AN ARCHITECTURE TO MASSIVELY DOWNLOAD AND DEAL WITH SATELLITE IMAGES

Thursday, October 3rd 2019

Johnny NGUYEN

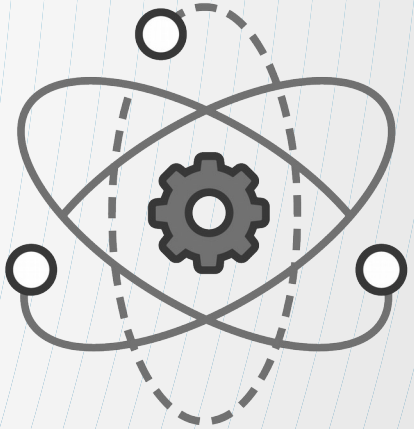
M2 SSTIM

Earth observation

Manage satellite data

Tools easily manipulated by scientists

ENVIRONMENT



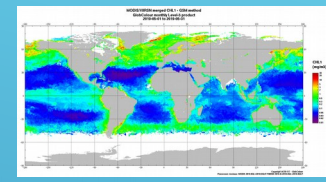
MANAGE SATELLITES DATA

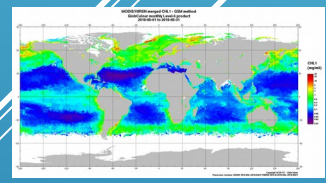
Gross product

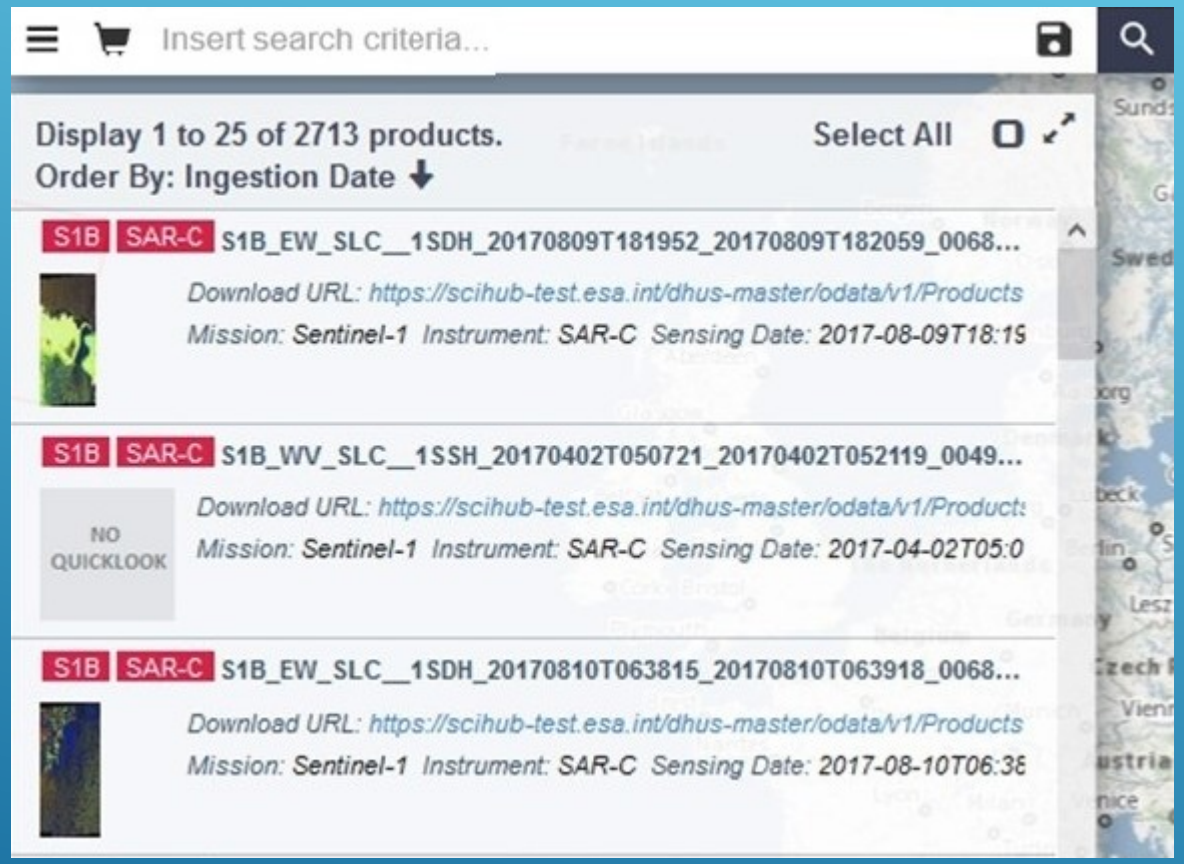
```
FFFF FFFF
AA99 5566
```

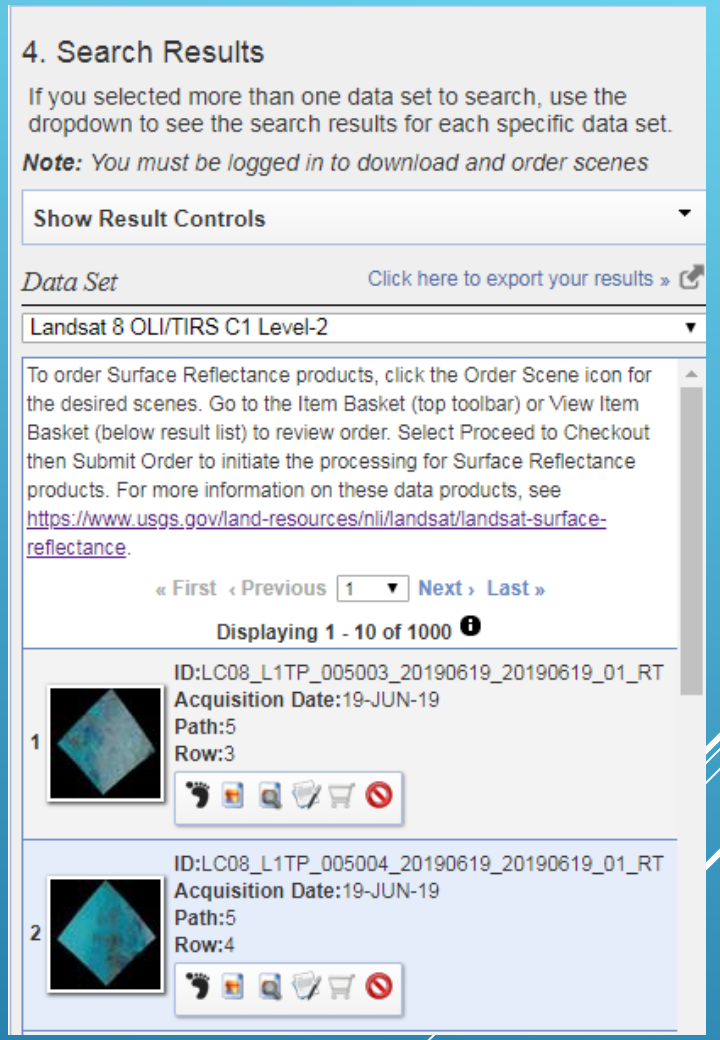
Antennas network

Providers

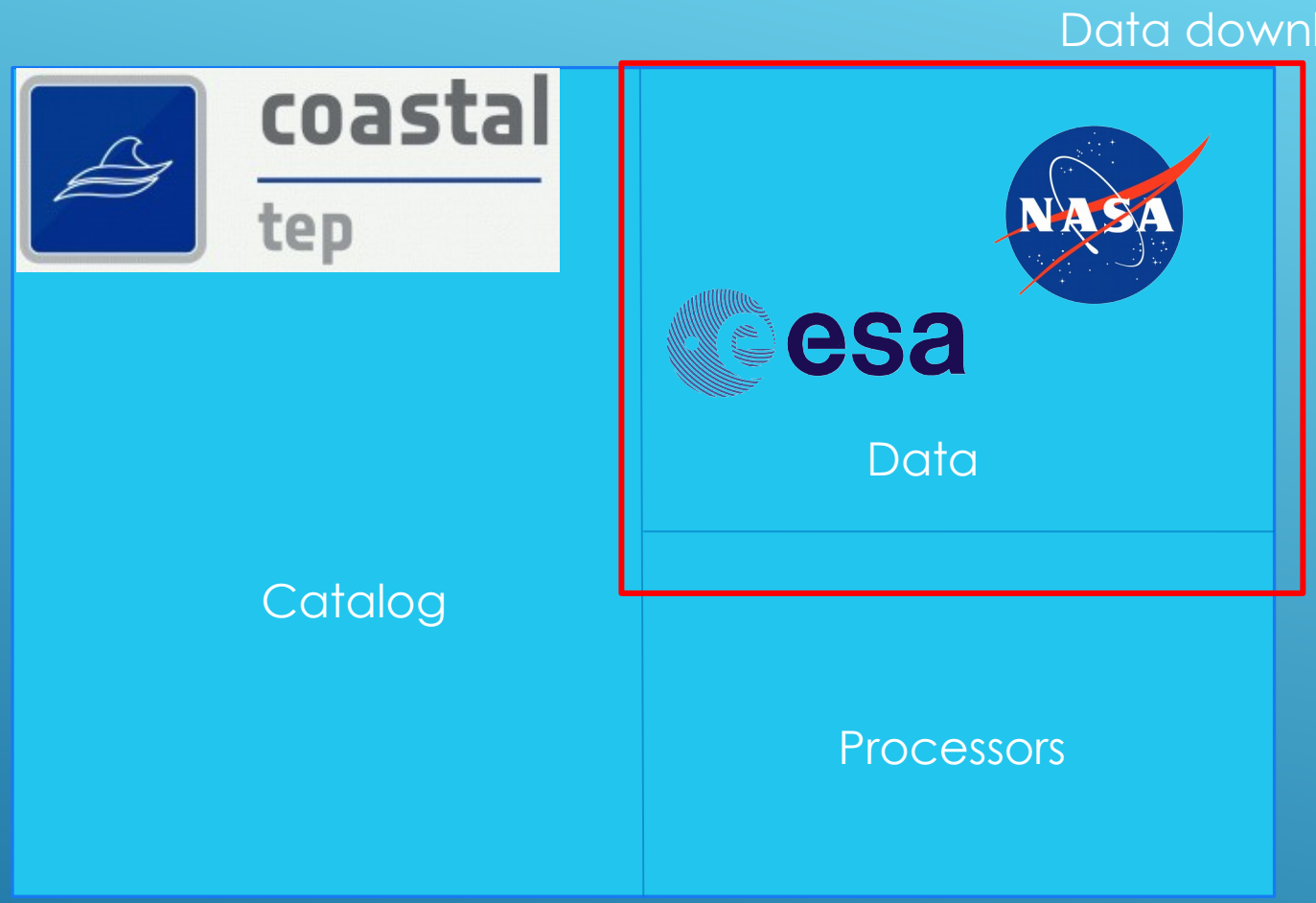





Insert search criteria...
Display 1 to 25 of 2713 products.
Order By: Ingestion Date
Select All
S1B SAR-C S1B_EW_SLC__1SDH_20170809T181952_20170809T182059_0068...
Download URL: https://scihub-test.esa.int/dhus-master/odata/v1/Products
Mission: Sentinel-1 Instrument: SAR-C Sensing Date: 2017-08-09T18:19
S1B SAR-C S1B_WV_SLC__1SSH_20170402T050721_20170402T052119_0049...
NO QUICKLOOK
Download URL: https://scihub-test.esa.int/dhus-master/odata/v1/Product
Mission: Sentinel-1 Instrument: SAR-C Sensing Date: 2017-04-02T05:0
S1B SAR-C S1B_EW_SLC__1SDH_20170810T063815_20170810T063918_0068...
Download URL: https://scihub-test.esa.int/dhus-master/odata/v1/Products
Mission: Sentinel-1 Instrument: SAR-C Sensing Date: 2017-08-10T06:38

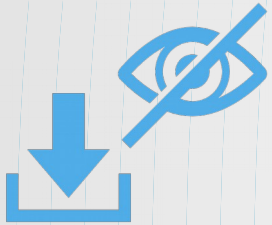

4. Search Results
