

Présentation de la plateforme Copernicus Data Space

connect
by
cnes



 **someware**

Contexte

→ **Projet européen DAS lancé en 2022 :**

- Accès centralisé aux données du programme Copernicus
- Services de traitement/calcul

→ **La plateforme est lancée fin 2022, sous le nom Data Space**

<https://dataspace.copernicus.eu>

→ **La mise à disposition des fonctionnalités/services est réalisée en plusieurs étapes, jusqu'à la fin 2023**

→ **Peu de communication sur l'offre de la plateforme**

→ **Difficile de se faire un avis sur les services proposés** : protocoles, technologies, performances, avantages/inconvénients, comparaison avec d'autres plateformes...

→ **Someware est mandaté en mai 2023 par le CNES pour réaliser une étude de la plateforme**

Expertise sur le cycle de vie

de l'information géographique



→ **Création en 2019**, après 15 ans dans des entreprises du secteur des SIG et de la 3D

→ **5 collaborateurs** à Rennes et Lyon

→ **Expertise sur l'ensemble du cycle de vie des données géographiques** : de leur production (audit, contrôle, traitement) au développement d'applications métier (web, mobile, desktop) permettant de les visualiser, représenter, gérer, aider à la décision.

→ **Intervention sur toutes les phases d'un projet, en autonomie ou en complément d'équipes existantes** : spécifications, étude technique, développement, POC, déploiement, maintenance/support...

→ **Depuis sa création, Someware se spécialise sur la diffusion des données spatiales avec la réalisation pour le CNES d'une étude des plateformes DIAS et l'accompagnement de sociétés ayant recours à données géographiques**

→ **Someware a aussi développé plusieurs sites valorisant des données spatiales et résultats de recherche :**

- [Tropisco.org](https://tropisco.org)
- [VietSCO.org](https://vietSCO.org)
- [Copernicus Regional](https://copernicus-regional.org)
- [Vu de l'espace](https://vu-de-l-espace.org)

Objectifs de l'étude

→ Donner de la lisibilité dans l'offre de la plateforme Data Space

→ Dresser un panel aussi complet que possible de son offre

→ Comparer la plateforme aux autres grandes plateformes de données spatiales

→ Permettre à de potentiels utilisateurs de comprendre l'adéquation de la plateforme à leurs cas d'usage

Méthodologie

→ Une première phase d'étude théorique (bibliographique) visant à recenser les caractéristiques de la plateforme.

→ Un entretien avec un des représentants de la plateforme (société Sinergise, rachetée par Planet en août 2023).

→ Quelques tests de la plateforme, de façon à analyser sa praticabilité

→ L'analyse de plusieurs cas d'usage concrets issus de 4 sociétés accompagnées par Someware en 2022



Historique et ambition du projet Data Space

Les plateformes DIAS

Objectifs et facteurs de succès

→ Le projet DIAS a constitué la première étape d'un chantier qui amène à la création de la plateforme Data Space

→ Lancé en 2017, la Commission Européenne a ainsi souhaité avec le projet DIAS :

- ouvrir l'usage des données du programme Copernicus à un panel nouveau d'entreprises ou organismes, de façon à créer de nouveaux marchés ou dynamiser des domaines existants
- proposer une solution compétitive et souveraine pour l'accès aux données du programme Copernicus, à même de rivaliser avec les plateformes américaines qui se développaient

→ Un appel d'offre est lancé en 2017 pour sélectionner plusieurs consortiums à même de mettre au point :

- une plateforme de stockage et diffusion des données du programme Copernicus,
- ainsi qu'une offre permettant à de futurs utilisateurs de réaliser des calculs sans avoir à investir dans des solutions lourdes

→ Cinq consortiums sont retenus en décembre 2017 pour mettre en oeuvre, de façon concurrente, le cahier des charges : CREODIAS, MUNDI WEB SERVICES, ONDA, SOBLOO, WEKEO

→ Les cinq consortiums délivrent la première version de leur solution aux trimestres Q1/Q2 de 2018, et sont engagés contractuellement avec l'ESA jusqu'en décembre 2020, puis prolongés jusqu'à la fin 2022.

→ Selon l'ESA, le succès des DIAS va être fonction :

- de l'attractivité de leurs services et de leur capacité à proposer des solutions offrant en standard plus de qualité / valeur que celles qui pourraient être développées en interne par leur client
- de la capacité de chaque DIAS à attirer une communauté de fournisseurs de services basés sur chaque plateforme

Les plateformes DIAS

Bilan

→ Les plateformes DIAS ont été une expérience de la Commission Européenne pour développer une offre souveraine et compétitive d'accès aux données du programme Copernicus.

→ L'offre proposée par les DIAS a surtout eu pour plus-value, par rapport à l'existant, de permettre un accès très performant aux données depuis le cloud de chaque plateforme

→ L'offre en termes de calcul est restée très pauvre : difficile par nature à définir car devant répondre à des besoins très variés, les plateformes (et l'ESA) n'ont pas réussi à apporter une réelle plus-value dans le domaine.

