**Appels à projets Horizon 2020 - Projets retenus et participants identifiés**

Ce jeu de données « fr-esr-projets-participants-h2020 » propose les informations présentées dans scanR sur le financement des projets de recherche par la Commission européenne.

Il liste tous les projets signés et tous les participants français et étrangers.

Il est composé de 37 colonnes :

* 14 variables descriptives du projet
* 10 variables descriptives du participant
* 13 variables de géolocalisation du participant

Dans le tableau ci-dessous le descriptif de chaque variable :

|  |  |
| --- | --- |
| **project\_type** | Le nom du programme-cadre |
| project\_id | Le numéro du projet |
| acronym\_projet | L'acronyme du projet |
| name\_project | Le nom du projet |
| start\_date | Le date de début |
| end\_date | La date de fin |
| budget\_financed | Les subventions allouées au projet par la Commission |
| duration | Durée de réalisation du projet |
| keywords\_en | Mots-clés |
| source\_url | La fiche du projet sur le site web de la Commission |
| call\_code | Le code de l'appel à projet |
| topic\_code | L'objectif de l'appel à projet |
| programme\_abbr | Le code du programme |
| programme\_lib | Le libellé du programme |
| participation\_id | Le numéro de la participation (project\_id + N° d'ordre) |
| participant\_id | L'identifiant du participant expliqué dans la suite du document |
| name | Le nom du participant |
| acronym | L'acronyme ou sigle du participant |
| role | Le role du participant (coordinator / partner) |
| participates\_as | Le type de participation (beneficiary/third party/partner organization/…) |
| co\_participant | Variable expliquée dans la suite du document |
| participant\_type\_code | La catégorie du participant (catégories de la Commission) |
| participant\_type\_name | Le libellé de la catégorie |
| funding | Subventions allouées à chaque participant |
| participant\_id\_type | Le type d'identifiant du participant (Siren, Siret, Grid,…) |
| country\_code | Le code du pays |
| country\_name | Le libellé du pays |
| country\_level\_1 | Regroupement des pays |
| gps | Les coordonnées de géolocalisation |
| com\_code | Le code commune |
| com\_nom | Le libellé du code commune |
| uucr\_id | Le code unité urbaine |
| uucr\_nom | Le libellé de l'unité urbaine |
| dep\_id | Le code du département |
| dep\_nom | Le libellé du département |
| dep\_num\_nom | Le code + libellé du département |
| aca\_id | Le code de l’académie |
| aca\_nom | Le libellé de l’académie |
| reg\_id | Le code région |
| reg\_nom | Le libellé de la région |

Une participation

Le **participation\_id** identifie une participation dans un projet. Cette participation peut être composée de plusieurs acteurs aux statuts différents.

Elle est obligatoirement composée d’un *beneficiary* (**participates\_as**) qui est aussi *participant* (**co\_participant**).

Elle peut être composée de plusieurs acteurs par exemple d’un *beneficiary* et d’un *thirdparty*. Ils se distinguent grâce à la variable **participates\_as.**

Chaque participation a un rôle *coordinator* ou *partner.* Dans les participations à multiples acteurs seul le *beneficiary + participant* est *coordinator.*

Le funding est la subvention allouée par la Commission à chaque participant. Dans le jeu de données, cette subvention est renseignée telle que transmise dans les bases eCorda.

Suite au besoin d’étudier la participation française au niveau des acteurs du programme-cadre de la Commission, le Ministère a développé un processus d’identification des acteurs participants.

Les besoins sont :

* De prendre en compte la spécificité de la recherche universitaire française et d’identifier les acteurs les plus proches de l’activité de recherche,
* De qualifier et classer tous les participants au programme,
* De localiser les activités de recherche sur le territoire à un niveau fin en limitant l’effet de siège.

Ce processus débute avec la collaboration des organismes de recherche (CEA, CNRS, INRAE, INRIA, ONERA). Ils labellisent chacune des participations affiliées à leur structure par un identifiant propre à leur organisme ou bien un numéro RNSR (Répertoire national des structures de recherche). Si nécessaire ils complètent avec une donnée géographique qui aide à atténuer les effets de siège des grandes structures.