If you selected more than one data set to search, use the dropdown to see the search results for each specific data set.
Note: You must be logged in to download and order scenes
Show Result Controls
Data Set
Click here to export your results »
Landsat 8 OLI/TIRS C1 Level-2
To order Surface Reflectance products, click the Order Scene icon for the desired scenes. Go to the Item Basket (top toolbar) or View Item Basket (below result list) to review order. Select Proceed to Checkout then Submit Order to initiate the processing for Surface Reflectance products. For more information on these data products, see https://www.usgs.gov/land-resources/nli/landsat/landsat-surface-reflectance.
« First ‹ Previous 1 Next › Last »
Displaying 1 - 10 of 1000
1
ID:LC08_L1TP_005003_20190619_20190619_01_RT
Acquisition Date:19-JUN-19
Path:5
Row:3
2
ID:LC08_L1TP_005004_20190619_20190619_01_RT
Acquisition Date:19-JUN-19
Path:5
Row:4

PROVIDERS



- Volume : 110 To
- Data transformation
- User interface

COASTAL THEMATIC EXPLOITATION PLATFORM



ISSUES

Complex download not tracked

Crash sensitive software

Different sequential downloads

Download → Error → Recovery

- Downloads state not tracked
- Complex download procedure

- How to automatically restart the software ?
- How to automatically restart an interrupted download ?
- How to manage different types ?
- Why not download simultaneously?