→ Sans doute trop ambitieuse au regard de son budget, et peut-être déviée de son intention par du lobbying (au départ, seules deux plateformes DIAS devaient être financées), l'initiative DIAS n'a au final pas réussi à trouver un modèle économique viable et à installer une offre européenne à même de concurrencer les géants du Cloud américains

→ De ce constat est apparu la nécessité, pour la Commission Européenne, de rebondir en finançant la mise au point d'une seule plateforme, Data Space

L'écosystème de plateformes de données spatiales

→ Les plateformes de données spatiales sont des portails web proposant a minima un large catalogue de produits satellitaires gratuits issus de programmes tels que Copernicus ou Landsat, ou payants tels Pléiades

→ Cette offre est généralement complétée par un ensemble d'outils facilitant l'accès à ces données ou la mise en oeuvre de calculs.

→ Ces portails se développent fortement depuis une dizaine d'années, sous l'impulsion notamment des géants du cloud, avec des ambitions et des moyens différents :

- **public cible** : entreprises, communauté scientifique, communauté "métier" spécifique
- **services, performances et accompagnement adaptés à des usages spécifiques**

Quelques plateformes

→ **Open Access Hub / SciHub** : plateforme de l'ESA permettant d'accéder aux données Sentinel

→ **Sentinel Hub** : plateforme de la société Sinergise proposant un catalogue et des outils pour accéder ou traiter des données de différents programmes (Copernicus, Landsat...)

→ **AWS Earth** : offre en ligne de données satellitaires, opérée par différentes sociétés (dont Sinergise) en tirant parti du cloud d'Amazon.

→ **Google Earth Engine** : plateforme basée sur le cloud de Google, proposant un catalogue de données spatiale, et des outils permettant de développer des applications / algorithmes de traitement

→ **Microsoft Planetary Computer** : plateforme basée sur Azure, le cloud de Microsoft, proposant un large catalogue de données spatiales et des outils de traitement



Copernicus Data Space

Copernicus Data Space

Enjeux et ambitions

→ Loin d'être simplement un portail permettant d'accéder aux données spatiales produites par l'Europe, **Copernicus Data Space se veut un écosystème combinant des données et des services pour exploiter ces données**

→ **L'ambition de la plateforme est de devenir incontournable à échelle mondiale pour accéder à des données du programme Copernicus**

→ **Data Space se veut donc un "Super DIAS" avec :**

- **une ambition très similaire aux DIAS,**
- **un financement de l'ESA beaucoup plus solide que celui des DIAS,**
- **un consortium intégrant plusieurs acteurs des DIAS : T-Systems, CloudFerro, Sinergise, VITO, DLR, ACRI-ST et RHEA**
- **une multitude d'APIs et services pour accéder aux données ou les traiter**

→ **Il apparaît donc plus que nécessaire de donner de la visibilité sur l'offre très large de Data Space.**

Copernicus Data Space

Données

Catalogue de données

→ L'ambition de Data Space est de fournir le plus vaste catalogue de données Copernicus disponibles en téléchargement direct en ligne.

<https://documentation.dataspace.copernicus.eu/Data.html>

→ Le catalogue comprend à la fois :

- **des données gratuites telles que les données Sentinel, Copernicus Core Services, Landsat ou quelques DEM.** Leur accès est possible gratuitement en respectant les quotas gratuits de la plateforme.
- **des données commerciales ou proposées par un organisme commercial, via la plateforme Creodias ou Sentinel Hub.** Leur accès est assujéti à des coûts d'acquisition des données et/ou coûts d'exploitation des services de chaque plateforme.

→ Pour comprendre comment accéder à chaque type de produits, quelques notions sont indispensables :

- **Archive Status** : format des données. "Packed" si les données sont zippées, "Unpacked" sinon. Certaines données sont disponibles au format COG, ou avec des traitements spécifiques au produit
- **Access Type** :
 - **IAD (Immediately Available Data)** : données disponibles "à chaud", sans temps d'attente
 - **DAD (Deferred Available Data)** : données disponibles en décalé, avec un temps d'attente, autrement dit ces données sont archivées ("à froid") et doivent donc être extraites par la plateforme pour être téléchargeable
- **Spatial Extent** : couverture spatiale proposée par la plateforme (World, Europe)
- **Temporal Extent** : couverture temporelle proposée par la plateforme
- **Available in Ecosystem from** : date de mise en ligne du produit sur la plateforme Data Space

Copernicus Data Space

Données

Localisation des données hébergées

→ Le consortium derrière la plateforme se compose de deux Cloud Providers (Open Telekom Cloud, de T-Systems, et Cloudferro) avec des **datacenters en Allemagne et Pologne**

→ Les données sont hébergées dans ces deux pays, sans réplication sur des clouds sur chaque continent comme peuvent parfois le proposer les grands Cloud Providers

Pour des utilisateurs basés hors d'Europe, les temps de téléchargement des données depuis Data Space sont donc plus longs, ce qui peut s'avérer réellement problématique pour certains cas d'usage

