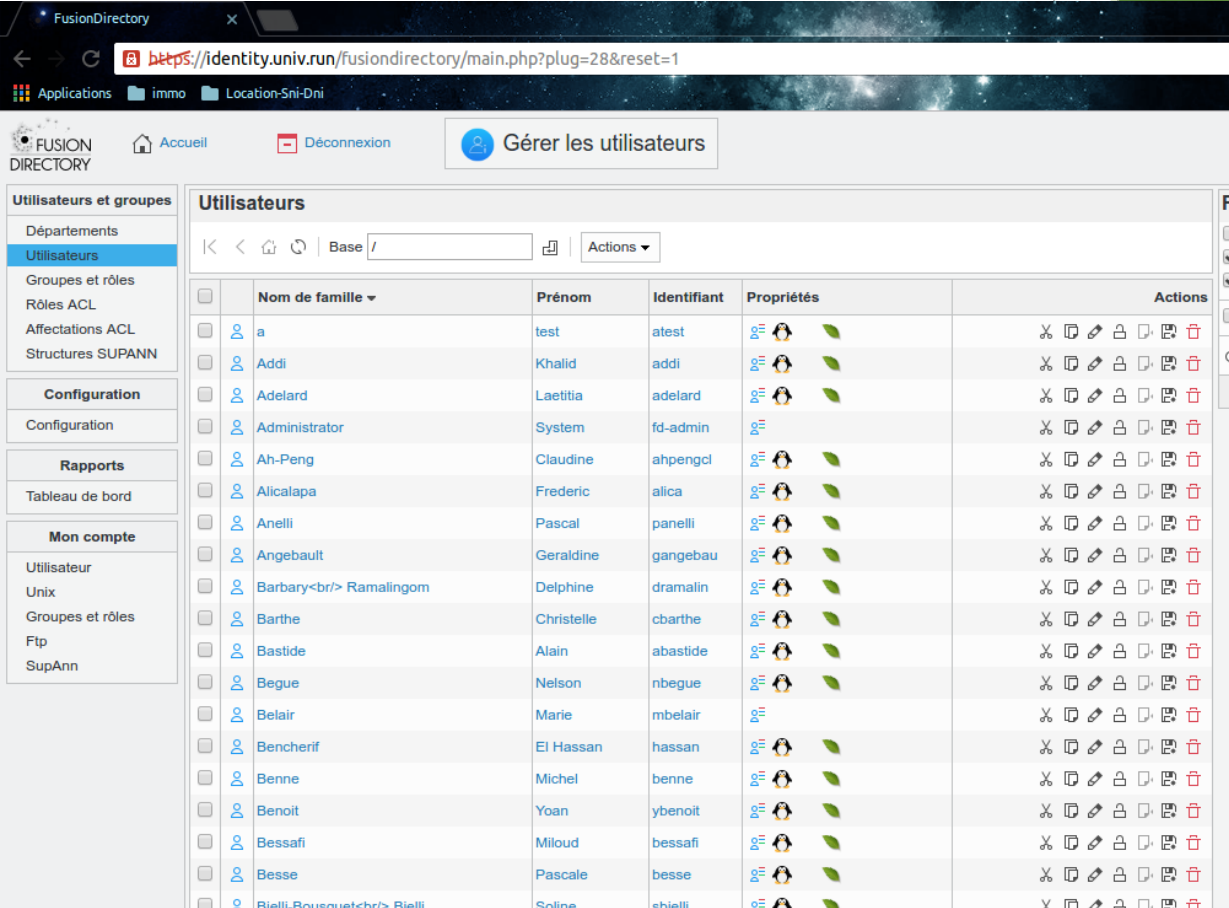
20 10 2014 / instrument: Stratospheric ozone lidar / product: Stratospheric_Ozone_Profile_level2a / version: 2

21 10 2014 / instrument: Stratospheric ozone lidar / product: Stratospheric_Ozone_Profile_level2a / version: 2

2/ Authentication

2/ Authentification openLDAP & fusion directory

- ▶ Serveur openLDAP contenant
 - Personnel de l'OSU-R
 - Stagiaires / Invités
 - Collaborateurs
- ▶ Trier par groupe pour donner +/- de permissions
- ▶ Saisie facilitée par interface web



The screenshot shows the FusionDirectory web interface. The browser address bar displays the URL: `https://identity.univ.run/fusiondirectory/main.php?plug=28&reset=1`. The interface includes a navigation menu on the left with sections like 'Utilisateurs et groupes', 'Configuration', 'Rapports', and 'Mon compte'. The main content area is titled 'Utilisateurs' and contains a table of user entries. Each entry includes a checkbox, a user icon, and columns for 'Nom de famille', 'Prénom', 'Identifiant', 'Propriétés', and 'Actions'.