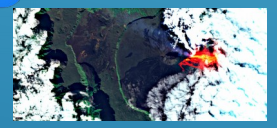
COMPLEX DOWNLOAD NOT TRACKED

System crash

HTTP

Software

Download interruption

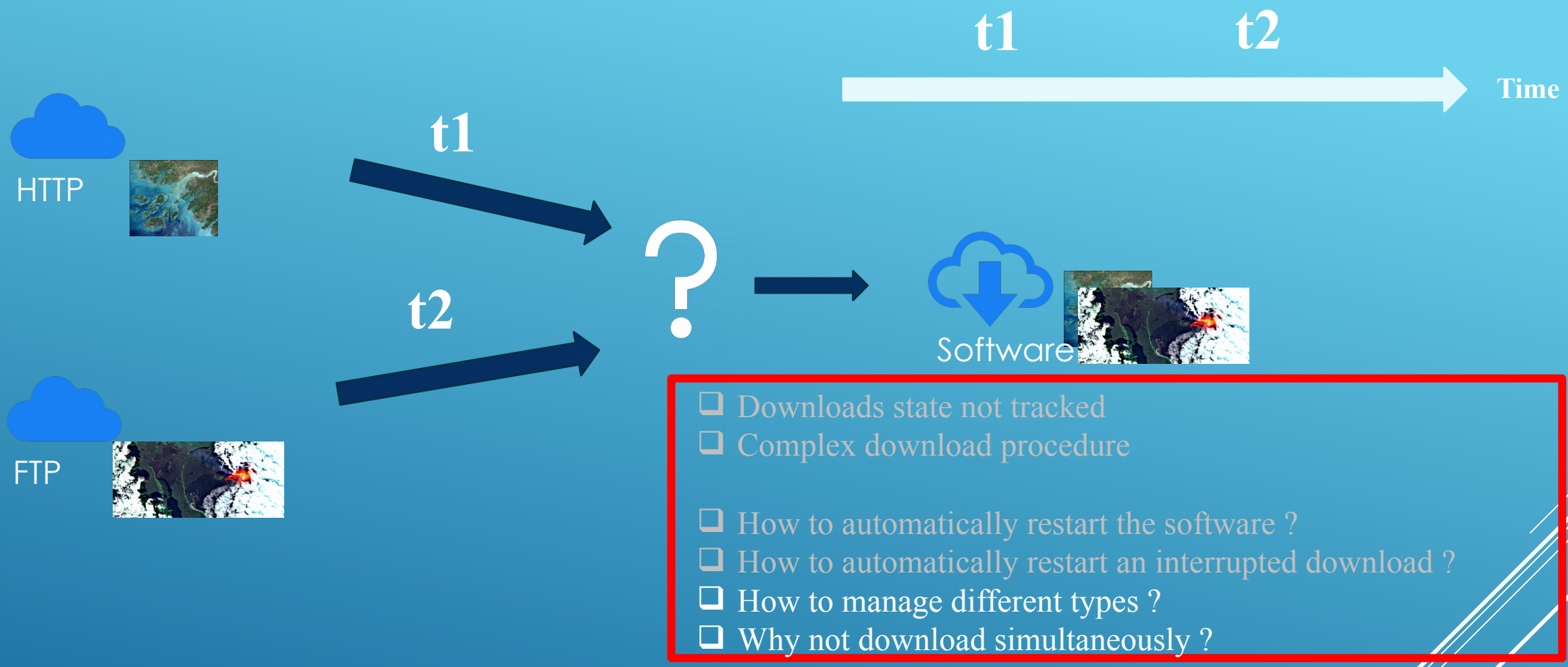


FTP

- Downloads state not tracked
- Complex download procedure

- How to automatically restart the software ?
- How to automatically restart an interrupted download ?
- How to manage different types ?
- Why not download simultaneously ?

CRASH SENSITIVE SOFTWARE



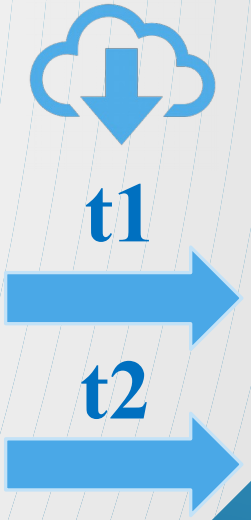
DIFFERENTS SEQUENTIAL DOWNLOAD TYPES

Download module

Automatic download recovery

Fast scalable interface

SOLUTIONS



Collection	
PK	id_collection
	collection

Provider	
PK	id_provider
	provider

Credential	
PK	id_credential
PK,FK1	id_provider
	api
	user
	password

datahub_tools

+ download(datahub, id_tag, zipfile, safe, output_dir): string
+ datahub_request(request): [], [], []
+ getListProduct(request): [], [], []
+ download(prod_n, path_p, path_l, credentials): string

collection text	provider text	user text	password text
Sentinel-2	cophub	s2mpc-ops	
Sentinel-3 Water	coda	sclerc	
Sentinel-5	s5phub	s5pguest	

Many providers managed with the database

Download state obtained

DownloadManager

+ timetosleep: int

__init__()
+ run(): void
- todoupdateto(idDownload, newstatus): void
- todo(requests): void

Request	
PK	IdRequest
FK	id_request_status
	datebegin
	dateend

id_request text	product_name text	status text
2867	S2B MSIL1C 20180216T152	Done
2869	S2B MSIL1C 20171121T153	Done
9	S3A OL 2 WRR 2018102	Downloading...
9	S3A OL 2 WRR 2018102	Downloading...
2818	S2A MSIL1C 20190220T150	Done
9	S3A OL 2 WRR 2018102	To do
9	S3A OL 2 WRR 2018102	To do

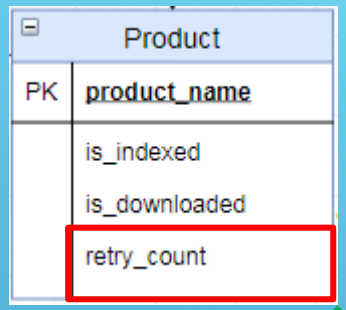
Download_status	
PK	id_download_status
	status

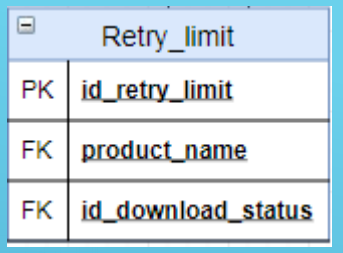
Download	
PK	IdDownload
FK	IdRequest
FK	id_download_status
	product_name

DOWNLOAD MODULE

Solutions

id_request text	product_name text	status text
2867	S2B MSIL1C 20180216T152619	Done
2869	S2B MSIL1C 20171121T153549	Done
9	S3A OL 2 WRR 20181029T6	Downloading...
9	S3A OL 2 WRR 20181029T6	Downloading...
2818	S2A MSIL1C 20190220T150711	Done
9	S3A OL 2 WRR 20181029T1	To do
9	S3A OL 2 WRR 20181029T1	To do





