ISTITUTO NAZIONALE DI FISICA NUCLEARE
Direzione Sistemi Informativi

INFN-23-22-DSI
31 May 2023

A time tracking system for EU-funded INFN projects

Antonello Paoletti[1], Marco Canaparo[1], Claudio Galli[1], Francesco Serafini[1]

[1)] *INFN, Direzione Sistemi Informativi, I-00044 Frascati (Roma), Italy*

Abstract

The aim of this work is to describe the structure of a web application to support EU-funded INFN projects in the process of time tracking and reporting. The system has been designed in the early 2010s, following the growth of the EU-funded projects and the need for a framework to handle consistent reports according to the EU guidelines.
One objective of this project is to guarantee and certify a valid distribution of hours declared on EU projects by INFN personnel in consistency with the official INFN clocking system, the scientific project management system and, overall, any signed agreements between the INFN and the founding bodies. The system allows the project participants to declare the hours worked on their projects, detailing their distribution across days, work packages and tasks.
Interfaces are provided for *Project Managers* and *Financial Officers* to manage a project and monitor its trend. This activity is also supported by the INFN Business Intelligence Service, integrated with this system, to provide comprehensive reports for the upper management.

DOI n. 10.15161/oar.it/77018

Published by
Laboratori Nazionali di Frascati

1 Introduction and Motivations

The growth of external funded projects has set new challenges and standards in the field of reporting. The funding bodies, in particular the **European Commission**, require that the beneficiary institution of a grant adopts integrated systems to comprehensively track and monitor members activities, not only on funded projects, but also considering any other activity that contributes to the annual number of hours, to avoid overexposure and double funding.

Quoting the "Guide to Financial Issues related to FP7 Indirect Actions"[1]: "*Only the hours worked on the project can be charged. Working time to be charged must be recorded throughout the duration of the project by timesheets, adequately supported by evidence of their reality and reliability. In the absence of timesheets, the beneficiary must substantiate the cost claimed by reasonable means (alternative evidence) giving an equivalent level of assurance, to be assessed by the auditor*".

Such guideline states requirements that must be met both by the underlying processes and the reporting systems, with a specific focus on the **consistency** and the **accountability** of the declared data. The same guide states: "*In cases where personnel work on several projects during the same period the time recording system must enable complete reconciliation of total hours per person, listing all activities (EU projects, internally funded research, administration, absences etc.)*" specifying that the timesheet must become a certified representation of the whole time spent on research, of which it is part the project of interest to the funder.

According to this set of guidelines, the INFN has requested to develop a system to track the worked hours declared by project members, integrated with non-project related activities (e.g., teaching), days of illness and leaves in a comprehensive report. This tool is part of a change of *vision* towards a *Project Management* approach in the process of management of a scientific experiment, as requested by the founding bodies.

The development has been commissioned to the **INFN IT Direction** and has requested 3 years to be completed. In 2014, the system has been tested and rolled out to the end users under the supervision of the **INFN Research Services Direction**.

2 Use Cases, Actors and Business Process

The **timesheet** system has been designed to implement five use cases, respectively dedicated to **researchers**, with a specific focus on the *data-entry* tasks, to **PMs** with a focus on *management* and *evaluation* tasks and **financial officers**, with a focus on *supervising* and *reporting* tasks. There is also a service use case dedicated to clocking data acquiring, that supports the integration with the **INFN clocking system**.

[1] https://ec.europa.eu/research/participants/data/ref/fp7/89556/financial_guidelines_en.pdf