Formats de données

→ **Les API d'accès aux données de la plateforme proposent par défaut les formats proposés par les différents programmes d'observation de la terre, autrement dit le format SAFE pour les données Sentinel (avec de l'imagerie au format JPEG2000)**

Néanmoins, certaines données peuvent parfois être proposées au format COG (Cloud Optimized GeoTIFF)

→ **Certaines APIs spécifiques peuvent aussi proposer d'autres formats**, comme l'API Sentinel Hub WMS qui propose des formats web (PNG, JPEG, TIFF, Shapefile, GeoJSON)

Copernicus Data Space

Données

Délai de disponibilité des données

→ L'intégration des nouveaux produits dans Copernicus Data Space se fait à partir de SciHub puis, à partir de 2024, sera réalisée à partir d'autres portails propres à chaque type de produit.

→ **Les temps d'intégration de chaque produit dans Copernicus Data Space varient de quelques heures à quelques jours selon le type de produit.**

Ces temps ne sont pas aujourd'hui communiqués officiellement. Ils sont en moyenne de 5 heures.

Données calculées

→ **La plateforme ne propose pas de données calculées spécifiques, hormis celles des Copernicus Core Services ou des données telles que Sentinel 2 L2A.**

→ **Des produits type NDVI avec une résolution de 10 mètres ne sont donc pas disponibles en téléchargement.**

Néanmoins, les APIs de calcul de la plateforme permettent de les calculer.

Copernicus Data Space

Données

API de catalogage et téléchargement de données

→ Les APIs de catalogage sont les suivantes:

API OData : basée sur le protocole OData, cette API permet de requêter l'ensemble des produits (sauf certaines collections uniquement disponibles via Sentinel Hub), qu'ils soient disponibles à chaud/froid, avec différents filtres possibles

API STAC : encore en beta lors de la réalisation de l'étude, cette API conforme au standard STAC permet de rechercher les produits disponibles à chaud

API OpenSearch : basée sur le protocole standard OGC OpenSearch, cette API propose un modèle de données basé sur GeoJSON et permet de filtrer selon quelques critères.

API Sentinel Hub Catalogue : aussi basée sur le standard STAC, cette API permet de rechercher les données disponibles dans la plateforme Sentinel Hub.

→ Pour le téléchargement de données, la principale API est la suivante:

API S3 : dédiée au téléchargement de la majorité des données proposées par la plateforme.

Autrement dit, il est nécessaire d'utiliser une des API ci-dessus pour rechercher le catalogue et identifier les données à télécharger avec leur adresse S3. Puis celles-ci peuvent être téléchargées via l'API S3

→ Pour accéder aux données via des protocoles standards OGC, comme **WMS, WMTS, WFS, WCS**, la seule solution est l'**API Sentinel Hub**, solution commerciale intégrée à l'écosystème Data Space

Cette API est à la fois une API de téléchargement et une API permettant de réaliser des calculs simples sur les données

Copernicus Data Space

Services de calcul

API de calcul

→ Deux APIs de calcul :

APIs Sentinel Hub : APIs HTTP permettant de calculer des statistiques, ou d'appliquer des fonctions de calcul simples (scripts Javascript) sur chaque pixel d'une image en sélectionnant les bandes souhaitées.

API openEO : permet d'écrire des traitements simples ou complexes sous la forme de workflows dans des langages tels que Python, R ou Javascript. Ces workflows sont ensuite exécutés par la plateforme, et permettent d'exporter les résultats dans des formats classiques dans le monde de la Data Science (GeoTIFF, GeoJSON, CSV, netCDF)...

L'API openEO peut être paramétrée pour déclencher un calcul dès la parution d'un nouveau jeu de données. En plus de l'API, openEO propose un éditeur web ainsi qu'une base d'algorithmes/scripts publics.

→ Service de calcul On-Demand Processing

Le service de **Production à la demande ('On-Demand Processing' ou ODP)** est une API employée par la plateforme pour produire des niveaux 1, 2 ou 3 à partir de niveaux supérieurs.

Cette API sera proposée aux utilisateurs en octobre 2023 pour leur permettre de spécifier les réglages fins de chaque service de calcul, de façon à produire des niveaux 1/2/3 répondant à leurs besoins

Copernicus Data Space

Offre cloud

CREODIAS

→ La plateforme propose une offre cloud, sous le nom CREODIAS (rebranding de l'ancien DIAS du même nom), qui correspond en premier lieu aux offres combinées d'Open Telekom Cloud (T-Systems) et Cloudferro.

Autrement dit, chaque société propose ses services cloud, similaires à d'autres opérateurs cloud tels qu'OVH, Amazon, Google... avec des services standards (machines virtuelles, S3/Object Storage...), quelques services propres et sa tarification.