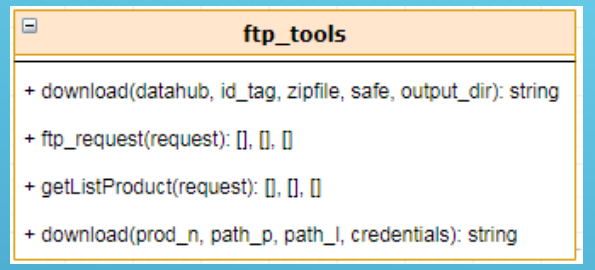
Without download interruption

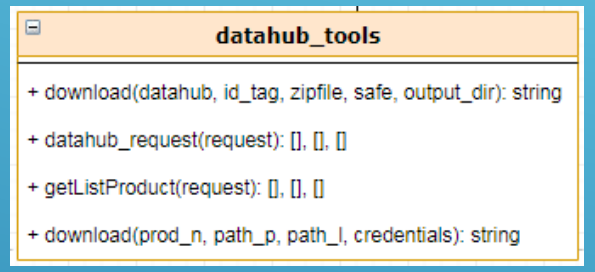
```yaml
services:
  cc-download-manager:
    logging:
      driver: "json-file"
      options:
        max-file: "10"
        max-size: "10m"
    user: "${ID}"
    privileged: true
    build:
      context: .
      dockerfile: Dockerfile.download_manager
    depends_on:
      - cc-postgres
    restart: always
```

```python
def run(self):
    """
    Launch every worker.
    Only if there at least one credential on the database
    """
    # At the beginning check if there was nothing to download
    restart_incomplete_downloads()
    while True:
        try:
            credentials = self.db.get_all_credential()
            # create rows on table worker only if there wasn't exist
            self.init_table_workers(credentials)
            break
        except:
            logging.info("DownloadManager : Not thread to launch, I wait 5 minutes")
            time.sleep(300)
    # launch the different thread
    self.launch_workers(credentials)
```

Cleanly restarted software

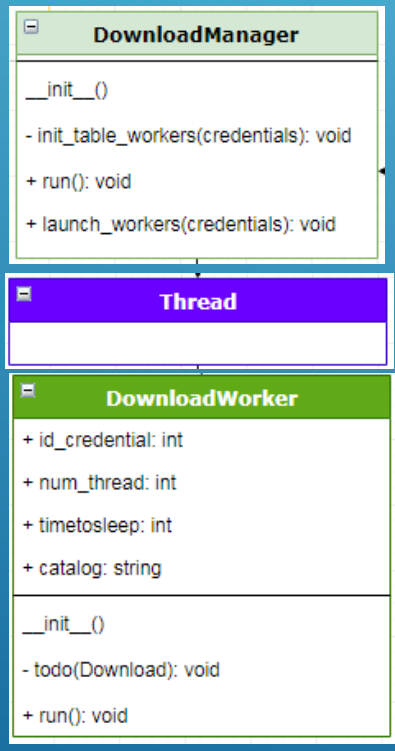
AUTOMATIC DOWNLOADING RECOVERY





id_request text	product_name text	path_provider text	status text
890	2019042100 6h	ftp://nrt.cme	Done
666	metoffice foa	ftp://nrt.cme	Done
764	metoffice foa	ftp://nrt.cme	Done
90	2019041806 6h	ftp://nrt.cme	Done
680	metoffice foa	ftp://nrt.cme	Done
63	S2B MSIL1C 20	https://scihu	Alread
44	S2B MSIL1C 20	https://scihu	Alread
45	S2B MSIL1C 20	https://scihu	Alread
43	S2B MSIL1C 20	https://scihu	Alread