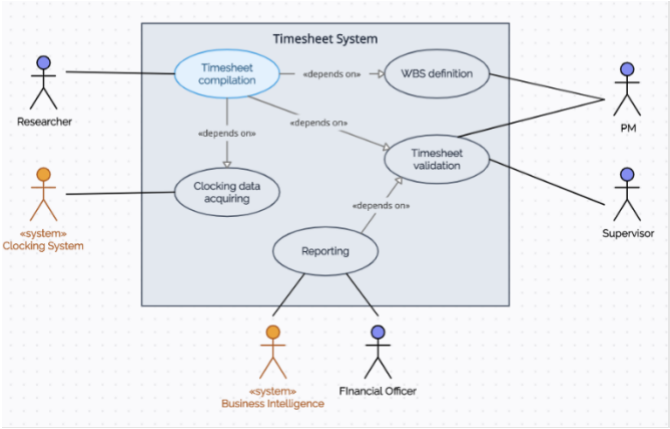

Figure 1: Use cases diagram

Such use cases share a common workflow that starts with the *Project Manager* (usually the *Principal Investigator* or the National Project Responsible) who defines the project structure (*Work Breakdown Structure*[2]) in terms of *work packages* and *tasks*. The flow continues with the *Project Participants* (*Researchers*), who are responsible for declaring the worked hours, then with one or more *Supervisors* (*WP Responsibles* or Local Project Responsibles) that validate the timesheets and ends with a *Financial Officer* that collects all the signed documents and reports back to the founding bodies.

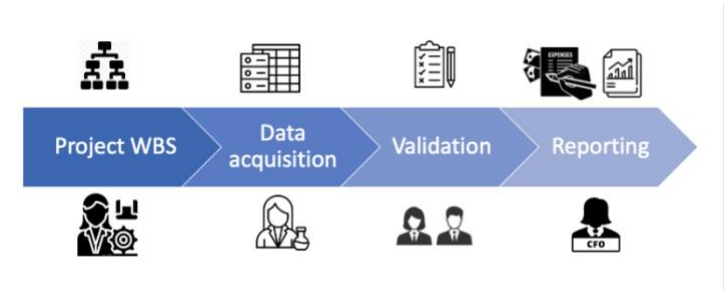

Figure 2: Workflow of a timesheet

3 Data Model

The data model of this system is structured around the main *entity* called **timesheet** that contains information about declared hours by a user, on a *Work Package* in a specific day of the year. This *entity* interacts with other *entities* derived from the **workflow management system** (approval statuses), the INFN **clocking system** and the scientific **projects/roles management system** with *n:m* relations that contribute to build the model.

It is important to note that this model is purely conceptual as information coming from external systems is acquired through *API* and integrated at the application layer.

[2] https://www.projectmanager.com/guides/work-breakdown-structure

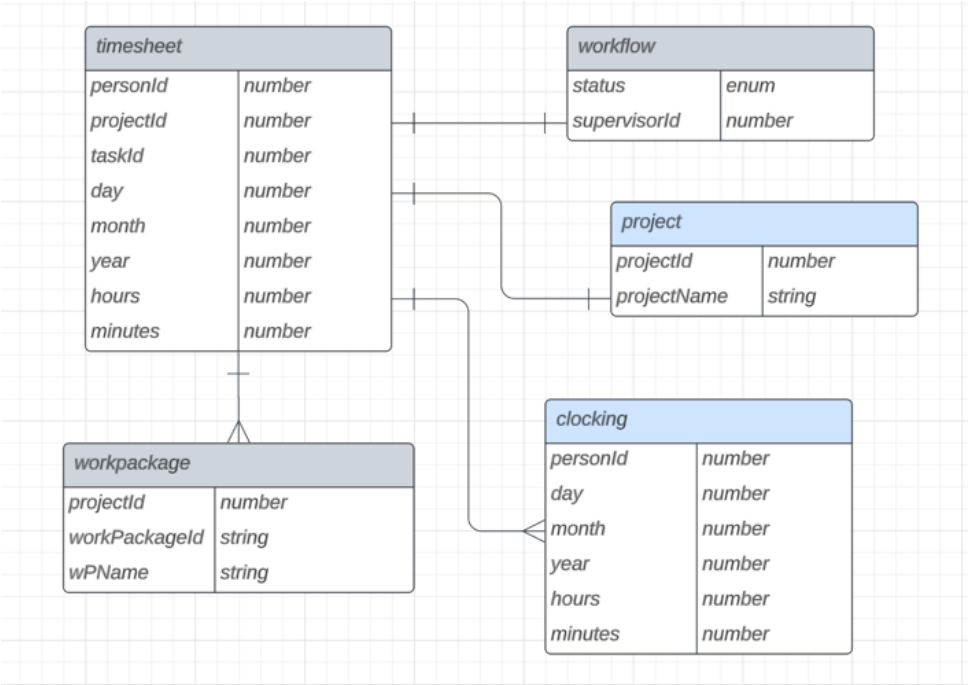

Figure 3: Conceptual ER diagram

The **timesheet** entity is represented as an *excel-like* matrix that stores the declared hours worked by a researcher, organized by day, project, and project/work-package/task. This entity is managed through a graphical user interface structured according to the project memberships of a user and prefilled with information coming from the INFN clocking system.

