Workplan

A. Context

In the EuroNanoLab network, we address the challenges of storage and exchange of data produced by the infrastructure by defining first an efficient **Common Data Format (CDF)** and then implementing an optimal design of the metadata infrastructure. Thus, the existing standards and recommendations for metadata will be reviewed in order to help with the unambiguous definition of metadata elements and their attributes.

For this purpose, we will refer to the many initiatives that are currently active, both in Europe and outside, to deal with the issue of an efficient management of data repositories. In particular, our goal will be to design metadata formats which ensures interoperability. This means to use a harmonized language through the MODA template, (https://emmc.info/moda-workflow-templates/), as done in several other H2020 projects.

An impressive work has been also done in this regard by the H2020-NFFA project (www.nffa.eu), in the framework of which a "draft metadata standard for nanoscience data" report has been released (http://www.nffa.eu/media/124786/d112-draft-metadata-standard-for-nanoscience-data_20160225-v1.pdf). However, this work is more dedicated to data associated to the description of nano-materials (composition, structure, properties) than to fabrication process data (parameters of nanofabrication machines).

B. Primary objectives/tasks and deliverables

Since nanofabrication processes are usually chains of « elementary » **process steps** that represent the action of a cleanroom machine and a cleanroom operator on a sample, we assume that it is possible to reduce a NF process control to the control of each process step. Therefore, in the following text, the word « process », will always refer to « nanofabrication process steps ».

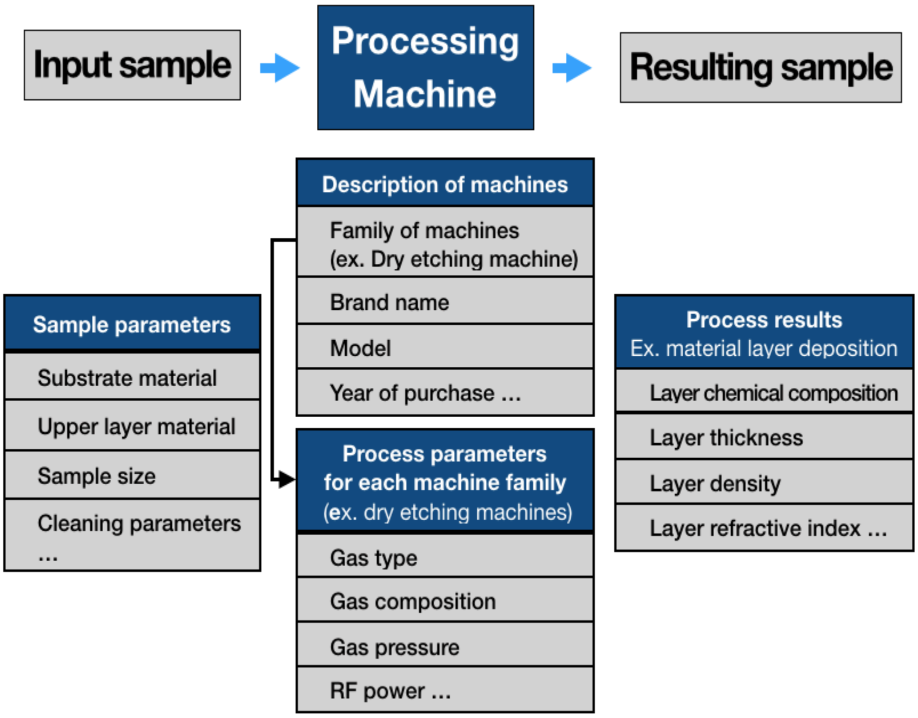
The goal of this work program is to study the possibility of describing NF processes with a high enough fidelity that will enable a relevant standardization of the description of NF processes, through data sets of reasonable size.

This goal requires a lot of experimental work that may take years. A first round has already been undertaken on a relevant category of process steps (dry etching processes) and the identified difficulties will help for further optimization of the workplan.

This mid-term program is divided in two main workpackages:

- **WP1: TOWARDS A HIGH FIDELITY OF NF PROCESS STEPS DESCRIPTION**

Any cleanroom **process step** involves a sample that is modified by a « processing machine » that produces a new « resulting sample » (see figure below).