Generic downloading

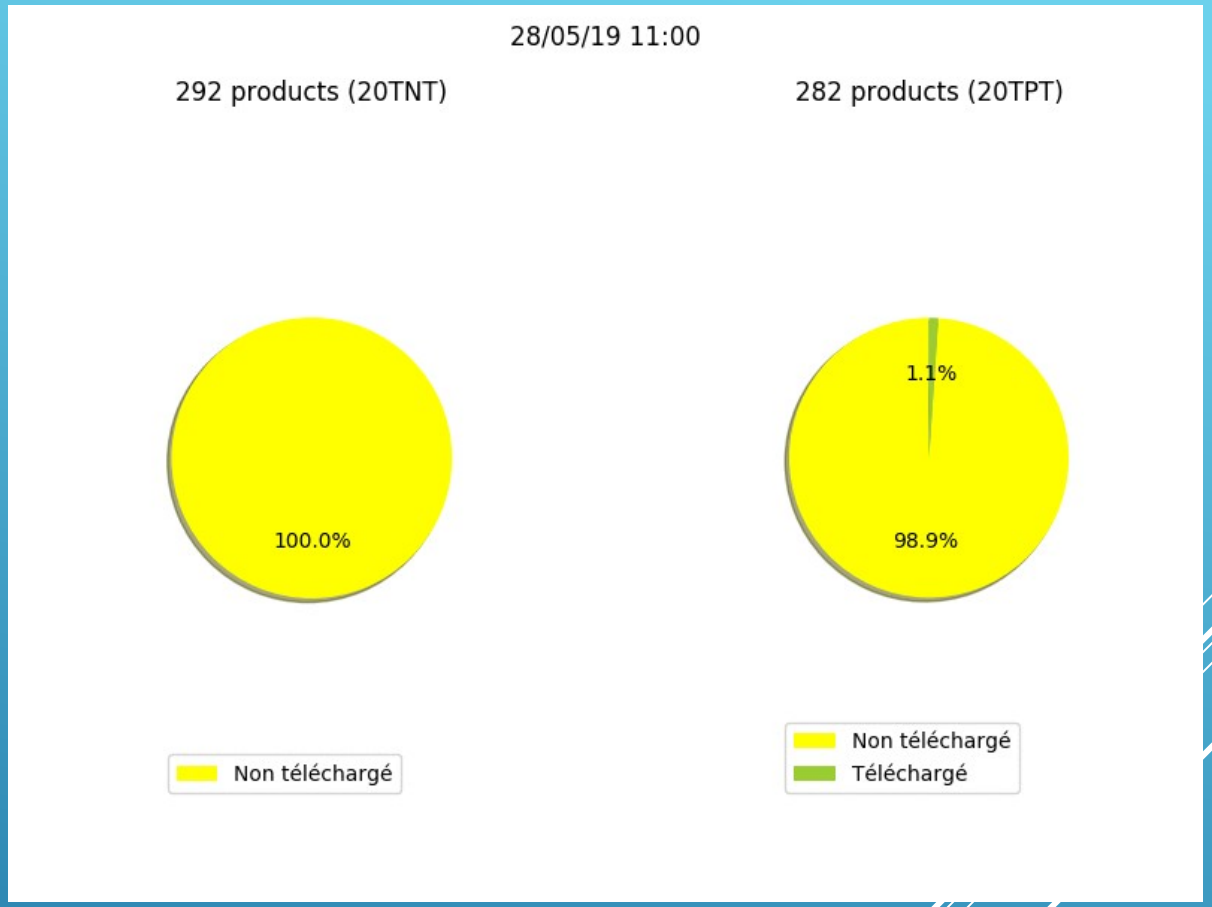


provider text	num_thread integer	product_name text	date_begin timestamp without time zone	nb_thread integer
scihub	1	S2B MSIL1C 20	2019-03-27 14:01:20.31516	1
s5phub	1			1
coda	1	S3A OL 2 WRR	2019-05-16 14:55:37.346532	1
cophub	1	S2A MSIL1C 20	2019-06-16 09:01:01.948784	5
cophub	2	S2B MSIL1C 20	2019-06-26 12:00:27.501211	5
cophub	3	S2B MSIL1C 20	2019-06-26 12:00:27.559485	5
cophub	4	S2A MSIL1C 20	2019-06-26 09:32:04.299751	5
cophub	5	S2A MSIL1C 20	2019-06-22 09:00:39.444981	5