The information provided by the INFN clocking system plays an essential role in the **consistency** validation process as it operates as a baseline/limit to the data-entry. For example: if the clocking system reports that a researcher has been at work for 7 hours and 10 minutes, the system will allow them to enter a total amount of 7:10 (no more, no less) distributed among their projects, their personal research, teaching, and other activities. This data structure has other constraints that contribute to its validation and consider a daily number of total hours (according to the profile level of the researcher) and a yearly amount of total declared hours (according to the agreements with the funder).

The user is required to fill in such information as pairs hours/minutes for every task or activity that is available for compilation. Completed this task and ensured that the total declared hours match the total worked hours from the clocking system, the user can submit the timesheet and "freeze" the information in the database. After the user submission, the system activates a validation workflow that allows the local supervisor and then the principal investigator to view and validate the timesheet.

Principal investigators and financial officers can access aggregated reports with lists of project members and their timesheets (total hours, validation status…) and can setup the *breakdown structure* of a project in terms of *work-packages* and *tasks*.

3.1 The INFN Clocking System

The INFN clocking system serves as a source of data pertaining to the activities of INFN employees. Specifically, the software generates a materialized view that contains the information extracted from each employee's attendance card. In the following paragraphs, we are going to describe how the data present in an example of attendance cards are reported into a materialized view.

Figure 4 shows an excerpt from an employee's attendance card. From left to right, the columns represent the day of the month, the clock-in timestamp, the clock-out timestamp, the type and duration of any absence or work leave, and the total number of hours worked for the day.

Table 1 presents the records from the materialized view corresponding to the data from the attendance card. It shows 4 records of type "*work*" each one of which corresponds to one day in the attendance card. A record of work type encompasses various types of work available in the INFN clocking system, including both remote and on-site work. Remote work is indicated in the attendance card as "*lavoro agile*" (IT for *smart working*), while on-site work is denoted by the presence of clock-in and clock-out timestamps. The other two records indicate the presence of an absence leave named "*festività abolite*" (IT for *holiday abolished*). For example, on March 6th, an employee worked on-site for 6 hours and 14 minutes and took 2 hours and 8 minutes of "holiday abolished" leave. On March 7th, the same employee worked remotely for 7 hours and 30 minutes.

The communication between the INFN tracking system and the clocking system operates on a polling approach. When an employee requests to view their timesheet through the web application, the information is retrieved by utilizing an API that collects data from the materialized view in the INFN clocking system.