	Nom de famille	Prénom	Identifiant	Propriétés	Actions
<input type="checkbox"/>	a	test	atest	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Addi	Khalid	addi	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Adelard	Laetitia	adelard	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Administrator	System	fd-admin	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Ah-Peng	Claudine	ahpengcl	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Alicalapa	Frederic	alica	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Anelli	Pascal	panelli	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Angebault	Geraldine	gangebau	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Barbary Ramalingom	Delphine	dramalin	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Barthe	Christelle	cbarthe	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Bastide	Alain	abastide	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Begue	Nelson	nbegue	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Belair	Marie	mbelair	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Bencherif	El Hassan	hassan	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Benne	Michel	benne	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Benoit	Yoan	ybenoit	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Bessafi	Miloud	bessafi	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Besse	Pascale	besse	deu	✖ 📄 🗑️ 🛡️ 📧 📧
<input type="checkbox"/>	Bielli-Bousquet Bielli	Soline	sbielli	deu	✖ 📄 🗑️ 🛡️ 📧 📧

2/ Authentification

Interfaçage LDAP avec les services

Geosur catalogue

Gérer les groupes

Gérer les utilisateurs

Utilisateurs

Filtre

- admin admin (Administrateur)
- Emmanuel Cordier (Administrateur)
- Franck Gabarrot (Administrateur)
- Pierre Stamenoff (Administrateur)
- Remy Decoupes (Administrateur)
- Yoan Benoit (Utilisateur enregistré)

+ Nouvel utilisateur

Browser tabs: FusionDirectory, Geosur catalogue - OSU R..., Utilisateurs - Nextcloud

URL: <https://geosur.univ-reunion.fr/filosur2/index.php/settings/users>

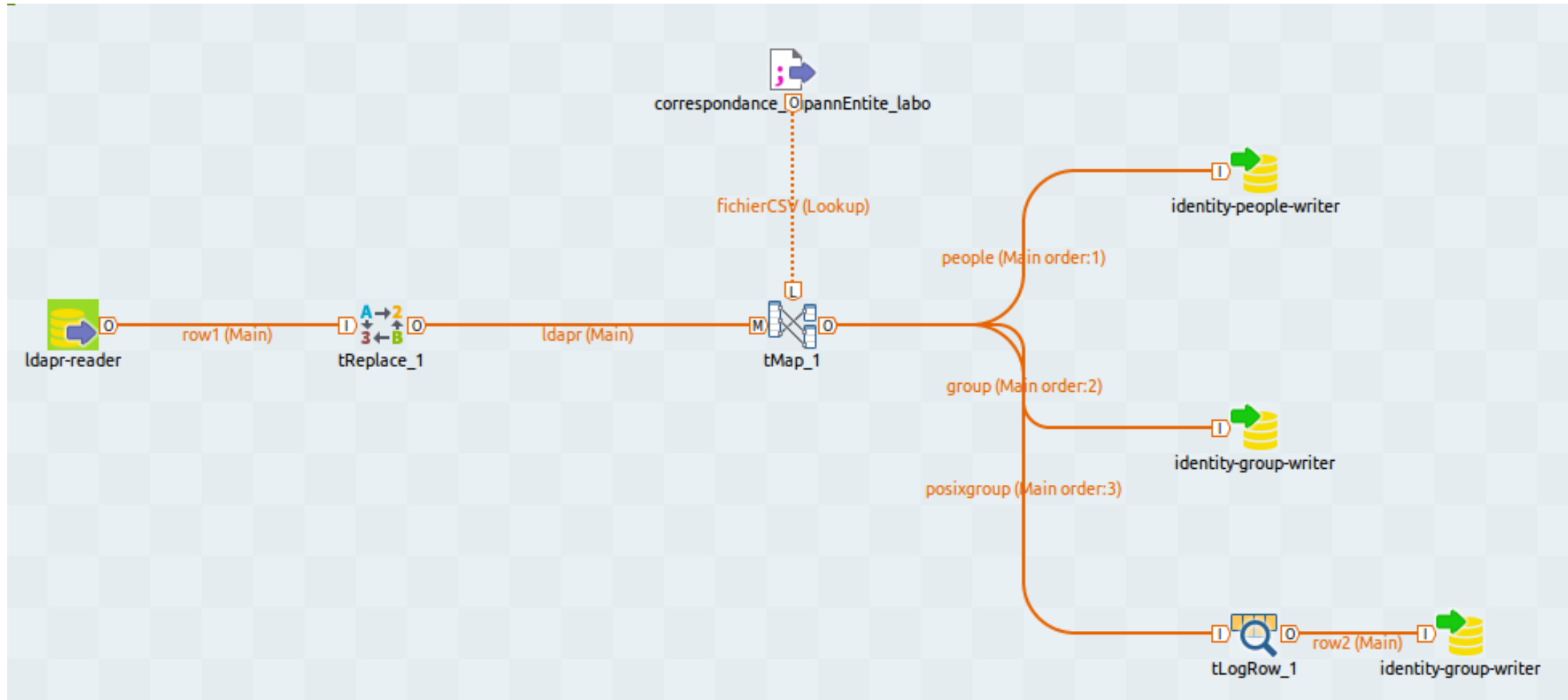
Buttons: Applications, immo, Location-Sni-Dni, +, Ajouter un groupe

