



Euro Argo Use Case

Euro Argo Use Case

DI4R 2017



Presenter: Claudio Cacciari
(c.cacciari@cinca.it)



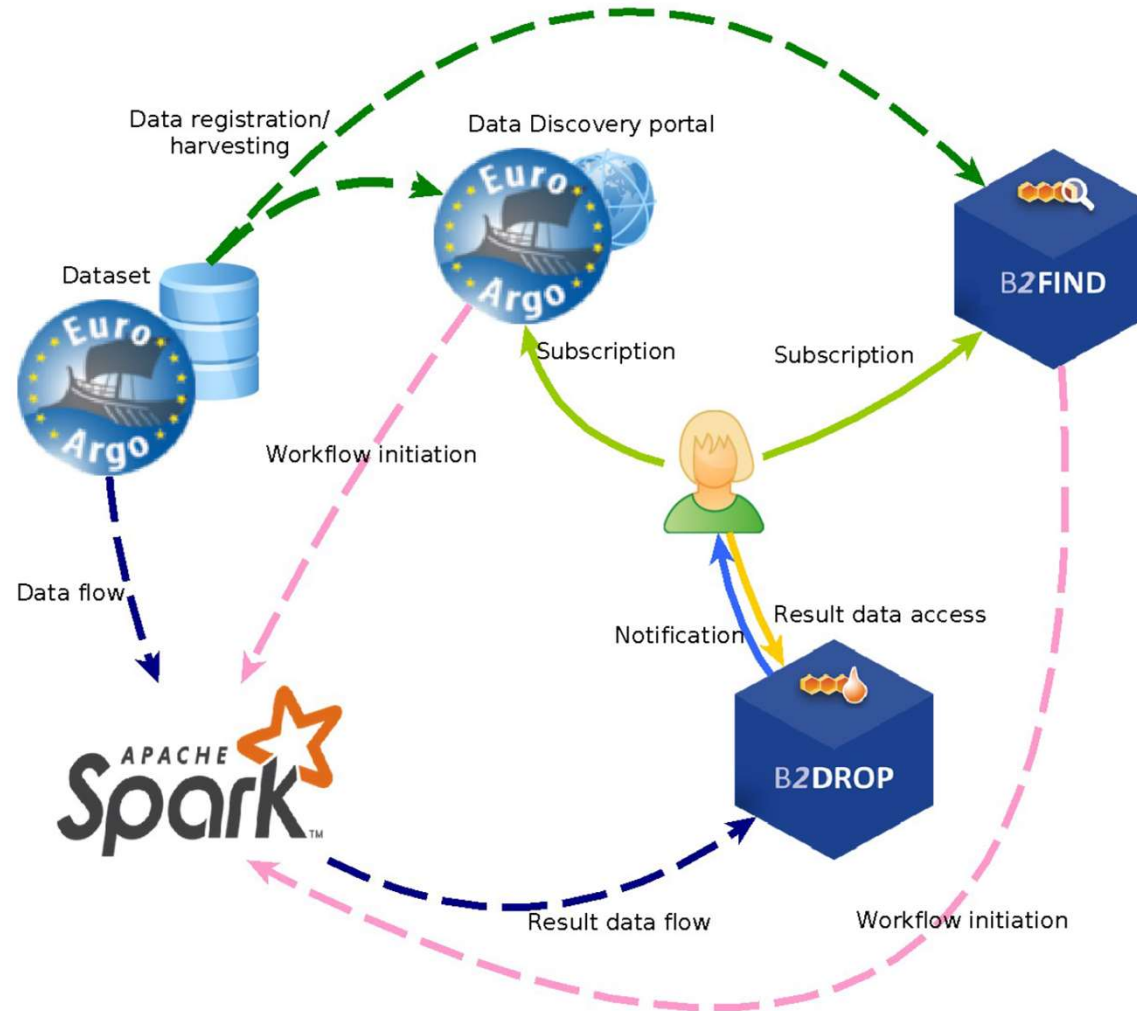
Euro Argo Use Case : Context Reminder

- 1 archive (15GB) of Monthly CSV files (>1300 files)
 - Available on IFREMER FTP
 - 500 GB expanded (biggest monthly csv file ~ 4GB)
 - 5,6 billion records
 - 5% growth per year
 - This archive was used for test purpose (Spark, Hive, etc ...)
- Evolution
 - The last month file updated daily + Historical (CSV files or archive)
 - Historical part updated monthly (or annually)



