FILLING GUIDE FOR PROJECT LEADERS

Fee Calculation Grid for Databases Registered in the HDH Data Warehouse

1 - Introduction

As part of the implementation of the HDH Data Warehouse (ex SNDS Catalog), the harmonization of fee grids for data access by private actors is essential. The objective is to meet a growing demand for transparency and consistency in pricing practices, while facilitating access to health data for private stakeholders, particularly within the framework of health research. This approach aligns with the orientations set by the Strategic Health Data Committee (Costrat) and builds upon the work initiated by the Financing Working Group (GT Financement) in 2023. A working group, composed of the data holders listed in the decree of May 22, 2022, met three times in 2024 to highlight the specificities of making data available via the HDH Data Warehouse and the necessary adaptations of the associated pricing principles.

A fee calculation grid for the SNDS Data Warehouse databases was therefore developed based on the two pricing grid models proposed by the Strategic Health Data Committee following the work of the Financing Working Group. This grid follows these principles: transparency, price accessibility, scope of activities covered by fees, valuation of activities related to data updates, equity, flexibility and adaptation to types of actors, as well as operational efficiency. It enables the calculation of fees in compliance with the framework provided by the Regulation of May 30, 2022, on European data governance (DGA) and Article L. 1461-5 of the Public Health Code, which provides for cases where access to SNDS data is free of charge.

This grid offers the user an *a priori* modeling of the workload associated with the various tasks of providing data and project support, based on the characteristics of the project as entered by the user. Indeed, it associates each task with one or more profiles and a workload in a standardized manner.

The purpose of this filing guide is to assist project leaders in completing their sections of the fee calculation grid.

The tabs in the document are presented in the order in which they must be completed.

2 - Application Procedures

This grid is used when a project leader submits a request for access to a data holder's database via the HDH Data Warehouse.
This grid is provided by the HDH and must be completed according to the following steps:

1. The data holder completes:
    a. The "Requirements" tab: the purple cells.
    b. The "Project Steps" tab: the purple cells.
        i. Column L to indicate the type of service the data holder offers to the project leader.
        ii. Column M to indicate the phases during which the data holder contributes to making the data available.
2. The data holder and the HDH agree on the grid completed by the data holder, which is then annexed to the partnership agreement signed between the HDH and the data

holder.

3. The cost calculation grid specific to each database is shared with all project leaders requesting access to the database in question.
4. The project leader completes:
    a. The "Requirements" tab: the yellow cells.
    b. The "Project Steps" tab: the yellow cells in column H.
    c. The "WP xx" tabs: the yellow cells in column K.

    The project leader sends the grid to the HDH with the completed "Requirements" tab and the associated study protocol, if available.
5. The HDH completes the red cells in the "Requirements" tab.
6. Finalization of the cost estimate with the data holder.
7. The cost estimate is sent to the project leader.

3 - "Requirements" tab

This tab is to be completed by both the data holder and the project leader. The cells are categorized as follows:

- Yellow cells are to be completed by the project leader.
- Purple cells are to be completed by the data holder.
- Red cells are to be completed by the HDH.

The columns function as follows:

- Column A corresponds to the field label or title.
- Column B is to be filled in according to the color code mentioned above.
- Columns C and D correspond to the weighting and determine the calculation.
- Column G provides an indication of the type of value expected when a cell is to be completed (free text, date, number, boolean, decimal, etc.). When there are multiple cells to complete on the same line, the expected value types are separated by "//".
- Column H indicates whether the information entered on that line impacts the calculation of the total project costs. If "Yes" is indicated, it means the value entered on that line is used in one of the Work Package tabs and dictates the involvement of certain profiles and/or the duration of their involvement in the project (i.e., the associated number of man-days). As with the expected value types, when there are multiple cells to complete on a single line, their impact on the final cost is also detailed and separated by "//".

This tab serves to characterize both the database being made available and the project for which data provision and support are requested, enabling the file to automatically generate a corresponding cost estimate. Indeed, the costing is based on a set of predefined tasks and sub-tasks organized into different Work Packages (*see Section 4 for details*). For each task and sub-task, a set of profiles to be mobilized is also predefined. Their involvement and the duration thereof are therefore automatically populated by the file based on the project characterization completed by the user in the "Requirements" tab.