Navigation: Tout le monde, Administrateurs, Désactivé, coastalstation, entropie, erorun, espacedev, lacy, le2p, lgsr, lim, piment, Paramètres

Nom d'utilisateur	Mot de passe	Groupes	Créer		
Nom d'utilisateur	Nom complet	Mot de passe	Groupes	Administrateur de groupe pour	Quota
addi	Khalid Addi	piment, restrictedData...	aucun groupe	Illimité
adelard	Laetitia Adelard	piment, restrictedData...	aucun groupe	Illimité
adennemo	Marie-Alice Dennemont	ums3365, restrictedData...	aucun groupe	Illimité
admin	admin	admin	aucun groupe	Illimité
arechou	Anne Rechou	lacy, restrictedDatasetUs...	aucun groupe	Illimité
benne	Michel Benne	le2p, restrictedDatasetU...	aucun groupe	Illimité
bessafi	Miloud Bessafi	le2p, restrictedDatasetU...	aucun groupe	Illimité
besse	Pascale Besse	pvbmt, restrictedData...	aucun groupe	Illimité
bgrondin	Brigitte Perez	le2p, restrictedDatasetU...	aucun groupe	Illimité
boyer	Harry Boyer	piment, restrictedData...	aucun groupe	Illimité

2/ Authentication

Talend : peuplement automatique du LDAP

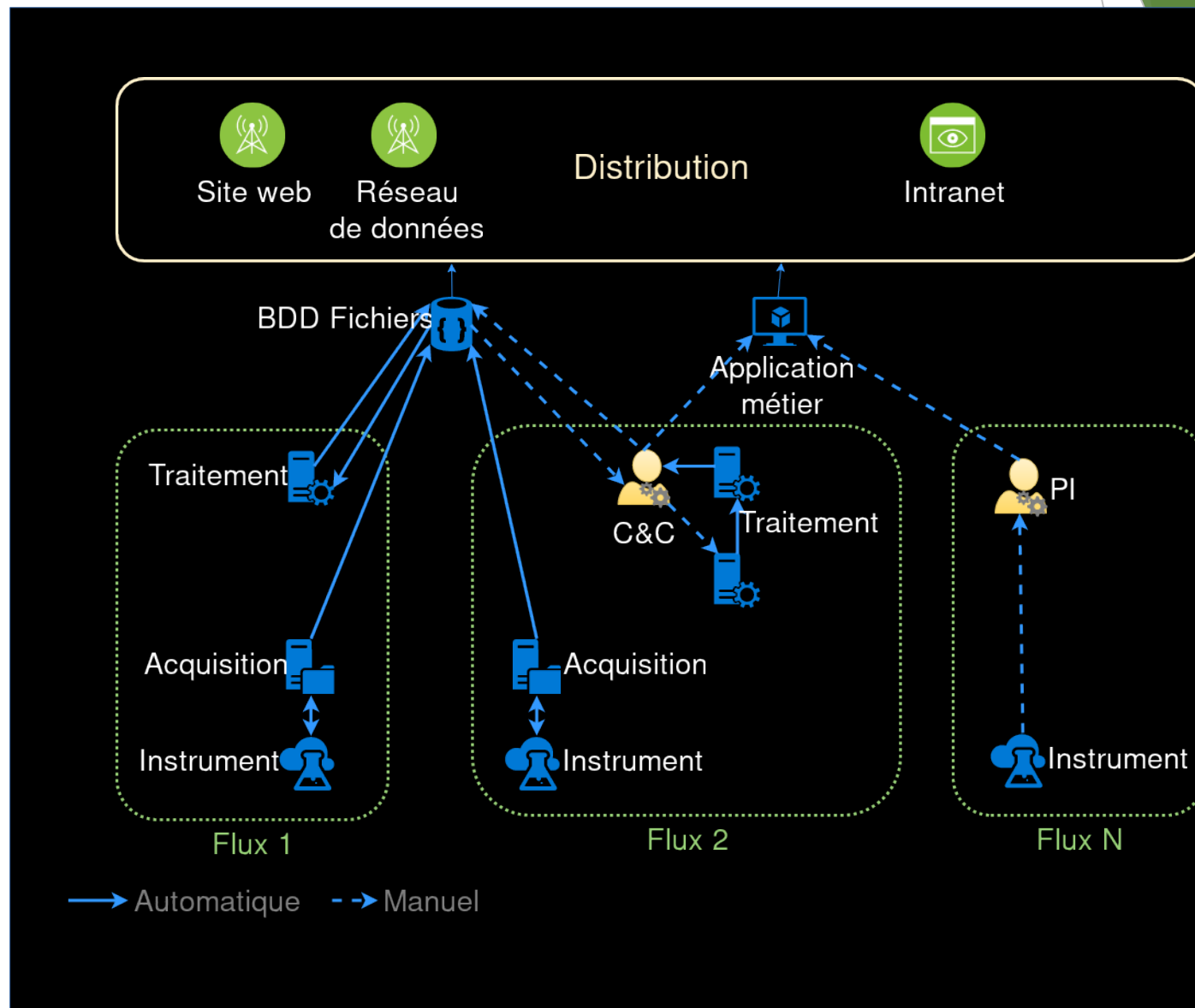


3/ Gestionnaire de production des données

3/ Gestionnaire des flux

Problématique

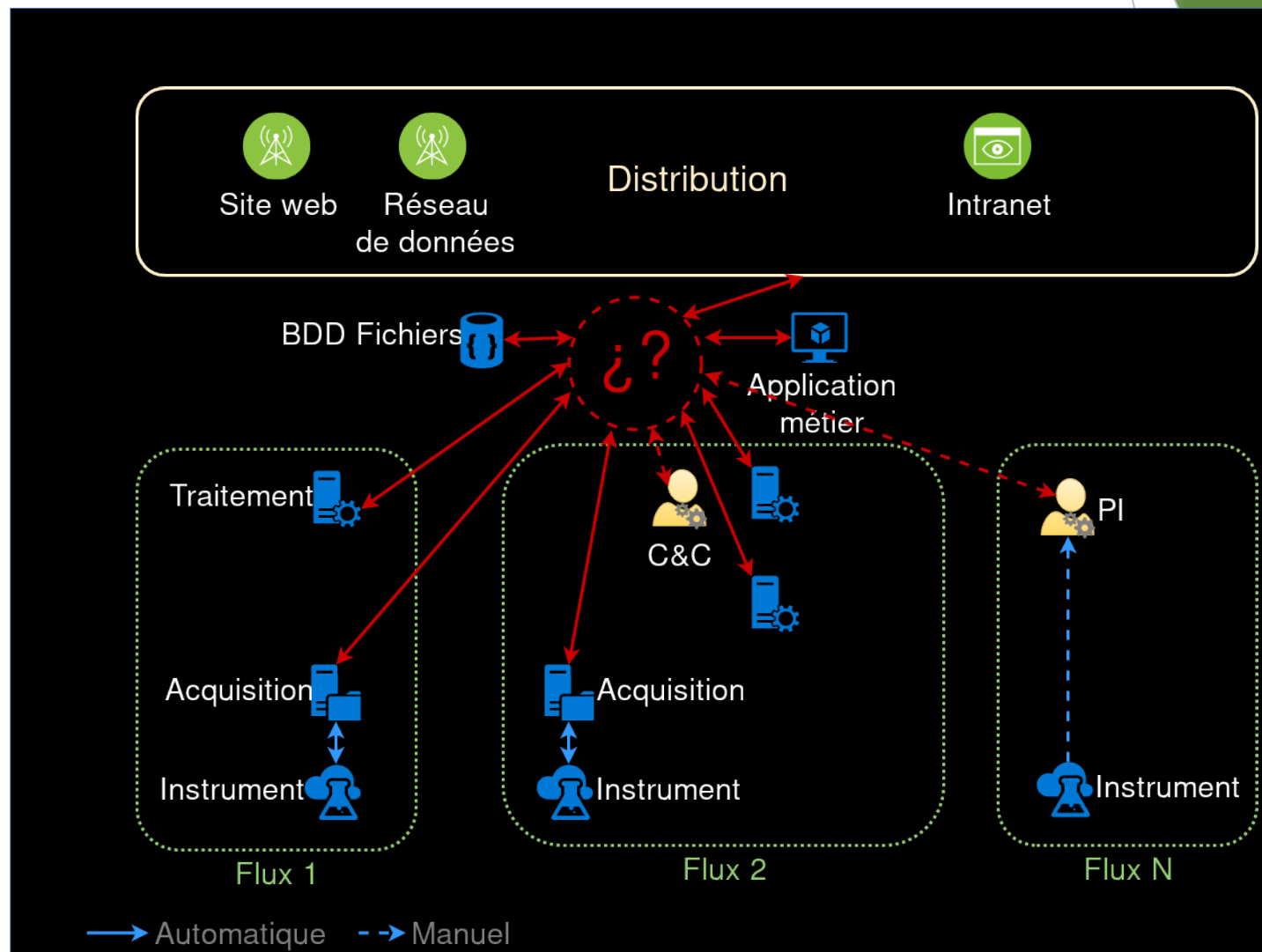
- ▶ Python, Fortran, Matlab, IDL, traitement manuel ...
 - beaucoup de langages différents
- ▶ Logique des flux distribués
 - informations/données distribuées
- ▶ Impossible de visualiser rapidement l'état de la production de données
- ▶ Difficulté de maintenance
- ▶ Charge inconnue