Euro Argo Use Case : Workflow

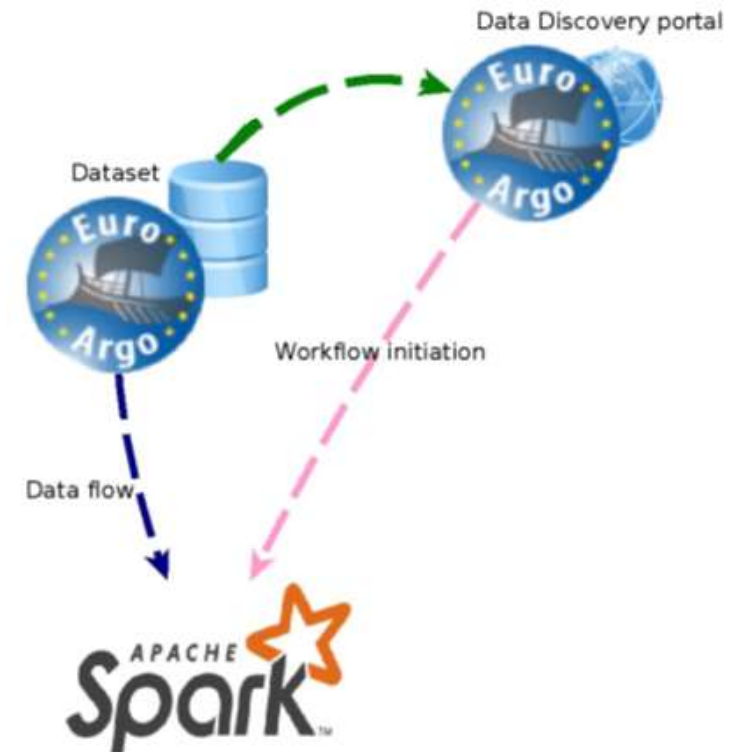
How DataSet is :
- Provided
- Initialized
- Updated (Daily)
- Processed (Spark)





Euro Argo Use Case : Meeting Focus

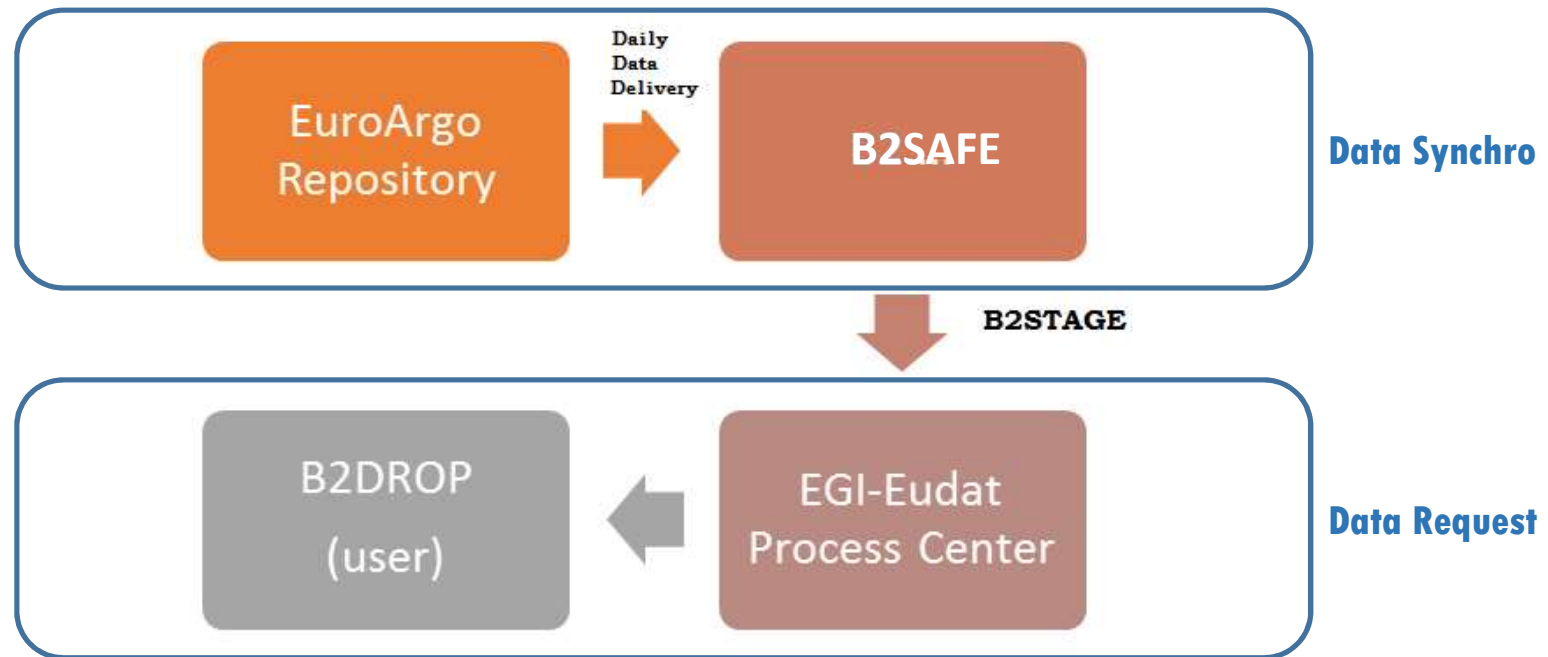
- Data Discovery Portal (DDP) in IFREMER
- DDP has to manage the Workflow initiation
- Abstract :
 - We need a hosted bigdata space to push the daily update and where request will be executed to extract subsets of data. (B2SAFE)
 - The results will be pushed on user's cloud services (B2DROP)