Required databases to describe a cleanroom process step

Four kind of databases are required to describe completely a cleanroom process step:

- The database containing the **description of the samples** to be processed (example: material chemical composition & structure, size, geometry, sample preparation parameters, etc.). This database should contain enough fields to allow describing almost any possible sample (and its preparation state) that may be processed by a specific family of machines.
- The database containing the **description of all the machines** of the network, grouped by families of similar machines (example: deposition, dry etching, wet etching, crystal growth, lithography, etc.). Any difference between two machines (for instance different options installed, different gas sources, etc.) should be recorded in this database. It is of uttermost importance to notice that the description of the machines should often also contain the initial state of the machine - **including the history of previous machine operations** - because traces of materials that have been processed previously in the machine may greatly influence the machine behavior: there is sometimes a « conditioning » of the machine by the previous usage that changes parameters settings to obtain the desired result and « contamination » issues that may influence the quality of the results, especially when a high level of purity is required.
- **Process steps parameters** should be described in a database containing **machine settings** for a given family of machines (for dry etch machines: gases partial pressures, gas flows, RF power, etc.). All possible variations of the machine settings should be included as well as any possible hidden parameter that is suspected to have an influence on the process step result.
- The database containing the **description of process steps results** for each family of process steps and input sample (ex. for deposition of a thin layer, the thickness/roughness of the obtained layer, its density, refractive index, conductivity, optical losses, etc.). This database is oriented towards the complete and precise description of materials and the definitions of its fields may use the work perviously done by NFFA and mentioned above.

The work to be done is to identify data sets of minimal size that can represent some specific NF processes with enough fidelity to ensure its reproducibility and reusability.

Il the following, a PROCESS REPRESENTATION will refer to a list of parameters names that describe the process- not to be confused with the values of these parameters.

The PROCESS DESCRIPTION is the list of parameters names with their values. A process description is linked to a sample and a desired result (there will be different process descriptions for different samples or desired results).

Objective 1.1: identify TEST CASES to study the possibility of obtaining a high fidelity REPRESENTATION of NF processes

Process representations will depend on the family of NF processes that is considered. Therefore, the parameters that should be stored in the databases mentioned above are different for different families of processes.

A study of different families of NF processes will be done to identify the families of processes whose description database may be enough similar to expect a possible standard description.

After establishing a classification of machines and NF processes for the certification procedure, a suitable test case should be identified (a machine and a family of machines) to be able to test the whole certification procedure for these specific Nanofabrication processes. This work has already started with the case of « Dry Etching » cleanroom machines: this family of machines could become the test case if it is found to be enough representative of a generic cleanroom machine.

Objective 1.2: develop a minimal data set for the NF process representation on the test case

This work is currently progressing through International Expert Groups (IEG) that are each studying a specific category of cleanroom processes (a group on « dry etching technologies » has already been working since 2018 and a group on « lithography technologies » will be launched at the end of 2019).

Theses IEGs are responsible for establishing a list of required parameters for the major process step that can be performed by the machines. This list should be as exhaustive as possible and the fields contents should be described as precisely as possible.

A first version of these elements should be delivered after less than 3 years for a set of nanofabrication processes.

Note that a process description is often aiming at a specific process result for a specific category of sample (for example minimizing roughness of the etched bottom surface on a specific material). **Therefore, the process description should explicitly refer to the kind of process result that is sought**. For this reason, different processes have to be identified for each relevant case depending on the most important criterion to fulfill on the result.

• WP2: CERTIFICATION OF THE TEST NF PROCESS REPRESENTATION

Objective 2.1: develop a certification procedure of process REPRESENTATIONS on the test case for a single model of cleanroom machine

The « quality » of the process representations can be certified using a test procedure that should be as simple as possible. The main quality expected from the process representation being to give a reproducible result whatever the operator and the machine, a minimal variability should be observed for a given sample and a given machine. Therefore, the « quality » of the process representation can be established by running **the process (**with controlled machine settings) on **reference samples** and compare process results (including first order derivatives of process results versus machine settings) for different machines of the network (same brand and same options). The process representation will then be certified if the differences on resulting samples remains within fixed boundaries (example variations of less than 1%). If it is not the case, it will mean that the process representation does not fill the requirements for a certification, probably because there is a parameter that remains uncontrolled and sensitive to environmental conditions. This parameter should then be identified and added to the process representation to obtain a process representation that can be certified.