3/ Gestionnaire des flux

Objectif

- ▶ Unifier l'architecture et les langages pour les orchestrations et la surveillance
- ▶ Centraliser la logique et la documentation
- ▶ Dashboard pour le monitoring du fonctionnement des flux et de la charge



3/ Gestionnaire des flux

airflow



- ▶ Airflow

<https://airflow.apache.org>

- ▶ Interface web de contrôle et de commande des flux / pipeline
- ▶ Distribue le travail sur différents nœuds de traitement

3/ Gestionnaire des flux

airflow



```
45 # Default argument that will be applied to all tasks by default
46 default_args = {
47     'owner': 'airflow',
48     'start_date': datetime(2018, 5, 25, 0, 0, 0),
49 }
50
51 dag = DAG(
52     dag_id='flux-exemple-sist2018',
53     description='',
54     schedule_interval='@daily',
55     start_date=datetime(2018, 5, 25, 0, 0, 0),
56     end_date=None,
57     default_args=default_args,
58     params={},
59     concurrency=2,
60     max_active_runs=2,
61     dagrun_timeout=timedelta(3600),
62     default_view='graph',
63     orientation='TB',
64     catchup=True)
65
66
67 # TASK CONFIGURATION
68 classify = PythonOperator(task_id='classify',
69                           python_callable=classify_files,
70                           provide_context=True,
71                           op_args=[SOURCE, DEST],
72                           dag=dag)
73
74 tpl_cmd = "rm %s/data_{{ ds_nodash }}.csv" % SOURCE
75 clean_input = BashOperator(task_id='clean_input',
76                             bash_command=tpl_cmd,
77                             dag=dag)
78
79 # WORKFLOW CONFIGURATION
80 classify >> clean_input
81
```

3/ Gestionnaire des flux airflow

Off DAG: specinduv_OA_L0L2_v1.0 Orchestration for specinduv_ET_L0_v1.0 and specinduv_ET_L2_v1.0

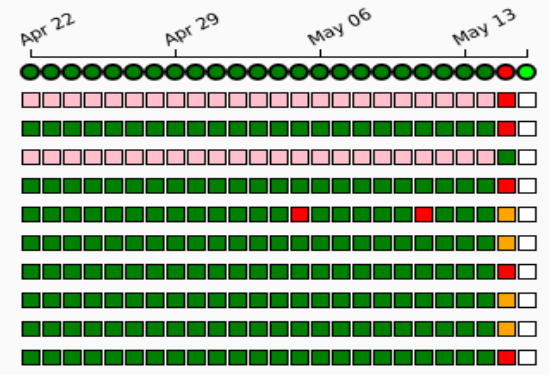
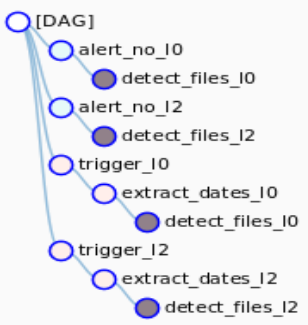
schedule: @daily

- Graph View
- Tree View**
- Task Duration
- Task Tries
- Landing Times
- Gantt
- Details
- Code
- Refresh

Base date: 2018-05-16 00:00:00 Number of runs: 25 Go

EmailOperator PythonOperator SSHFileSensor TriggerDagRunOperatorLoop

success running failed skipped retry queued no status



Conclusion

Merci pour votre attention

Remy.Decoupes@univ-reunion.fr

Osureunion-informatique@univ-reunion.fr