Euro Argo Use Case : Data Flow

- Monthly Data files
- Current Month Data File need to be updated daily
- Best B2 service that match our needs → B2SAFE
- How to set a process that daily push our file on B2 service → B2Drop and Cadaver





Euro Argo Use Case : Data Synchronization

- Data Upload and Synchronization :
 - Ifremer generate daily's update of current month data file
 - Data files are updated/synchronized with B2SAFE

IFREMER

- 500 GB
- Monthly Update

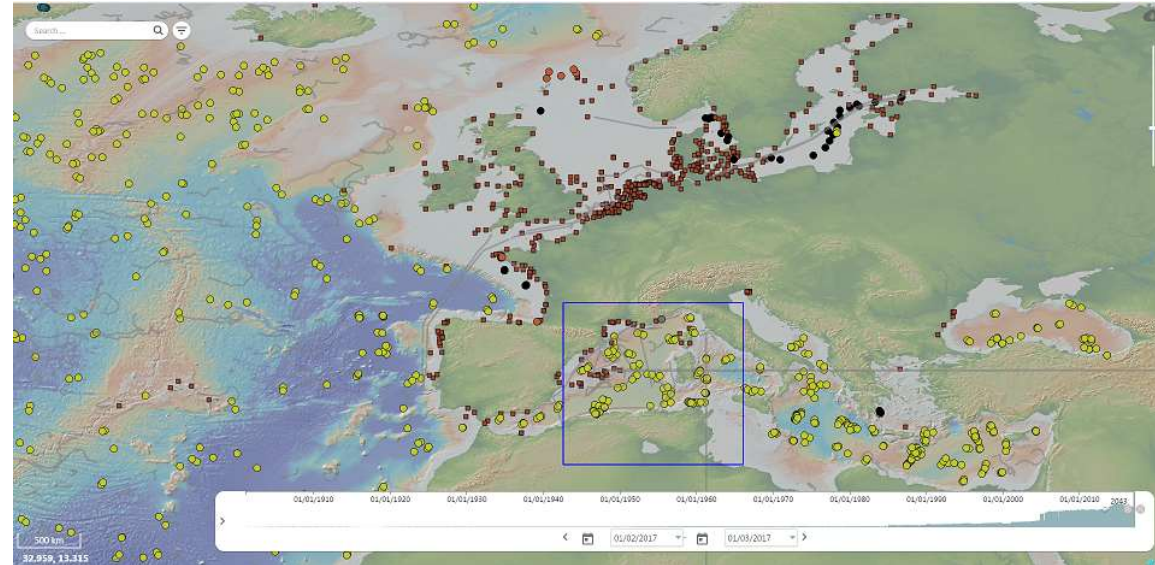
B2SAFE

- Store Data
- Data is synchronized or available for COMPUTER CENTER



Euro Argo Use Case : Data Subscription Process

- Data Subscription Process is managed by IFREMER
 - Data Visualization Interface :
 - Unique Download Capability
 - User can subscribe:
 - Daily (minimum period), Monthly
 - Criteria are stored
 - IFREMER application :
 - Store user's requests
 - Manage subscriptions
 - Application initiate workflow by sending request to Computation Center :
 - API on Computaiton Center's side to communicate
 - Request are sent 10 to 100 times a day
 - Data extraction could be zipped NetCDF and CSV files
 - Results data set are pushed on user's B2DROP space