→ Pour les utilisateurs, l'intérêt de cette offre cloud est de proposer :

- des téléchargements illimités réalisés depuis une VM d'un des clouds
- la possibilité d'exploiter des services tels qu'openEO, Sentinel Hub, Sen4CAP ainsi que des services de génération de produits préconfigurés (sous le nom PGaaS)
- des crédits et une offre packagée, plutôt que des coûts d'utilisation fonction de l'usage et de l'évolution des prix, permettant à des organismes publics (peu flexibles en termes de budget) de s'appuyer plus facilement sur la plateforme

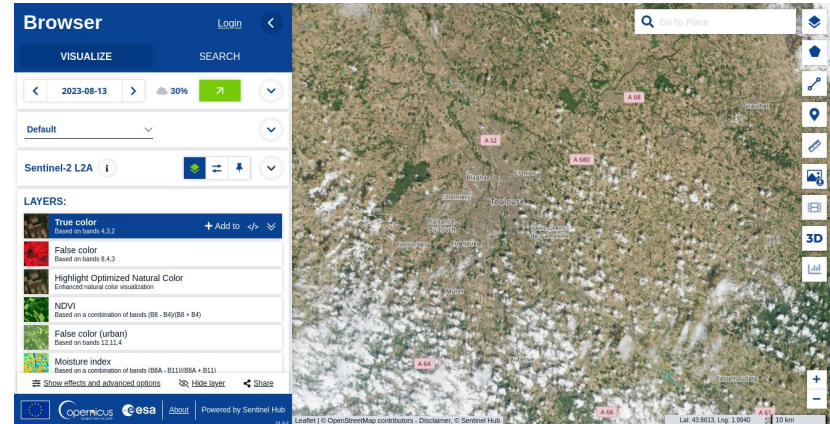
Copernicus Data Space Applications

→ La plateforme propose trois types d'applications web pour explorer ses données et tester ses services:

Browser : catalogue web permettant de rechercher des données à chaud ou froid, et de les consulter dans une interface de WebMapping. Les données peuvent ensuite être ajoutées au Data Workspace, qui fait office de "Favoris".

Data Workspace : application web permettant de consulter son catalogue de données online/offline (IAD / DAD) et de commander des données calculées, qui peuvent ensuite être téléchargées.

JupyterHub : environnement interactif très populaire dans l'univers de la Data Science, avec les Jupyter Notebooks. La plateforme permet à chaque utilisateur de disposer d'un container (équivalent d'une machine virtuelle, avec des limites en termes de CPU/RAM) pour accéder aux données et développer des traitements.



→ Enfin, comme toute plateforme de ce type, Data Space proposera une **Market Place** pour mettre en avant les applications s'appuyant sur ses données et services, de façon à montrer à quel point elle peut être considérée comme fiable.

Copernicus Data Space

Autres services

Bring your own data

→ Prévu pour juillet 2023, ce service doit permettre aux utilisateurs d'intégrer leurs propres données dans les services de catalogage, téléchargement et calcul de la plateforme

API de traçabilité

→ Cette API permet à un utilisateur d'obtenir des informations sur l'ensemble du cycle de vie d'un produit, comprendre les traitements qui ont été appliqués et données source employées

Copernicus Data Space

Tarification et quotas

→ La philosophie générale est de proposer des données et services gratuits, dans la limite du respect de certains quotas

Le catalogue comprend à la fois des données gratuites telles que les données Sentinel, Copernicus Core Services, Landsat ou quelques DEM. Leur accès est possible gratuitement en respectant les quotas gratuits de la plateforme.

Il comprend aussi des données commerciales ou proposées par un organisme commercial, via la plateforme Creodias ou Sentinel Hub par exemple. Dans ce cas, l'accès comprend des coûts d'acquisition des données et/ou des coûts d'exploitation des services de chaque plateforme.

→ Plusieurs types d'utilisateurs sont gérés par la plateforme avec, pour chaque type, des quotas différents:

Utilisateurs généraux de Copernicus : utilisateurs enregistrés sur la base d'un auto-enregistrement

Utilisateurs Copernicus Services : utilisateurs issus d'institutions et organes de l'Union européenne, autorités publiques nationales ou régionales de l'UE ou d'un pays tiers

Partenaires du segment terrestre collaboratif de Copernicus : utilisateurs comprenant les États participant au programme Copernicus, après la signature d'un accord avec l'ESA et la Commission européenne

Partenaires internationaux de Copernicus : typologie d'utilisateurs ouverte aux partenaires internationaux, après signature d'un accord de coopération avec la Commission européenne et d'un accord d'exploitation avec l'ESA.