Project characteristics to be completed by the project leader only

1. General information	
Project name	*Name given to the study by the project leader*
Date	*Date of completion of the fee calculation grid*

HEALTH DATA HUB

Name of the project-leading center	*Name of the center the project leader belongs to*
Name of the project's partner center(s)	*Name of the partner center(s) involved in the project*
Name of the project's coordinating center (if there are multiple stakeholders)	*Name of the coordinating center, if the project involves multiple stakeholders*
Type of project leader	*Internal Academic; French Academic; Other public actor; European Academic; International Academic; Mixed; Industrial/Private; Association; Micro-enterprise; Small enterprise; Medium-sized enterprise*
Name of the medical lead	*Name of the person responsible for the study*
Project duration (months)	*Project duration as estimated by the project leader*
Use of the technology platform (opening of a project space)	*Yes / No depending on whether the project leader's Health Data Warehouse hosts the project on its technological platform*
2. Project purposes	
Study type (use case)	*Feasibility study, population targeting, statistical analysis, etc.*
Name of the pathology	*Name of the pathology being studied*
Rare disease	*Yes / No*
Data recovery period, depth (in number of years)	*To be indicated in number of years*
Patient information procedure	*Individual or collective, noting that individual information results in a higher complexity score as it requires greater effort to inform each patient*
Request for access to data updated on the HDH platform less than 1 year ago	*Yes / No* *Noting that access to data updated on the platform within the last year triggers the addition of an optional WP, as it requires specific actions on the part of the data holder. (Information regarding database updates can be found here: https://www.health-data-hub.fr/catalogue-de-donnees)*
5. Algorithms development *These costs may be borne by the HDH and/or the data holder.* *Costs borne by the HDH do not result in any billing for the project leader.* *Costs borne by the data holder result in billing for the project leader. This applies if the data scope requested by the user includes data updated on the HDH platform within the last year.*	
Linkage with SNDS database	
Linkage algorithm with SNDS database wished	*Yes / No* *Depending on whether the project requests a linkage between the HDH Data Warehouse database and SNDS database*

6. Provision of project space *(These costs are borne solely by the HDH and are not billed to the user)*	
Platform requirements	
Duration of analysis (project space availability) - in months	*Total in months*
Data archiving period (in years)	*Total in years*
Disk space used in TB	*Total in TB*
GPU server (number * months)	*Total number of GPU servers made available to the project leader*
Disk space used for archiving (TB)	*Total in TB*
Number of accesses requested by the project leader	*Total number of platform access requests submitted by the project leader*
Provision of tools	
Available tools	*List of tools already available on the HDH technological platform (https://www.health-data-hub.fr/offre-technologique) that the project leader intends to use to carry out their project*
Determining requirements	*Describe which project need the tool addresses*
Requires adaptation	*Yes / No* *If the project leader requests an adaptation of the tool to carry out their project*
Tools to be developed	*If the project leader requests the development of new tools to carry out their project*
Complexity of development	*Simple / Complex*
License costs new tools	*Total cost of licenses related to the new tools*

4 - "Project Steps" tab

This tab is structured around six main work packages (1 to 5 and one optional WP) which include the tasks required for data provision and project leader support. To these is added a complementary work package (0), which completes the set and corresponds to the cross-functional project management work necessary to monitor the progress of the tasks to be completed during the data provision process.

Regarding Work Package 0, however, the related costs are calculated directly within this tab. It is important to note that this section is fixed for all databases and all projects; no modifications are necessary, and the cells are locked. Accordingly, Column C indicates the profiles that will be mobilized by the data holder for the cross-functional project management related to making the database available via the HDH Data Warehouse. For each profile, the annualized FTE (Full-Time Equivalent) percentage is indicated on the corresponding line in Column D.

Using the information entered in Columns C and D, the file calculates the total cost associated with the service of each mobilized profile based on the time dedicated to the project; this total is displayed in cell F9.

Coordination globale du projet						
		Coordinateurs	% d'ETP annuali	Coût annuel	Total projet	Imputation
WP0	Coordination projet	Chef de projet	5%	134 400,00 €	6 720,00 €	Centre coordinateur
		Directeur de projet	2%	149 040,00 €	2 980,80 €	Centre coordinateur
				Total projet	9 700,80 €	

This is an example.