Video demonstration

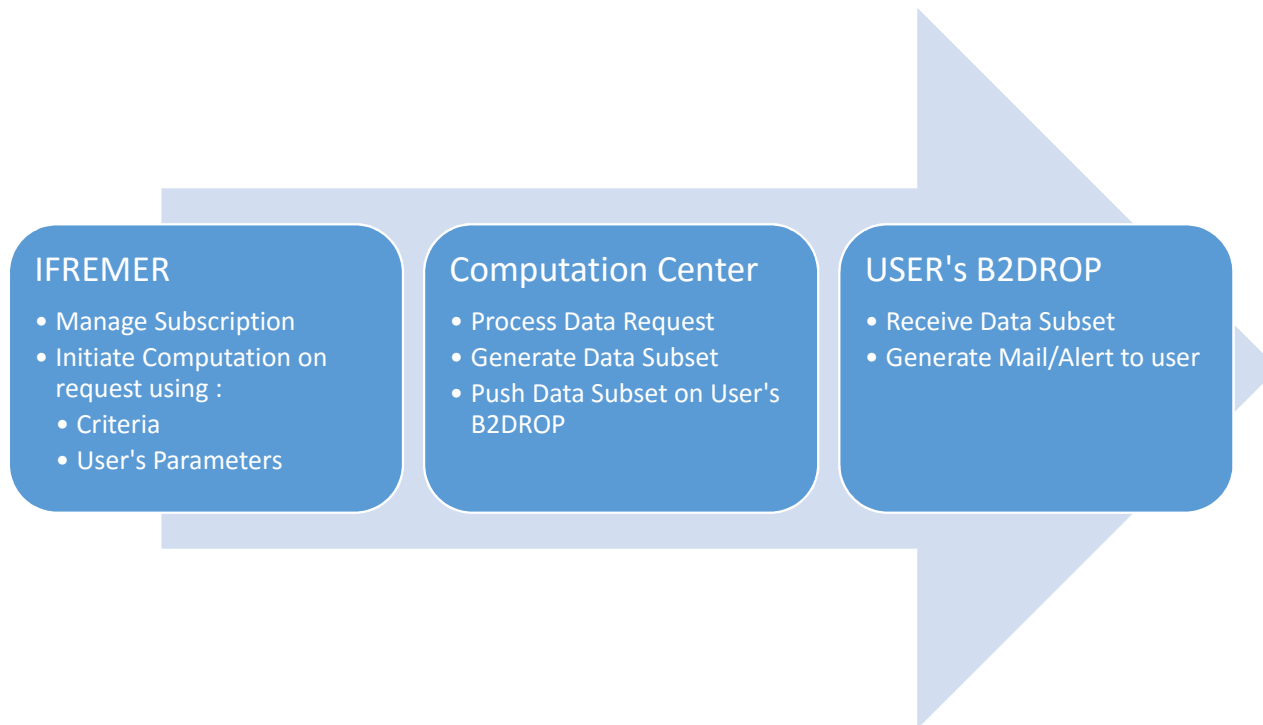
- https://www.youtube.com/watch?v=PKU_JcmSskw



Euro Argo Use Case : Data Request

■ Computation Center :

- Required API to communicate with IFREMER's application
- Manage to get access/synchronize IFREMER's data
- Process request sent and push results on user's B2DROP





Interoperability & reusability

- The Data Subscription Service is
 - **Interoperable:**
 - Data storage space based on EUDAT B2SAFE and B2DROP, which offer standard interfaces like webdav, http api and GridFTP.
 - Data analysis space based on EGI Dynamic Real-time Infrastructure Planner (DRIP), which can be synchronized through GridFTP/http protocols.
 - **Reusable:** the only specific component is the community portal for data discovery, and the connector between the portal and the DSS, therefore everything else is reusable



Questions?