Objective 2.2: extend the certification of process REPRESENTATION to the whole family of machines belonging to the test case

This certification procedure may be extended to other types of machines of the same family (for instance similar machines from another brand). An extended certification of the process representation can be tested on series of machines belonging to different categories (for example different brands) and should give reproducible results for each category, even though the process description may be different for two categories of machines. This kind of « family certification » (a kind of higher-level certification) may ensure that the certified process representation allows « translation » of the machine settings for a category of machines to another category of machine.

This procedure may therefore be used on the long-term to be able to export processes from one cleanroom to another one when the machines belong to a different category within the same family.

Objective 2.3: on the test case, characterize the validity domain of a process DESCRIPTION

From the process description that is used to obtain a certain result, it should be possible to deduce the process description to obtain another result if these desired results are not too much different. By using measurements of first derivatives of process parameters variations, it may indeed be possible to predict the machines settings that are necessary to obtain another result, as long as a linear extrapolation is possible. For instance, if the etching depth is proportional to the etching time, it is easy to find the machine settings that are necessary to obtain different etching depths for the same kind of process. However, extrapolation of reference processes to obtain different results may only be valid in a certain domain.

This validity domain may be investigated and this kind of investigation would be very useful to understand the different behaviors associated to different parameters variations. At least it should be possible to identify the most critical parameters whose first order derivative are the largest and that should be much better controlled than others to obtain reproducible results.

• WP3: STANDARDIZATION OF NF PROCESS REPRESENTATIONS

Once a process representation is certified, it can become a standard for a given community. This standardisation will then enable reliable exchanges of processes that should be easy to reproduce elsewhere.

All GO NANOFAB cleanrooms using these 4 categories of databases should share the same kind of process representations, which means that **the databases of all cleanrooms should have the same fields** (that should be very precisely defined to allow precise and non-ambiguous understanding of their contents).

The work on this WP will be mainly to wrap-up the obtained knowledge on the test case and use it to write the documentation for the standard.

• WP4: DATA MANAGEMENT ISSUES AND METADATA GENERATION

Use case analysis is the starting point to build the metadata architecture in GO NANO, which can be assisted by formal languages, like UML. UML, with its different diagrams, address from first questions, like “who are the users of our system” to the development of entity relationships diagrams corresponding to metadata models, ready to be implemented in a database.

The first use case refers to building a standardized representation of a NF process.

GO NANO Starting and general aim and USE CASE

Transfer a process from one cleanroom to another – DESCRIBE THE PROCESS IN A STANDARDIZED and SHARED WAY

According to the first Expert Group, on Dry etching, examples will be on this technology. The question to be answered are (example of):

- how to describe a dry etching – parameters, rules, constraints
- how to describe process characterization
- what parameters are common to different equipment
- which is the minimum set of parameters to be managed to certify a process
- which are the information to be equipped with a PID – persistent identifier
- where do we find a certain process, or equipment
- who can access the data, and according to what policy
- tariffs: do we have to pay to use facility and data

Answering to these questions address a first set of entities to be managed shown in Fig.1