5x faster downloading for cophub

FAST SCALABLE INTERFACES

« Îles de la madeleine data recovering with satellite Sentinel-2 since 2016 to nowadays»



RESULTS

- 5x faster download
- Download without interruption
- 400 Go data downloaded within 12h instead of 3 days

CONCLUSION

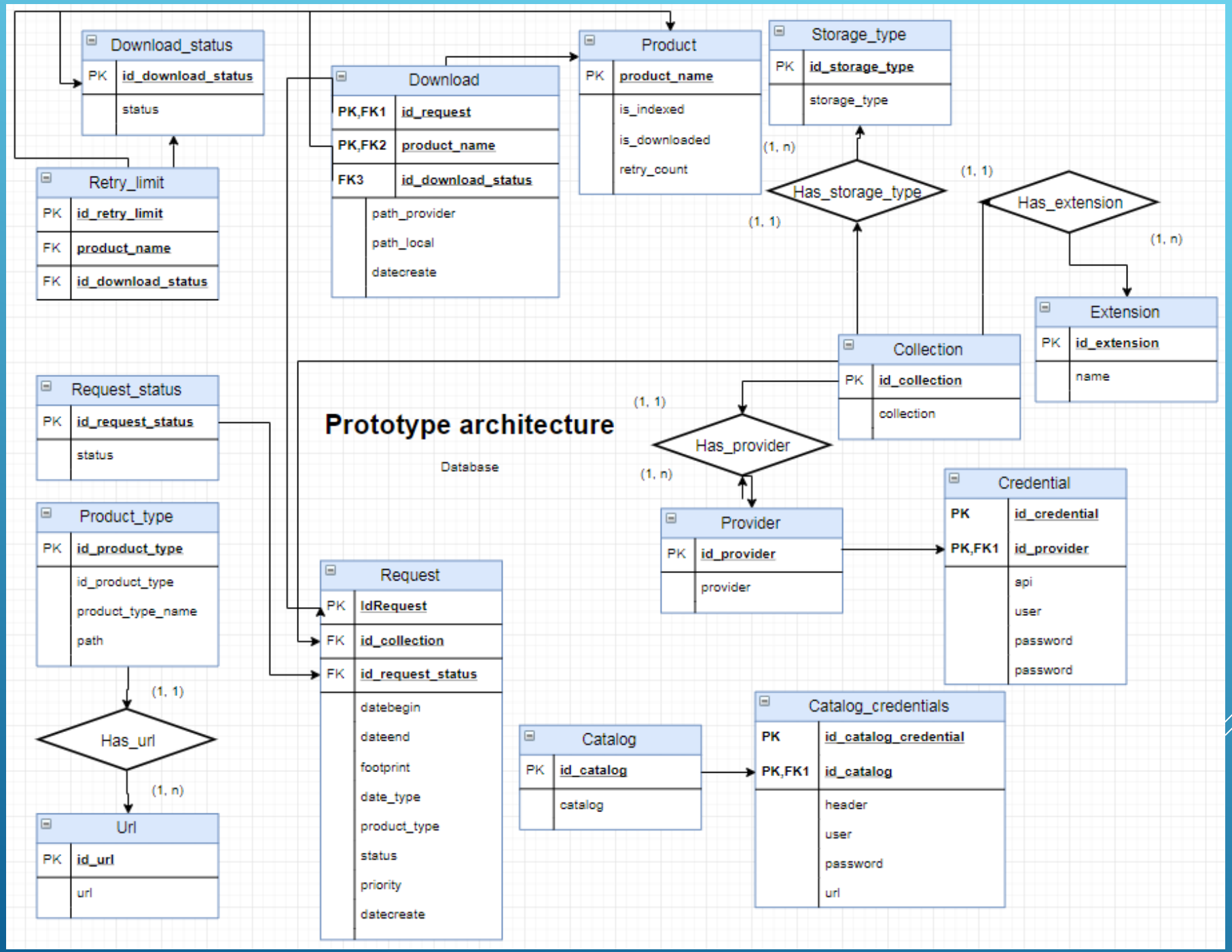
- Creation of the data manager software
  - Knowledge of the satellite data flow
  - Provider (for example, scihub) => Service (Data manager) => Application (detection, map)
- Technologies transferable to other domains