The six main work packages are as follows:

- Work Package 1: Project Feasibility Review and Quotation. This corresponds to tasks associated with receiving the data access request file and its review by the HDH front desk and the relevant HDH Data Warehouse database team. It also includes the drafting of the scientific protocol and the project feasibility assessment to establish a quote for the project team and, subsequently, sign a partnership agreement.
- Work Package 2: Administrative and Regulatory Tasks. This includes the tasks required for the project to obtain the authorizations necessary for its implementation. It involves procedures with the database's CSE (Scientific and Ethical Committee), as well as the CESREES and the CNIL.
- Optional Work Package: Data Update Actions. This corresponds to actions performed by the data holder when updating data in the HDH Data Warehouse. This work package becomes mandatory when the project leader requests access to data that has been on the platform for less than one year.
- Work Package 3: Preparation of Data and Project Environment. This focuses on the preparation within the HDH platform. Tasks included in this work package are data extraction, quality processing, transmission of data to the project space, and making it available to the project team. For HDH Data Warehouse databases that were not initially linked to the SNDS database, and in cases where the project team wishes to link the database with that of the SNDS database, the tasks involved in performing the data linkage are also included in this work package.
- Work Package 4: Analysis. This concerns the analysis the project team intends to perform and, as such, includes tasks related to drafting the analysis plan and the various subsequent analyses that may be conducted (statistical and medical).
- Work Package 5: Scientific Promotion and Project Closure. On one hand, this corresponds to the scientific promotion of the study results. It includes tasks associated with drafting and reviewing scientific articles as well as publishing results. On the other hand, this work package covers project closure actions and the archiving of the mobilized data.
- To ensure each work package calculates the relevant costs, project leaders only need to complete the yellow cells in Column H of the "Project Steps" tab if the tasks comprising each package are indeed to be carried out within the framework of the considered project. When the user selects "No," all costs related to that task are set to zero in the tabs corresponding to the different work packages involved.

Column K indicates the actions that are mandatorily performed by the HDH (or not) within the framework of a request for access to a HDH Data Warehouse database. Column L indicates the actions performed by the data holder that are subject to fees. Column M indicates the services offered by the data holder that you are free to select or not. These services require an additional contracting phase between the project leader and the data holder to define the procedures for performing these services and the associated costs.

5 - "WP #" tabs

These tabs are automatically populated based on the information provided in the "Requirements" and "Project Steps" tabs. Neither project leaders nor data holders need to manually complete these tabs, with the exception of Column K, which can be updated by the project leader to define which project stakeholder (project leader, project coordinator, or project partner) is responsible for the expenses.

Each of the tabs dedicated to Work Packages 1 to 5 and the optional WP is organized as follows:

- A first table summarizes the total cost associated with the work package, followed by a table providing the details for each task that comprises it.

				PDS Cost	DH Cost
WP1	Project feasibility review and quote	1.1	Receipt of application	- €	na
		1.2	Protocol drafting	na	na
		1.3	Review of the protocol	na	na
		1.4	Feasibility study	- €	na
		1.5	Quote	- €	448,00 €
		1.6	Contractualization	- €	112,67 €
				- €	560,67 €

This is an example.

- A set of tables—one per task—provides a detailed cost breakdown for each task by dividing them into sub-tasks. These tables present information according to the following structure:
    - The type of expense: personnel, overheads, subcontracting, infrastructure contribution, or purchase.
    - The position or profile mobilized for the sub-task in question (*if applicable to the type of expense considered*).
    - The description of the sub-task.
    - The cost of one working day according to the mobilized profile (*if applicable to the type of expense considered)*.
    - The theoretical units correspond to the number of man-days required to complete the task. This is calculated in a standardized manner based on scores derived from the information entered in the "Requirements" tab for relevant sub-tasks. In cases where expenses do not involve human resources, a number of theoretical units may still be entered (e.g., for a purchase).
    - The "Task to perform" column indicates whether the task is to be carried out ("TRUE") or not ("FALSE"), based on the selections made in the "Project Steps" tab.
    - The number of units corresponds to the number of man-days actually allocated to perform the sub-task (or the number of units required if the sub-task does not involve human resources); i.e., it equals the theoretical units if the task is conducted, and 0 otherwise.
    - The "Total HDH" column presents the total costs associated with the sub-task for the project, performed by the HDH. In practice, for tasks performed by the HDH, no costs are billed to the project leader. The amount entered is automatically €0. This detail is provided for informational purposes only.
    - The "Total DH" column presents the total costs associated with the sub-task for the project, performed by the data holder. These costs are billed to the project leader.
    - **<u>Note</u>**: Costs related to services offered by the data holder are not calculated in this grid. This part is to be discussed directly between the data holder and the project leader without HDH intervention.

- ○ The "Imputation" column is the only one to be completed in the Work Package tabs. It indicates whether the costs related to the sub-task should be allocated to the coordination center, the leader center, or the project partner centers.
- ○ Finally, the "Multicenter cost" column specifies if the cost associated with the task is multicentric. These costs will only be allocated if centralized coordination is necessary (i.e., if "Yes" was indicated in cell B33 of the "Requirements" tab).