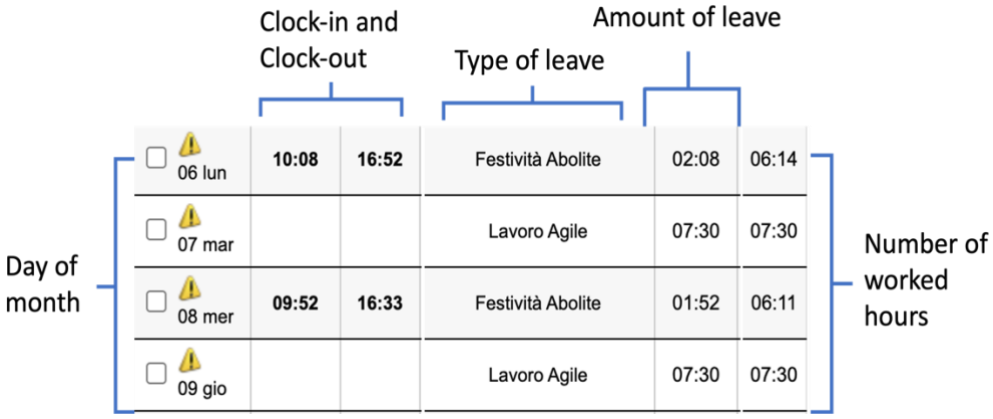

Figure 4: Excerpt from an employee's attendance card

EMP	YYYY	MM	DD	Activity	Amount (HH)	Amount (MIN)
1	2023	03	06	Leave	02	07
1	2023	03	06	Work	06	14
1	2023	03	07	Work	07	30
1	2023	03	08	Leave	01	52
1	2023	03	08	Work	06	11
1	2023	03	09	Work	07	30

Table 1: Excerpt from a materialized view

4 Architectural Design and Development Process

The architecture of the system reflects a common design pattern of the 2010s, based on the *Model View Controller* paradigm in conjunction with *CodeIgniter*[3], a widespread PHP framework. Currently, the source code supports version 8.x of **PHP** thanks to a full upgrade carried out in 2022.

The application has a layer of data persistence based on **Oracle RDBMS** and offers authenticated/authorized access using the **INFN-AAI** infrastructure with *oAuth2* protocol.

The scaffolding of the code reflects the separation among the three types of classes responsible for *Data Modeling*, *Flow Control* and *Presentation* as stated by the MVC/MVP standard. Being a monolithic PHP application, the *Webservice* layer is not implemented *by-design*, even if several controllers class are designed to ease a transition towards a (micro) services paradigm. The framework uses *Composer*[4] as a dependency management system to include library and software components in a modern and solid fashion.

Each use case is modeled as a *Controller* that governs the business logic and access business data through *Data Access Objects* that wrap the persistence layer. Each DAO handles an *entity* on the DBMS and models every interaction with the logic layer.

[3] https://codeigniter.com/

[4] https://getcomposer.org/

5 User Interface and browsing

Upon logging in, the user is greeted with a dashboard that provides an overview of the projects in which the user has a role of scientific coordination or monitoring. At the top of the interface there is a navigation bar that remains accessible across all pages and contains links to various sections of the application, such as timesheets, projects, reports, settings, and support. The bar also includes a search utility for quickly finding specific projects or researchers.

Figure 4: Landing page

The "Go to my timesheets" link redirects a user to a page dedicated to the timesheet creation or management. The interface allows users to select a specific month and choose the project or task they worked on during that period. The timesheet interface is structured as a calendar view, allowing users to visualize their work schedule in a specific month. The user can navigate through different weeks or days, view/hide the *work packages* or *tasks* related to their projects and fill in the number or hours worked.

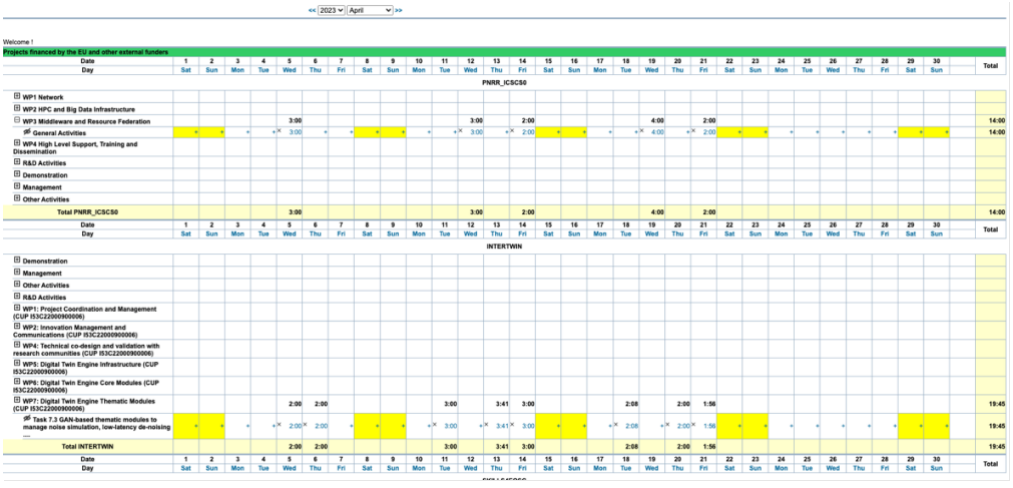

Figure 5: Timesheet table

The page contains a section dedicated to the approval process where the scientific supervisors can review and approve the timesheets submitted by their team members in a three-step *workflow*. Such *workflow* is activated by the researchers themselves, upon

completion of the timesheet for a specific month. After this first step, the system freezes the information in the database and sends an email to the local project supervisor, waiting for their approval that allows the principal investigator to “close” the timesheet.