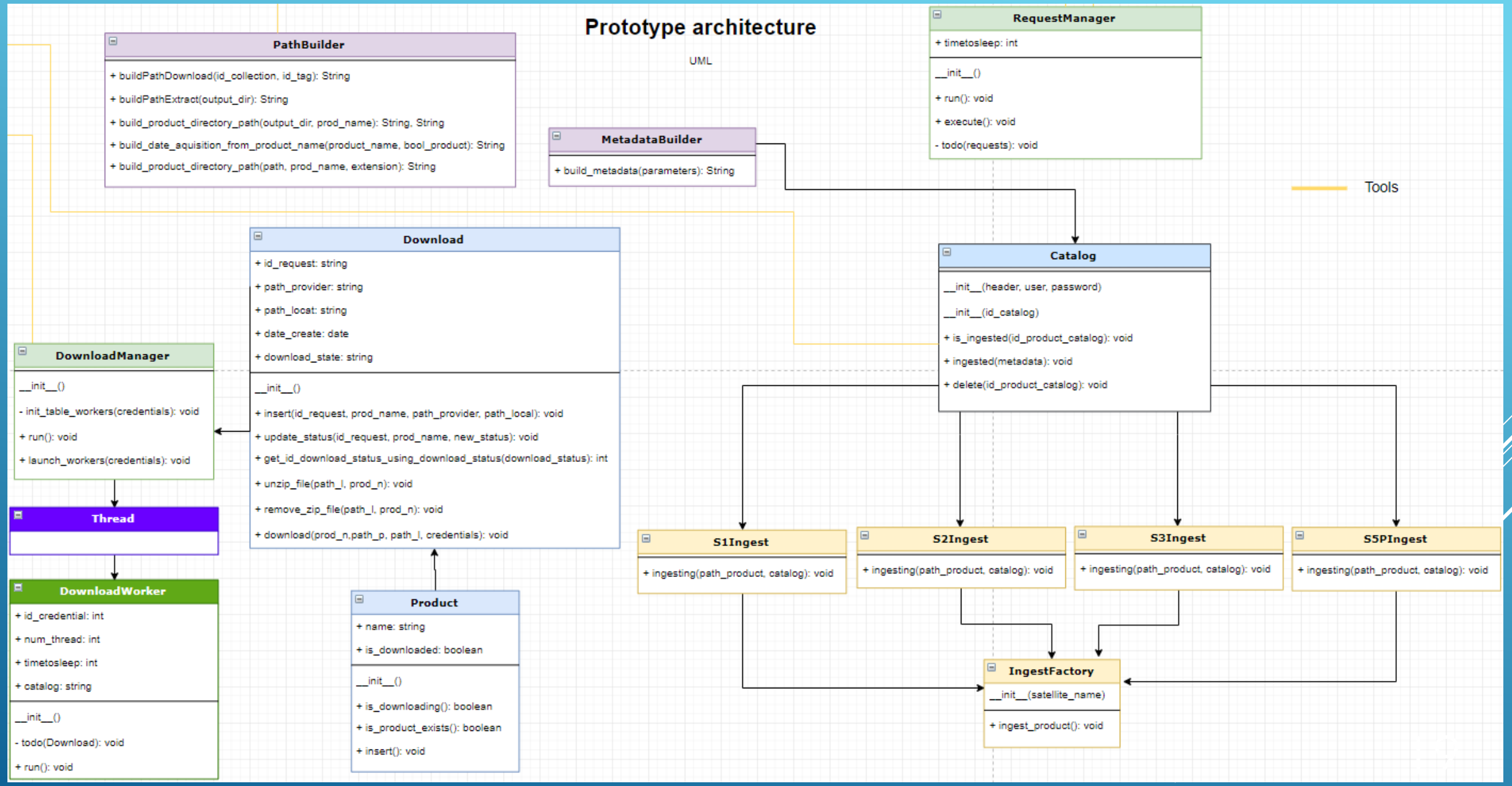
- The next steps
  - Broadcast on other services
  - Create visualization tools