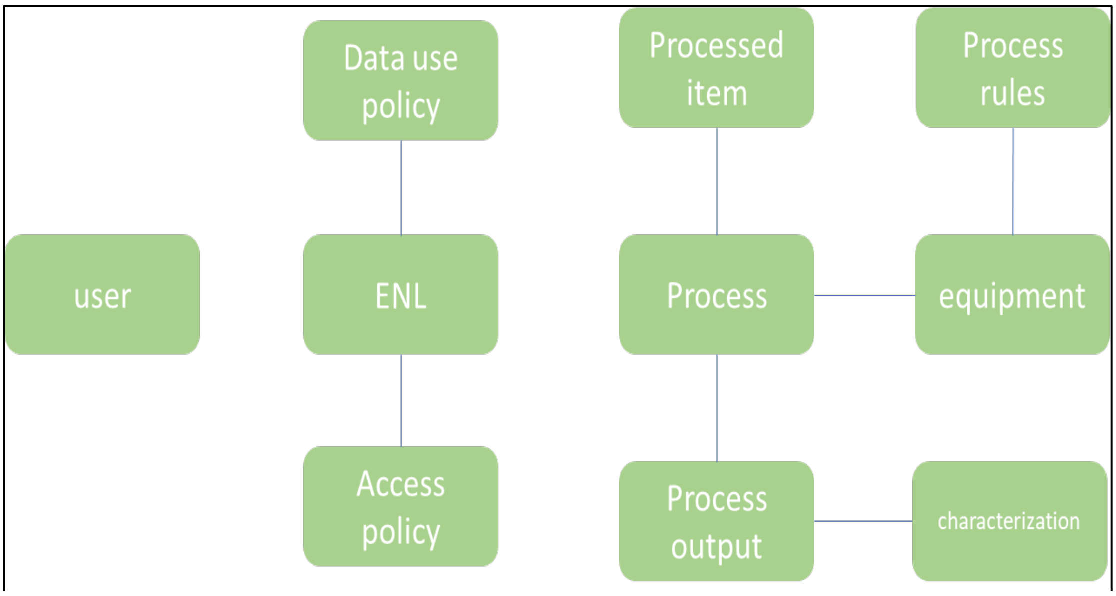


Fig. 1 – main entities of the GO NANOFAB project

To manage all information generated from the entities in an EOSC and FAIR compliant way, several choices must be addressed. In Table 1 some preliminary solutions are listed.

Architecture	Some options	Notes

Metadata standard	vCard, DublinCore, CSMD,….	According to different entities, and attributes, different standard could be used. Eg., vCard for users, CSMD for measure raw data, or equipment parameters
Metadata format/syntax	XML, JSON, RDF	
Persistent identifier - PID	DOI, DataCite.org, ORCID	To be applied to selected entities; among them: users, facilites, equipment, materials

Table 1 – initial formalism under evaluation to implement the ENL GO NANO project

The workplan describes an iterative process, where metadata models are used to generate a database of real equipment, process, materials and rules, which allow testing how the process representation is approaching a standard.

Each iteration will modify the metadata according to the results of real tests, made on samples. The test will say how controlled and standard and optimized the process results according to the metadata. All the WPs in the workplan are interested to this mechanism, starting from WP1, where the building databases are described, to arrive to WP5, the evolution of the architecture including the management of IP and access policies.

This developing and optimization loop will produce a growing architecture requiring two main components:

1. Data repository. Most of the managed data are elementary types – strings, numbers, links, some images, a few videos for tutorials – so that it is possible to estimate non-critical storing needs, which could be found in cloud services like CloudSQL, or Azur, or Aruba, just to mention some examples, offering very popular relational database management systems - Postgres, MySQL, SQL Server.

   These services not only guarantee high reliability of the servers and availability of data everywhere, but also strong security, using encryption and sophisticated algorithm to protect data. All modern database management systems allow to define different levels of access for different groups of users, and this is fundamental to further protect data.

   The choice of a centralized repository is in line with the distributed infrastructure project, where different cleanrooms are collaborating to share technology data, information about equipment etc.

2. Interfaces and clients using the data, the repository. A natural way to imagine interfaces to the database, especially one in the cloud, is via Web applications, with clients running on common browser, and further safety layer like the https protocol. Clients can be developed to easily manage the data and the metadata, and to connect these data to other tools, like the cleanroom management software, process simulators and optimizers, or directly to equipment. Client application can also manage IP and compliance with access policy, building filters to data and users.

- **WP5: COLLECTION OF PROCESS DATA OR METADATA**

Objective 5.1 achieve agreements on appropriate Persistent Identifiers (PIDs)

A correct analysis of process data requires taking in account the variability of some parameters such as nanofabrication machines or individual operators. This kind of information should thus be systematically integrated in the nanofabrication process descriptions.

For that purpose, it is necessary to open a dialogue and pathway for achieving agreement on the identification and implementation of appropriate Persistent Identifiers (PIDs), as well as best practice in data capture. We will aim to identify the key PIDs for metadata content (site, equipment, operator, source of finance, process agents and eventually recipes or processes) and collaborate with GFISCO experts to work on common PIDs in line with recommendations from other initiatives (RDA, EOSC, NIST etc.) and conventions for nomenclatures and versioning. ENL will build and initiate pilot studies to examine technical or context-specific challenges concerning PIDs. Then a joint review document on these pilot studies will be produced and publish after 12-18 months.

GO NANOFAB will seek consensus on PIDs for NF machines (date of purchase, operator, model, serial number, etc.) and ORCIDs for operators, and on the metadata process descriptors necessary to ensure the reusability of steps and modularization of steps from independent cleanrooms with minimal redevelopment.