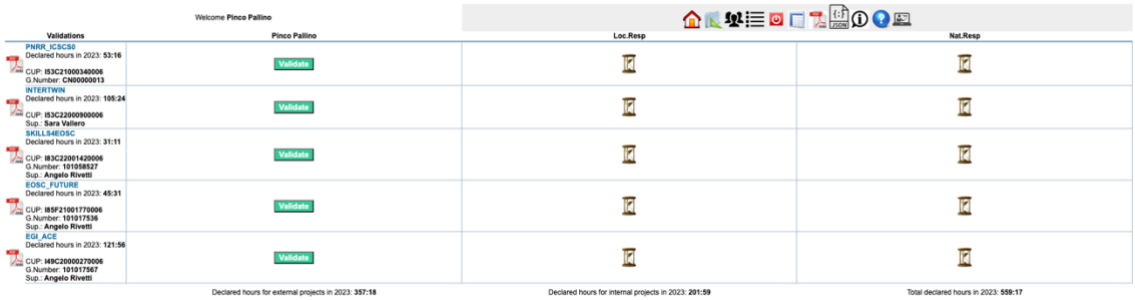

Figure 6: Project validations

Principal investigators, supervisors and financial officers can access a project dashboard showing aggregated data (e.g., total hours declared per work package) and the approval status of any participant with different colors according to the *workflow* status. From this page it is also possible to configure a project in terms of *work packages* and *tasks* to be activated for the compilation (*Work Breakdown Structure*).