- Johnny NGUYEN
- M2 SSTIM

THANK YOU FOR YOUR ATTENTION

APPENDIX


Prototype architecture
Database
Download_status
PK id_download_status
status
Retry_limit
PK id_retry_limit
FK product_name
FK id_download_status
Download
PK,FK1 id_request
PK,FK2 product_name
FK3 id_download_status
path_provider
path_local
datecreate
Product
PK product_name
is_indexed
is_downloaded
retry_count
Storage_type
PK id_storage_type
storage_type
(1, n)
Has_storage_type
(1, 1)
(1, 1)
Has_extension
(1, n)
Extension
PK id_extension
name
Collection
PK id_collection
collection
(1, 1)
Has_provider
(1, n)
Provider
PK id_provider
provider
Credential
PK id_credential
PK,FK1 id_provider
api
user
password
password
Request_status
PK id_request_status
status
Product_type
PK id_product_type
id_product_type
product_type_name
path
(1, 1)
Has_url
(1, n)
Url
PK id_url
url
Request
PK IdRequest
FK id_collection
FK id_request_status
datebegin
dateend
footprint
date_type
product_type
status
priority
datecreate
Catalog
PK id_catalog
catalog
Catalog_credentials
PK id_catalog_credential
PK,FK1 id_catalog
header
user
password
url


Prototype architecture
UML
PathBuilder
+ buildPathDownload(id_collection, id_tag): String
+ buildPathExtract(output_dir): String
+ build_product_directory_path(output_dir, prod_name): String, String
+ build_date_aquisition_from_product_name(product_name, bool_product): String
+ build_product_directory_path(path, prod_name, extension): String
MetadataBuilder
+ build_metadata(parameters): String
RequestManager
+ timetosleep: int
__init__()
+ run(): void
+ execute(): void
- todo(requests): void
Tools
Download
+ id_request: string
+ path_provider: string
+ path_locat: string
+ date_create: date
+ download_state: string
__init__()
+ insert(id_request, prod_name, path_provider, path_local): void
+ update_status(id_request, prod_name, new_status): void
+ get_id_download_status_using_download_status(download_status): int
+ unzip_file(path_l, prod_n): void
+ remove_zip_file(path_l, prod_n): void
+ download(prod_n,path_p, path_l, credentials): void
Catalog
__init__(header, user, password)
__init__(id_catalog)
+ is_ingested(id_product_catalog): void
+ ingested(metadata): void
+ delete(id_product_catalog): void
DownloadManager
__init__()
- init_table_workers(credentials): void
+ run(): void
+ launch_workers(credentials): void
Thread
DownloadWorker
+ id_credential: int
+ num_thread: int
+ timetosleep: int
+ catalog: string
__init__()
- todo(Download): void
+ run(): void
Product
+ name: string
+ is_downloaded: boolean
__init__()
+ is_downloading(): boolean
+ is_product_exists(): boolean
+ insert(): void
S1Ingest
+ ingesting(path_product, catalog): void
S2Ingest
+ ingesting(path_product, catalog): void
S3Ingest
+ ingesting(path_product, catalog): void
S5PIngest
+ ingesting(path_product, catalog): void
IngestFactory
__init__(satellite_name)
+ ingest_product(): void

SOURCES

- Images :
  - https://geospatialmedia.s3.amazonaws.com/wp-content/uploads/2019/07/The-big-picture.jpg
  - https://prd-wret.s3-us-west-2.amazon
  - https://d1.awsstatic.com/Marketplace/solutions-center/icons/data-science-tool-gray.2a0f018c1fac411125a862eea46bb79b97aff3b6.png
  - aws.com/assets/palladium/production/s3fs-public/styles/landing_page_custom_blocks/public/thumbnails/image/dm-data-integration.jpg?itok=990B2w70
  - https://twitter.com/hashtag/sentinel2
  - https://twitter.com/usgslandsat?lang=fr
  - https://www.python.org/static/img/python-logo.png
  - https://scihub.copernicus.eu/twiki/pub/SciHubUserGuide/GraphicalUserInterface/gui-07.jpg
  - https://upload.wikimedia.org/wikipedia/fr/f/fa/Logo_ESA.png
- CTEP :
  - https://www.coastal-tep.eu/portal/wp-content/uploads/2016/11/CTEP-Thematic-Workshop-S.-Clerc.pdf
- Job offer data libarian :
  - https://www.indeed.fr/voir-emploi?cmp=acri-st&t=Data+Librarian&jk=69e071dcaa232aea&q=Data+librarian&vjs=3