Copernicus Data Space

Tarification et quotas

Description des tarifs / quotas

https://documentation.dataspace.copernicus.eu/_doc/CDSE-SDE-TSY_Service%20Description%20and%20Evolution_1.1.pdf

→ Pour les utilisateurs généraux, les quotas sont les suivants :

Application / API	Quotas / Prix
Copernicus Browser	Recherche illimitée, consultation / téléchargement / export assujettis aux quotas de l'API Sentinel Hub
JupyterHub	Chaque utilisateur dispose d'un container avec 2 vCPU et 8 Go de RAM
APIs de catalogue	Aucune limite, hormis quelques sécurités sur le nombre max de requêtes/minute pour éviter une saturation des services
APIs de téléchargement	Aucune limite si les téléchargements sont réalisés depuis l'un des clouds de la plateforme (nécessite la location d'a minima une machine virtuelle). Si les téléchargements sont réalisés depuis l'extérieur, il existe des limites, notamment sur la bande passante (max 20 Mo/s), sur le nombre de connexions en parallèle (4) et le total téléchargé / mois (6 To). Pour l'accès aux COG via HTTP, la limite est de 50K requêtes/mois.
APIs Sentinel Hub	Pour calculer des données, les coûts se basent sur la notion d'unité de calcul. Ex: pour des produits calculés tels que Sentinel 2 NDVI ou True Color, les coûts sont au km ² traité, entre 0.00000004 et 0.00025 €/km ² . Les quotas d'accès gratuits sont de 10K requêtes/mois, 300 par minute, 10000 unités de calcul/mois et 300 unités de calcul/minute
API openEO	Chaque utilisateur reçoit 1000 crédits à la création de son compte
Production à la demande	25 produits calculés par mois, 2 en parallèle

→ Autrement dit, la plateforme propose une offre gratuite ("Free tier") permettant d'évaluer son offre et de tester ses APIs.
Pour certains utilisateurs, cette offre gratuite peut être suffisante.

→ A partir d'un usage dépassant les quotas de la plateforme, il n'existe à ce jour pas de prix publics pour l'usage de chaque type de services (hormis pour les services commerciaux tels que Creodias ou Sentinel Hub).

Par exemple, pour télécharger des volumes de données de plus de 6 To / mois, ou avec une bande passante de plus de 20 Mo/s, via l'API S3, aucun prix public n'est annoncé.

Dans ce cas, il est nécessaire de contacter les représentants de la plateforme.

→ Pour d'autres types d'utilisateurs, en particulier pour la communauté scientifique, il est nécessaire de se référer aux documentations en ligne de la plateforme pour connaître les quotas en vigueur.



Test de la plateforme

Test de la plateforme

→ Les tests de la plateforme ont porté exclusivement sur l'utilisation du catalogue (via l'application Browser et quelques APIs) et le téléchargement de données

→ **Semblant assez simple de prime abord, l'utilisation de la plateforme s'est avérée plus complexe en pratique, avec une documentation succincte et l'absence de tutoriels.**

→ L'utilisation de la plateforme a d'emblée buté sur quelques problèmes :

- le non-fonctionnement (lors du premier test) du Browser, pourtant le premier outil employé par les utilisateurs d'ordinaire
- l'API STAC peu ou pas opérationnelle

Néanmoins l'API OData a parfaitement fonctionné, tout comme le téléchargement de données.

Le browser s'est avéré très pratique et ergonomique à partir du moment où il s'est mis à fonctionner.

→ **Il est clair, à la date du test (août 2023), que certains des services de la plateforme sont encore peu utilisés et pas encore très robustes.**

→ **Le manque de tutoriels, et de documentation pratique (par exemple pour l'API STAC), est révélateur d'une plateforme encore en plein développement, avec peu de temps consacré à la documentation puisque les services ne sont pas encore pleinement opérationnels.**



Comparatif avec d'autres plateformes

Comparatif avec d'autres plateformes

→ **Qu'est-ce que Data Space propose de plus ou de moins que les autres plateformes, dont certaines sont aujourd'hui très utilisées ? Est-ce que Data Space est plus ou moins cher, plus ou moins efficace ?**

→ **Il est difficile d'apporter une réponse simple à ces questions sans étudier des cas d'usage spécifiques. La présente étude n'a pas vocation à réaliser de tels comparatifs.**

→ Néanmoins, l'ambition et les fonctionnalités de l'écosystème Data Space permettent de lister quelques caractéristiques différenciant la plateforme de ses concurrentes, attestées via un entretien avec un de ses représentants :

- **Data Space va proposer la plus large collection de données Sentinel, avec une couverture spatiale et temporelle totale pour la plupart des produits**
- **Data Space aura aussi les données les plus récentes, et servira d'ailleurs à alimenter ensuite les autres plateformes**
- **Data Space intègre Sentinel Hub, l'offre de SciHub et la quasi totalité de l'offre des DIAS**

→ **Alors que l'ESA misait il y a quelques années sur une offre basée sur plusieurs plateformes, telles que SciHub et les DIAS, l'ambition est donc désormais de proposer un point d'accès unique pour les données Copernicus avec Data Space**

Comparatif avec d'autres plateformes

Open Access Hub / SciHub

→ Jusqu'à présent le portail de référence pour accéder aux données Copernicus, SciHub va disparaître fin 2023 pour que Data Space soit le seul point d'accès européen.

→ Data Space propose en outre des services de calculs, ce que SciHub n'a jamais proposé.

Comparatif avec d'autres plateformes

Sentinel Hub

→ Sentinel Hub est désormais intégré à l'écosystème Data Space, bien qu'il s'agisse d'une brique commerciale opérée par Sinergise et VITO

→ Sentinel Hub va donc jouer un rôle central dans Data Space, et gagne ainsi en importance et longévité.