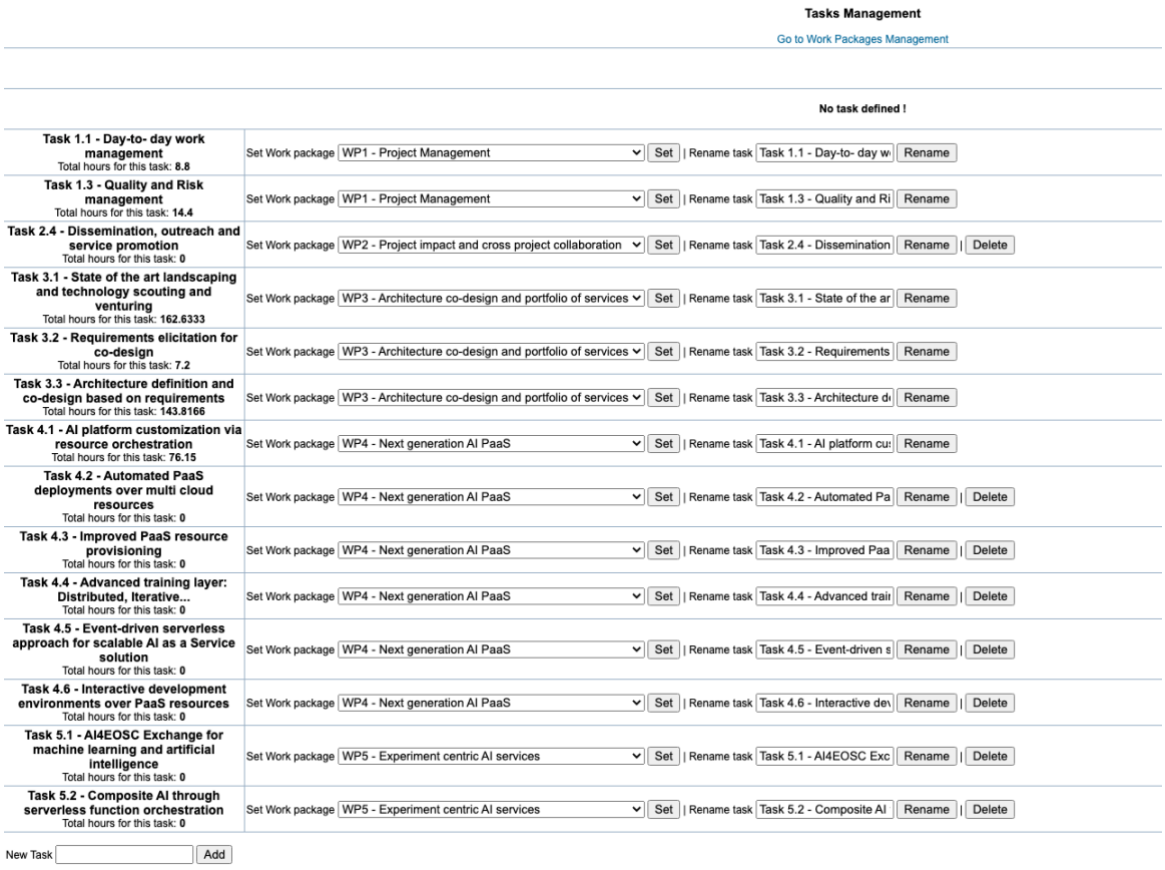

Figure 7: WP and tasks management

6 Business Intelligence

In early 2022, significant effort was dedicated to the implementation of a complex *data warehouse* to support the need for a comprehensive reporting system. This demand has grown within the Italian institutions, like INFN, due to the recent EU framework programs (including NextGen EU and the Italian PNRR) which have provided substantial fundings and necessitate detailed reporting on fund utilization. The BI unit of the INFN IT Direction has carried out a full recognition over the timesheet data model to design a solid ETL process and build a reliable *datamart*.

6.1 ETL and Data Warehouse

As shown in the following image, the *Extract, Transform, Load (ETL)* process is quite complex as it involves DB connections, data transformations and integrations with other *datamart* that are conceptually connected. To give an example, information on projects, people's profiles, and employee contracts are correlated with typical information managed by the time tracking system, such as “timesheets”, its completion status, and approval levels, to create a *datamart* enriched with information from other application areas.

The result of this work is quite useful for many use cases and contributes to building a coherent research *data warehouse*. Such *data warehouse* is a compact representation of a complex and distributed data model that allows data scientists and managers to build reports upon.

Figure 8 highlights three components related to different aspects of the data model: experiment configuration, contractual and personal information of employees, and certification documents for the work they have done on behalf of various experiments.

Particular attention has been given to aligning employees’ contractual periods with the roles within the experiment and the details of monthly work certifications.

To achieve this result, several data processing steps have been implemented, aiming to execute them in parallel as much as possible to reduce processing times. However, certain phases were kept in sequence when they had a strong functional dependency.

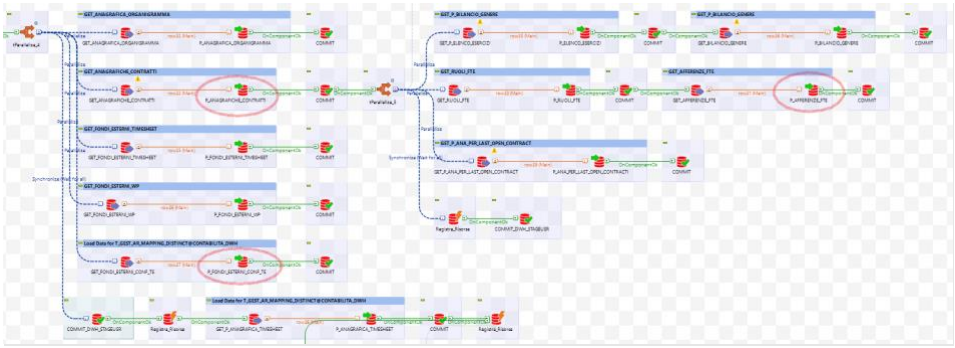

Figure 88: ETL workflow

6.2 Reports

The main usage of the pre-built *datamart* is building specific reports to support the management and the founding bodies in the activity of monitoring. The INFN BI System is structured around the **Jasper Suite**[5] and consists of a plethora of active reports designed by the BI Unit of the IT Direction. Such reports are the final deliverables of the entire process and are accessible through a web interface exposed to PMs, Financial Officers and the INFN Board.