La suite de l’identification est réalisée au sein du département des outils d’aide à la décision. Différentes méthodes semi-automatisées ont été mises en place (par exemple : rapprochement de chaînes de caractères avec des annuaires et référentiels disponibles).

Les variables qui sont issues de ces traitements sont les suivantes :

Participant\_id

L’objectif de cette identification est d’attribuer à chaque participant un identifiant appartenant à un référentiel institutionnel auquel nous avons accès. Cela permet de qualifier et de classer les structures.

Dans ces traitements nous nous appuyons sur les sources suivantes :

* [SIRENE](https://public.opendatasoft.com/explore/dataset/sirene_v3/table/?flg=fr&disjunctive.libellecommuneetablissement&disjunctive.naturejuridiqueunitelegale&sort=datederniertraitementetablissement&disjunctive.etatadministratifetablissement&disjunctive.sectionetablissement),
* [RNSR](https://appliweb.dgri.education.fr/rnsr/) Répertoire national des structures de recherche du Ministère et [jeux de données Open-Data](https://data.enseignementsup-recherche.gouv.fr/pages/explorer/?q=%22repertoire%20national%22),
* La [plate-forme Open-Data du Ministère](https://data.enseignementsup-recherche.gouv.fr/pages/explorer/?q=%22%C3%A9tablissements%20d%27enseignement%20superieur%22) pour différents jeux de données sur les établissements,
* [Finess](http://finess.sante.gouv.fr/fininter/jsp/index.jsp) Répertoire de la Santé,
* [RNA](https://www.journal-officiel.gouv.fr/associations/recherche/) Répertoire national des associations,
* [Grid](https://grid.ac/) et [ROR](https://ror.org/facts/) pour les acteurs étrangers.

Le **participant\_id** est typé selon son origine dans **type\_id**, par exemple le 775878390 est typé *siren*.

*Attention : l’identification n’est pas exhaustive.*

Tous les acteurs participants ne sont pas affiliés à une entrée de référentiel. Soit :

* L’information dans les bases eCorda n’est pas suffisante pour être formelle sur l’identité de la structure pour les acteurs français,
* Aucun matching n’a été réalisé entre un acteur participant et une entrée de référentiel,
* Le volume des participants étant conséquent, seuls les acteurs étrangers très présents dans les projets signés sont identifiés pour l’instant.

Co\_participant

La notion de *co-participant* est née suite au besoin d’identifier, pour la France, les acteurs les plus proches de l’activité de recherche notamment les unités de recherche telles les UMR. Il peut aussi s’agir d’autres types de structures comme les CHU de Paris qui sont affiliés à l’AP-HP dans eCorda.

Ces co-participants ne sont pas signataires de la convention de financement avec la Commission européenne et les informations les concernant dans les bases eCorda sont très disparates ou absentes.

A la suite de ces traitements ces acteurs sont classés comme co-participant (**co\_participant**).

Les co-participants ne sont pas typés dans la variable **participant\_type\_code** et **participant\_type\_name**.

Subventions allouées

Deux variables subventions sont disponibles **budget\_financed** et **funding**.

Le **budget\_financed** est le total des subventions allouées par la Commission au projet.

Le **funding** est la ventilation de cette subvention allouée à chaque participant.

Les thirdparties ne sont pas signataires de la convention de financement. Ils ne sont donc pas directement financés par la Commission mais par le bénéficiaire.

Localisation

Pour les participants identifiés comme unités de recherche et affiliés à un numéro RNSR, l’adresse prise en compte est celle de l’unité.

Pour les autres participants, l’adresse prise en compte est celle communiquée dans des tables complémentaires du SI eCorda.

Certaines adresses manquantes ont été renseignées grâce à l’affiliation du participant à un référentiel.

Pour certains participants étrangers les coordonnées de géolocalisation ont été ajoutées au niveau de la ville.