→ Cela signifie au passage que les architectes logiciels ayant conçu des solutions basées sur Sentinel Hub vont non seulement pouvoir continuer de les exploiter, mais vont aussi bénéficier de toute l'offre Data Space.

Comparatif avec d'autres plateformes

Amazon Web Services (AWS)

→ L'offre Amazon est exclusivement centrée sur la mise à disposition des services du cloud AWS pour répondre à des besoins propres au domaine du traitement de données spatiales.

→ Développés et administrés par des organismes/sociétés très divers, ces jeux de données et services de catalogage ont juste pour point commun d'être hébergés chez AWS.

→ Il est donc difficile de considérer cette offre comme stable et pérenne, puisqu'il n'existe aucun pilotage stratégique clair et affirmé de la part d'Amazon.

→ Chaque société proposant un service, que ce soit un catalogue de données ou des mécanismes de calcul, est libre de fixer ses conditions d'accès et prix.

→ La seule véritable action d'Amazon a été de regrouper ces offres sur une seule et même Market Place de façon à donner de la visibilité aux différents services et applications tirant parti d'AWS dans le domaine du spatial.

→ Cette offre déjà ancienne a pour intérêt de s'appuyer sur le cloud le plus populaire, pour lequel de très nombreux architectes et développeurs logiciel sont formés. Néanmoins, elle se compose d'une myriade de services sans recherche de cohérence globale et de pérennité, à l'inverse de Data Space.

→ A l'avenir, pour l'accès aux données Copernicus, Data Space devrait devenir la solution de référence.

Il est cependant possible que certaines sociétés développent des répliques d'une partie du catalogue de Data Space sur des datacenters AWS localisés sur chaque continent, de façon à permettre un accès performant aux données à des organismes situés hors d'Europe.

Comparatif avec d'autres plateformes

Google Earth Engine (GEE)

→ La particularité de l'environnement GEE est de fournir une expérience utilisateur/développeur complètement intégrée dans l'environnement technique Google pour visualiser et traiter des données spatiales.

→ Basée sur assez peu de standards (hormis S3 pour l'hébergement/téléchargement de données), l'objectif de la plateforme est de proposer un environnement facile à utiliser (via des outils propriétaires Google) pour rechercher des données, les visualiser et développer des traitements.

→ Plus qu'un espace de stockage, GEE est donc plutôt un environnement de conception de traitements, très efficace et simple d'utilisation.

→ Dans l'esprit des environnements propriétaires intégrés typiques de Microsoft et ESRI, le gain en simplicité d'utilisation s'accompagne de coûts généralement élevés et de possibilités d'optimisation/extension très réduites.

→ Au sein de l'écosystème Data Space, l'outil le plus proche est sans doute openEO, avec son éditeur web et sa logique de scripts de traitements.

Parmi les différences notables, il y a le support du langage R côté openEO (seulement Javascript et Python côté GEE), mais cette comparaison n'est pas suffisante car elle n'aborde pas dans le détail les fonctionnalités/technologies disponibles pour chaque langage.

D'autre part, openEO propose une extension GEE, de façon à appeler le moteur Google pour accéder à des données et réaliser des calculs. Son ambition semble donc supérieure à celle de GEE.

Il sera intéressant à l'avenir de comparer GEE et openEO pour comprendre plus finement les bénéfices et inconvénients de chacun.

→ Enfin, tout comme avec AWS, il est possible que Google capitalise à l'avenir sur des répliques d'une partie du catalogue de Data Space sur des datacenter Google Cloud localisés sur chaque continent, de façon à permettre un accès performant aux données à des organismes situés hors d'Europe.

Comparatif avec d'autres plateformes

Microsoft Planetary Computer

→ Basée sur différents standards et briques répandues (S3, STAC, JupyterHub), Microsoft a conçu cette plateforme, hébergée sur son cloud Azure, pour proposer une expérience intégrée permettant de rechercher des données, les visualiser, les traiter et les intégrer à des applications

→ Cette offre est donc assez proche dans l'esprit de celle de Data Space, qui se compose aussi de briques open source et propriétaires pour proposer l'ensemble des fonctionnalités de catalogage, visualisation et calcul.

→ Avec un catalogue administré en partie par Microsoft, une stratégie claire et ambitieuse (contrairement à Amazon) et l'emploi de standards (contrairement à Google), Microsoft Planetary Computer est un bon challenger pour Data Space et dispose de datacenters répartis dans le monde entier pour proposer ses services (contrairement à Data Space).

→ Néanmoins, pour la fourniture de données Copernicus, Microsoft s'est appuyé, comme Amazon, sur Sinergise pour l'administration des jeux de données, qui offrent aujourd'hui une couverture spatiale et temporelle moins bonne que Data Space.