A first work on process descriptors and PIDs will be done within the EuroNanoLab consortium, following recommendations of International projects such as FREYA. Then the results will be published and promoted through seminars and International conferences. Discussions with other NF centers from the academic or industrial world at international level will result in the constitution of standardization expert groups including all players.

Objective 5.2: automate the capture of key process parameters

After having identified the parameters that influence the most the desired result, it is possible to list a finite set of parameters that should be collected to ensure to the first order a reliable process control. Sensors should be implemented whenever possible to collect automatically these most important parameters. Through an interaction with equipment makers, it should be possible to obtain the installation of the required sensors in the standard configurations of the machines instead of having them on request on a case-by-case basis. These sensors will also enable a much better process control by locking dynamically the key process parameters to the desired values.

Therefore, it is necessary to open a dialogue with cleanroom equipment makers through an international expert advisory panel that can advise commercial actors on the trends and needs related to data capture.

Notes

1. The work mentioned above requires a well-structured community with a lot of cleanrooms that agree to collect and share data with the same format. Process knowledge will increase by monitoring rigorously a large number of trials and by extracting relevant process behaviors from these large multidimensional datasets. If correctly organized, the EuroNanoLab consortium can already provide a significant amount of data that can be analyzed to extract process behaviors. However, it will also be necessary to convince international partners to participate to such studies in order to improve the process knowledge and control.

2. The process results should include material parameters that are often numerous. The description of materials parameters will integrate and follow the work done by NFFA for Nanomaterials.

• WP6: DEVELOPMENT OF RULES AND AGREEMENTS FOR DATA SHARING

GO NANOFAB will seek to promote agreement on process standards, data sharing, accessibility of a large number of available validated processes with results stored in databases, and to promote best characterization practice throughout process development.

Objective 6.1 Establish rules for the sharing of nanofabrication know-how

Knowledge of process parameters usually takes a long time and a lot of efforts. Moreover, it can bring an advantage in the competition between scientists and for the valorization of new businesses based on scientific knowledge. Therefore, nanofabrication centers are often reluctant to share their fabrication process knowledge.

On the other hand, many basic nanofabrication processes are already known from our international competitors, but would still be worth sharing to save a lot of efforts for the development of new processes in individual nanofabrication centers.

To create a positive dynamic, it is therefore necessary to imagine a set of rules that will guarantee a balanced exchange of know-how between public nanofabrication centers and a fair return to know-how providers (acknowledgement of sources, financial return in case of commercial exploitation etc.). There is a particular need to identify optimal conditions for NF data release: since NF data can describe commercially sensitive know-how, embargoes or licensing restrictions may be necessary to enable an effective protection of nanofabrication process providers, which would consequently foster dissemination of the know-how.

The first step will be to write a charter for the sharing of a number of elementary process steps among academic partners. By signing this charter, the partners will agree to comply to regulations guaranteeing a balanced exchange of nanofabrication know-how. A set of shared processes will be thus defined inside the network and may be used to work on standardization or certification issues. The final goal is in any case to test the degree of reusability of nanofabrication processes that can be obtained.

Objective 6.2 Study licensing issues

The next steps will be to study the conditions for sharing process data between academic partners and RTOs or industrial players - taking in account the implications of a possible commercial exploitation of the work - and then define the conditions to transfer NF know-how to partners outside the network.

Transfer of know-how from the network to outside partners should also comply to general regulations concerning the dissemination of know-how and should be strong enough to allow an effective protection in case of usage of know-how without acknowledgement of the sources or commercial exploitation without return to the sources.

Principles for licensing should be established and shared among academic users. Such licensing principles should also be accepted by RTOs and industrial partners that will benefit form the academic work. Data release may be phased or delayed and precise conditions should be established through dialogue and agreement on shared general principles. It is envisioned that licensing nanofabrication know-how could make use of licensing principles from the world of Open source software.

• W7. ENCOURAGE THE DIFFUSION OF FAIR APPROACHES FOR NF PROCESSES

EuroNanoLab will seek to develop a sustainable communication plan for the promotion of FAIR data approaches within the Nanofabrication community.