Type of expense	Profile	Description	Cost/day	Theoretical number of units	Task to perform	Number of units	Total PDS	Total DH	Imputation	Multicenter cost
1.1 Receipt of application										
Staff	F PM	Examen du dossier par le guichet de la PDS	- €	0,70	FALSE	0,00	- €	na	Coordination center	No
						Subtotal	- €	na		
1.2 Protocol drafting										
Staff	SME	Protocol drafting	Refer to the list of services offered by the data holder.				na	na	Leader center	No
Staff	Data S	Protocol drafting					na	na	Leader center	No
Staff	BioStat	Protocol drafting					na	na	Leader center	No
						Subtotal	na	na		

This is an example.

- The final two tables present the work package costs by type of center using two different methods. The first displays the total costs broken down by expense type, while the second displays personnel costs only, broken down by profile. These costs do not include the "Total HDH" column. They therefore represent the amounts that will be billed to the project leader.

Detail by type of expense	Coordination center	Leader center	Partners centers	Total
Staff	448,00 €	112,67 €	- €	560,67 €
Subcontracting	- €	- €	- €	- €
Purchase	- €	- €	- €	- €
Infrastructure and DWH contribution	- €	- €	- €	- €
Overhead costs	- €	- €	- €	- €
Total	**448,00 €**	**112,67 €**	**- €**	**560,67 €**

Details by staff type	Coordination center	Leader center	Partners centers	Total
BioStat	- €	- €	- €	- €
PM	- €	- €	- €	- €
F PM	224,00 €	- €	- €	224,00 €
T PM	224,00 €	- €	- €	224,00 €
Data I	- €	- €	- €	- €
Data S	- €	- €	- €	- €
Devops	- €	- €	- €	- €
DPO	- €	- €	- €	- €
Counsel	- €	112,67 €	- €	112,67 €
SME	- €	- €	- €	- €
CISO	- €	- €	- €	- €
Total	**448,00 €**	**112,67 €**	**- €**	**560,67 €**

This is an example.

6 - "Costing" tab

This tab contains two tables. The first is a request characterization table; indeed, the tab is designed to centralize all information related to the project: both its characteristics and its cost. As such, this first table incorporates the elements provided about the project in the "Requirements" tab.

Only cell D9 must be completed with "Yes" or "No" by the data holder to indicate whether or not they have chosen to bill for the project. This indicates whether the project is billable: for instance, if it is an internal project (where the project leader center is also the data holder, the project is non-billable; this is the case when a data holder seeks a cost estimate for a grant application or call for projects) or if it is a project initiated by an external request (i.e., the project leader center is different from the center acting as the data holder, in which case the project is billable and invoiced to the leader center).

The second table includes:

- In Column F: all project costs borne by the HDH, broken down by work package (blue cells) and by individual tasks (white cells). As the HDH service offering is free of charge, the total cost is €0.
- In Column G: all project costs borne by the data holder, broken down by work package (blue cells) and by individual tasks (white cells). The total project cost for the data holder is indicated in cell G61. This amount is billed to the project leader.

Note: Costs related to services offered by the data holder are not calculated in this grid. This part is to be discussed directly between the data holder and the project leader without HDH intervention.

7 - "Settings" tab

This tab centralizes both the rate scales used to calculate the mobilization costs of the various profiles involved in the project, as well as the scores related to the different options proposed for the project description criteria in the "Requirements" tab.

Regarding the rate scales, please note that several assumptions are made:

- 210 working days per year are considered; this value can be adjusted according to the number of working days in the specific year.
- A coefficient of 1.6 is used to convert gross salary into total payroll costs.
- Outsourced payroll costs are used to calculate costs in the various Work Package tabs. Indeed, it is assumed that approximately 50% of the tasks are performed by outsourced personnel. Furthermore, a second assumption is that outsourced personnel costs are double those of internal staff. Therefore, the costs used for all profiles correspond to total payroll costs multiplied by 1.5: this is the outsourcing factor indicated in cell B3, which can be modified if different assumptions are adopted.

The tables starting from Column J serve to reference the scores allocated to the various options proposed in the "Requirements" tab.

A price reduction may be offered to certain actors, such as civil society and startups/SMEs. The level of this reduction is left to the discretion of the data holder. This reduction is adjusted in Column L of the "Settings" tab.