Comparatif avec d'autres plateformes

Les DIAS

→ La véritable différence entre Data Space et les DIAS réside dans :

- **la couverture spatiale/temporelle des données Copernicus proposées**, quasi totale pour Data Space alors que les DIAS étaient contraintes de proposer une petite partie du catalogue pour des raisons budgétaires
- **la mise à disposition d'APIs de calculs**, ce qu'aucune plateforme DIAS ne proposait et qui était très peu clair dans le cahier des charges de l'ESA
- **un financement par la Commission Européenne / ESA garanti jusqu'en 2028**, avec une extension possible jusqu'en 2032

→ Autrement dit, Data Space comble les trois lacunes des DIAS :

- **le financement court termiste des DIAS a freiné bon nombre d'utilisateurs potentiels**, peu enclins à investir de leurs temps et capitaliser sur une solution dont la pérennité semblait hypothétique
- **la couverture spatiale et temporelle des données Copernicus**, quasi totale, rend possible le développement de nombreux services basés sur de fortes profondeurs temporelles et des données hors d'Europe
- **proposer des APIs et un environnement de calculs permet à des développeurs de se focaliser sur leur métier**, plutôt que devoir mettre au point, installer et administrer une infrastructure de calcul comme c'était le cas avec les DIAS

→ **Les DIAS sont donc bel et bien morts (sauf deux d'entre eux, décrits ci-dessous)**, moins par leur faute que par un manque de vision de la Commission Européenne et l'ESA quant à la manière de développer un écosystème souverain de services d'accès et traitement de données spatiales.

→ **Néanmoins, deux DIAS vont continuer d'exister :**

WEKEO : considéré dès le départ comme un DIAS particulier, avec un cahier des charges allégé par rapport à d'autres DIAS, Wekeo se focalise sur les données environnementales, avec une clientèle identifiée dès le début sur ce sujet (monde académique en premier lieu), et va poursuivre ses activités de façon indépendante

CREODIAS : désormais partie prenante de Data Space, en tant que branche commerciale fournissant des services cloud, il ne reste véritablement de cet ancien DIAS que le Cloud Provider CloudFerro (leader du DIAS CREODIAS), auquel est désormais associé l'autre Cloud Provider Open Telekom Cloud (T-Systems).



Etude de cas d'usage

Etude de cas d'usage

Contexte

→ En 2022, dans le cadre d'action DIAS Priming (financée par le CNES), SOMEWARE a été amené à conseiller quatre entreprises, sélectionnées par le CNES, souhaitant implanter leurs chaînes de traitements sur les DIAS

→ A partir des besoins propres à chaque entreprise, Someware a cherché à identifier les facteurs clé déterminant le choix d'une plateforme en fonction d'un cas d'usage : nature des données nécessaires, performances de téléchargement / calcul attendues, caractéristiques de l'environnement cloud...

→ Ensuite, Someware a réalisé une étude comparative des plateformes DIAS au regard de ces facteurs, en cherchant à obtenir auprès de chaque plateforme les informations nécessaires à l'identification de la ou des plateformes les plus adaptée(s) à chaque cas d'usage.

→ Les slides ci-après proposent une réflexion sur l'adéquation de la plateforme Data Space aux cas d'usage de ces quatre entreprises.

Etude de cas d'usage

Société A

Contexte/besoins

- La société A met au point des produits satellitaires à échelle mondiale, et se spécialise sur le traitement des couleurs.
- Elle développe une chaîne de traitement combinant les outils Maja/Wasp ainsi que Sen2core pour produire un niveau L3A mondial. En bout de chaîne, certains traitements colorimétriques nécessitent l'intervention d'un opérateur.
- Les difficultés qu'elle rencontre aujourd'hui sont liées aux temps de téléchargement / traitement longs des données pour une telle échelle de traitement.
- D'autre part, elle ne maîtrise pas le déploiement de calculs dans le cloud.

Solutions possibles avec Data Space

- Les volumes de données téléchargés par la société A, ainsi que la bande passante attendue, dépassent très certainement les quotas de la plateforme. Il est sans doute souhaitable pour cette société de contacter le service commercial afin de connaître les prix pour son usage.
- Elle peut aussi se pencher sur le déploiement cloud de ses calculs, de façon à bénéficier de quotas illimités de téléchargement quand les calculs sont déployés sur une des VMs des deux clouds proposés par Data Space. L'essentiel des calculs pourraient ainsi être réalisés sur des VMs, puis les données résultantes pourraient être rapatriées en local pour réaliser les derniers traitements colorimétriques.

Etude de cas d'usage

Société B

Contexte/besoins

→ La société B met au point des outils d'aide à la décision basés sur la télédétection à destination des territoires insulaires du Pacifique Sud (où la société est basée).

→ Elle a pour projet de mettre au point et diffuser des données L2A (outils Maja / Sen2core) à l'échelle du Pacifique Sud.
D'autre part, elle souhaite accéder aux données Sentinel-1 dès leur mise en ligne pour un projet de surveillance maritime.

→ De façon générale, son besoin est de pouvoir accéder à des produits bruts récents ou données traitées à l'échelle du Pacifique Sud, avec de bons temps de téléchargement.