Below is presented the report as it appears for a Financial Officer who wishes to have both a compact and analytical view of the completion status of timesheets related to people involved in the projects they are responsible for.

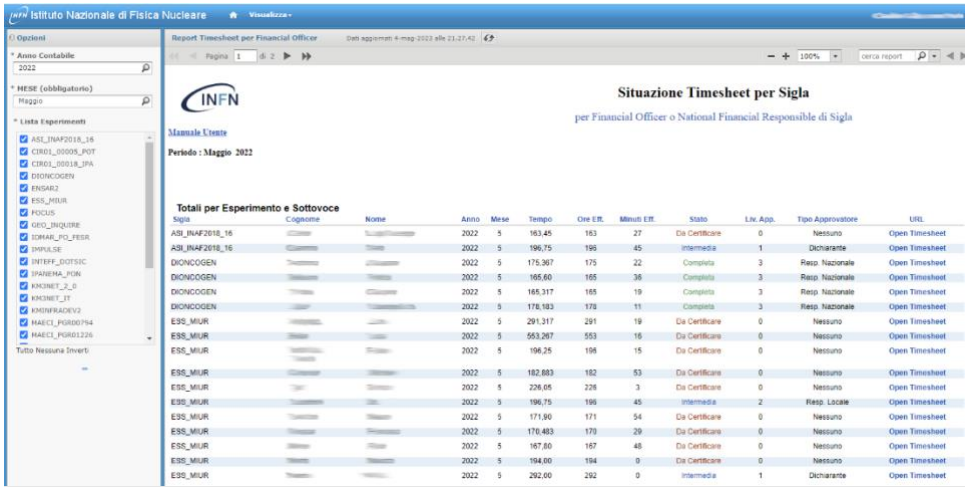

Figure 9 9: Timesheet report

The URL column provides a direct link to the timesheet for which the summary data is presented. This useful functionality allows the Financial Officer to have a simple access to both aggregated view and analytical details within the time tracking system that represents the data source for the *datamart.*

[5] https://community.jaspersoft.com/project/jasperreports-server

7 Usage and Statistics

The software is considerably mature and has a maintenance policy based on a *backlog* of fixes and evolutions co-managed with the **Research Services Direction** of INFN. As of 2022, the system has been used by more than 1200 people (in red) and more than 320 projects (in pink) with a constant growth from 2014.

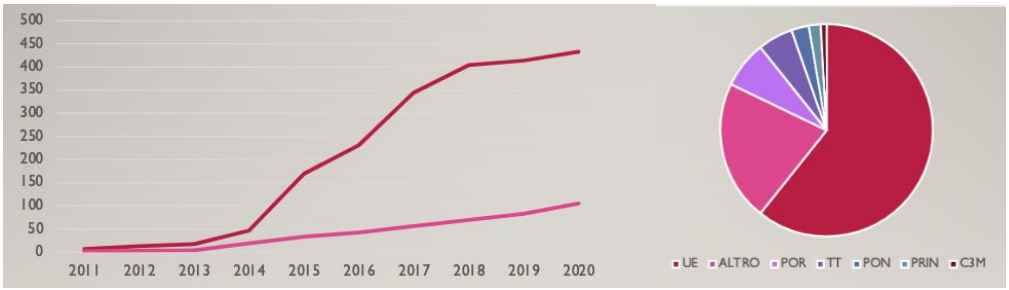

Figure 10 10: Trend of timesheet usage

The system is mainly used to report EU projects (in red) but the number of other projects, especially national ones, is also growing.

8 Future Developments

The main areas of intervention are concentrated on the already mentioned *backlog* and a long term technological refactor based on the current trend of (micro) services applications, even if the version 8.x of the PHP language has given new life to the system. A major improvement can come from a change of perspective on the persistence layer, switching from a relational approach to a *NoSQL* one that now dominates the market.