- The advantages of adopting GO FAIR principle for individual NF centers will be highlighted
- EuroNanoLab will organize specific meetings dedicated to data management and promotion of open and FAIR data approaches for the nanofabrication community
- The international biannual meeting on cleanroom management (ENRIS), organized by EuroNanoLab, will also be an opportunity to promote the GO FAIR approach for Nanofabrication internationally
- Documents showing the progress of EuroNanoLab data management activities will be put on the GO NANOFAB website and on EuroNanoLab member's websites to promote the FAIR Data approach.
- The links on the EuroNanoLab webpage dedicated to data management from associated partners will be highlighted

C. Long-term objectives

In the view of an ongoing process of enrichment it will be necessary to:

• DEVELOP CERTIFIED NF DATA MANAGEMENT CENTERS AS RECOGNIZED DATA STORAGE REPOSITORIES

By considering the existing opportunities (mapping of usual actions/providers within the community, etc.), GO NANOFAB will promote the development of certified centers for data collection. These certified centers will become referents for the others. A list of commitments with which any certified center should comply for the collection of process data will be written.

In the longer term we hope to work with other GO FAIR INs to study options for creating a common storage capacity, catalogue and mapping service for the EOSC. For this purpose, based on the work done by other initiatives, it will be necessary to formalize and harmonize the requirements for certified NF data storage repositories to be in-line with other INs.

• PROMOTE APPROACHES THAT FACILITATES EFFICIENT REUSABILITY OF PROCESS STEPS THROUGH INTEROPERABILITY OF GO NANOFAB DATABASES

We will collect and develop use cases demonstrating the gain on process development time through effective reusability of processes and modularization. For example, we will identify recently developed process steps that could have used standard building blocks in other cleanrooms.

The idea is to establish standard elementary process steps associated to "assembly rules" facilitating the development of a validated process chain. For instance, we will try to perform the fabrication once more by using standard building blocks developed elsewhere and compare the required time

• DEFINE DATA TYPES, METADATA CONTENT AND PROCESS DESCRIPTORS TO ENABLE THE INTEROPERABILITY OF PROCESS STEPS AND DATABASES

The GO NANOFAB vision is to encourage exchange and interoperability between sectors and across borders, allowing complex process chains to work efficiently across different sites, minimizing additional developments and speeding up fabrication services for research. This work will take in account legal aspects, principles, depending on the status of users (public, private, etc.).

In the long term GO NANOFAB will work towards improved interoperability of databases (machine parameters, process steps) with data from other INs: we will aim at convergence on metadata content, data types, and process descriptors to enable the interoperability of process steps and databases. For example, we will seek agreement on minimal standards for machine readable metadata, database legal interoperability (licenses), and links with developments or approaches from other INs. For this purpose, we have to:

- Promote dialogue on how to achieve agreement on data type conventions (names, versioning etc) towards establishing an internationally acceptable data model for NF metadata.
- Through pilot studies and use cases from different countries we aim to work towards agreement on common standards for metadata content. These should include a reflection on (example of success indicator: test a limited set of metadata and test the process reproducibility in another cleanroom with the same machine).

GO NANOFAB WORKPLAN: SUMMARY & TIMELINE		START YR.	END YR.
WP1: TOWARDS A HIGH FIDELITY OF NF PROCESS STEPS DESCRIPTION			
	Objective 1.1: identify TEST CASES to study the possibility of obtaining a high fidelity REPRESENTATION of NF processes	0	1
	Objective 1.2: develop a minimal data set for the NF process representation on the TEST CASE	0	3
WP2: CERTIFICATION OF THE TEST NF PROCESS REPRESENTATION			
	Objective 2.1: develop a certification procedure of process REPRESENTATIONS on the test case for a single model of cleanroom machine	0	3
	Objective 2.2: extend the certification of process REPRESENTATION to the whole family of machines belonging to the test case	3	5
	Objective 2.3: on the test case, characterize the validity domain of a process DESCRIPTION	0	5
WP3: STANDARDIZATION OF NF PROCESS REPRESENTATIONS		4	4
WP4. DATA MANAGEMENT ISSUES AND METADATA GENERATION		0	5
WP5: COLLECTION OF PROCESS DATA OR METADATA			
	Objective 4.1 achieve agreements on appropriate Persistent Identifiers (PIDs)	0	3
	Objective 4.2: automate the capture of key process parameters	3	10
WP6: DEVELOPMENT OF RULES AND AGREEMENTS FOR DATA SHARING - 3 yrs			
	Objective 5.1 Establish rules for the sharing of nanofabrication know-how	0	3
	Objective 5.2 Study licensing issues	0	3
W7. ENCOURAGE THE DIFFUSION OF FAIR APPROACHES FOR NF PROCESSES		0	10