→ Elle dispose de compétences en déploiement de calculs dans un data center local.

Solutions possibles avec Data Space

→ Pour cette société basée en Océanie, Data Space ne propose pas de nouveautés par rapport aux DIAS

Les données de Data Space sont hébergées en Allemagne et Pologne, sans réplication sur d'autres clouds répartis dans le monde, permettant un temps d'accès très performants pour des sociétés basées hors d'Europe.

→ Cette société devrait peut-être étudier la location d'une VM dans un des clouds de Data Space, pour télécharger des données sans limite, les traiter puis rapatrier seulement sur son DataCenter les résultats (encodés de la façon la plus efficace possible pour réduire les temps de téléchargement).

Etude de cas d'usage

Société C

Contexte/besoins

- La société C met au point un service d'analyse géographique, essentiellement en Afrique, à partir d'imagerie Sentinel-1 et Sentinel-2.
- Elle maîtrise l'automatisation ainsi que la conception d'algorithmes de traitement, mais pas le déploiement de calculs dans le cloud.
- Elle souhaite globalement optimiser ses temps de téléchargement / traitement et, si possible, pouvoir déclencher ses traitements dès qu'une donnée est disponible (notion de "triggers").
- Elle cherche aussi à accéder à des données sur de vastes zones, avec une forte profondeur temporelle.

Solutions possibles avec Data Space

- La plateforme propose l'ensemble du catalogue Sentinel-1 et Sentinel-2 en accès immédiat (IAD), autrement dit accéder à des données sur l'Afrique, avec une forte profondeur temporelle, est aisé.
- Pour disposer de temps de téléchargement très rapides, sans limite de taille, la solution sans doute la moins coûteuse est de louer une VM dans le cloud d'une des plateformes avec, dans ce cas, des quotas de téléchargement illimités.
- La plateforme ne propose pas de mécanisme de type "trigger" mais prévoit d'en proposer à l'avenir. Elle permet néanmoins, via l'API openEO, de développer des traitements qui se déclenchent dès la parution d'un nouveau jeu de données. Il s'agit sans doute d'une bonne réponse aux besoins de la société C.

Etude de cas d'usage

Société D

Contexte/besoins

- La société D développe des solutions de cartographie multi-source, basée notamment sur des données spatiales.
- Elle exploite notamment des données Sentinel-1, Sentinel-2 et Pléiades Neo en Europe et Afrique.
- Son usage le plus courant est le chargement de données en fonds de plan (tuiles / WMTS) dans une interface de WebMapping.

Solutions possibles avec Data Space

- Pour cette société, Data Space propose les APIs Sentinel Hub (brique pré-existante, faisant néanmoins partie de l'écosystème Data Space). Une des APIs Sentinel Hub permet d'accéder à des données selon le protocole WMTS (utilisé par le Browser de Data Space)



Conclusion



Conclusion

→ Copernicus Data Space s'annonce comme un challenger de poids dans l'écosystème des plateformes de données spatiales.

Avec une ambition, et surtout des moyens, supérieurs à ceux octroyés aux DIAS, l'ESA fait le pari d'une plateforme de référence pour les données Copernicus, plutôt que plusieurs petites plateformes en compétition.

Avec un catalogue et des fonctionnalités de calculs rivalisant aisément avec les solutions concurrentes (Google, Amazon, Microsoft), la plateforme devrait vraisemblablement attirer de nombreux usagers.

→ Parmi les freins possibles à son succès, il y a néanmoins :

- **des quotas gratuits, et coûts d'utilisation au-delà, encore assez difficiles à évaluer**, fortement dépendants de chaque cas d'usage. Un travail de documentation et de pédagogie doit être réalisé par l'ESA et le consortium.
- **des temps d'accès aux données potentiellement problématiques hors d'Europe**, rendant la plateforme moyennement efficace pour des usagers d'autres continents (quand le temps de téléchargement est critique pour leurs cas d'usage)

Ces inconvénients seront certainement corrigés à moyen terme.

Conclusion

→ **Le véritable questionnement qui subsiste concerne le modèle économique de la plateforme, entremêlement de services publics (essentiellement gratuits) et de services privés (payants), et ses orientations futures.**

En construisant Copernicus Data Space à partir de nombreuses briques propriétaires, même si éprouvées et de qualité, le consortium a clairement cherché à être incontournable à l'avenir pour opérer et maintenir la plateforme.

En particulier, au sein du consortium, la société Sinergise (acquise par Planet durant l'été 2023) a su se positionner derrière pratiquement toutes les plateformes de données spatiales, et peut donc influencer de nombreuses évolutions en fonction de ses intérêts propres.

→ **A quel point l'ESA souhaitera-t-elle, ou pourra-t-elle, à l'avenir imposer certaines orientations stratégiques au consortium ?**

→ **Comment pourra-t-elle mettre le consortium en compétition dans le cadre de futurs marchés publics ?**

Merci !

Bertrand GERVAIS

bertrand.gervais@someware.fr

+33 6 81 36 81 11

