

Work package number	3	Start Date or Starting Event				M1
Work package title	Tools for market and impact evaluation.					
Participant number	1					
Short name of participant	ILL					
Person/months per participant¹:						
Start month	1			End month	48	

Objectives:

The objective of this WP is to provide tools for the project and the ILL scientific management to monitor the scientific trends and measure the impact of decisions, actions and more generally scientific work carry out at the ILL.

In particular, these tools should bring global maps of potential fields of research and potential users that could lead to innovative usage. They should also allow to identify laboratories and countries yet unexplored that could open the possibility of new partnerships.

The most precise and accurate results will be provided by studying the ILL related publication in relation to our users and experiments databases, nevertheless regarding the potential fields of research this study needs to cover worldwide related publications and activities.

In order to identify which publication is resulting from which experiment, data DOI references in the publications, as requested to our users, is the most efficient and promising solution. In the same line, the matching of ILL users and publication authors thought the use of ORCID is extremely promising. Unfortunately, and despite of our actual effort, these solutions are still emerging and developing concepts which have not yet fully established in the current scientific practise. WP2 will be working on fostering the data DOI usage, nevertheless we need at the beginning of this project tools to realize these matchings and therefore provide the valuable link between publications and experiments on one side and publication authors and facility users on the other.

The ILL is engaged in Open Access not only for data (the ILL was pioneering with its open data policy released in 2011) but also for scientific articles. A specific budget is requested to ignite and promote immediate access to publication resulting of work carry out at the ILL. We would like to measure the impact of Open Access for a facility like ILL using an innovative process. When our users submit scientific proposal at the ILL they also provide a list of recently published articles that explain and highlight the importance and need of their proposal. By measuring the delay between article publication in journals and first appearance of the same article in the lists that sustain new proposals and by comparing the delay for Open Access articles and to the one of articles publish using traditional (non-open access) model we propose to measure the influence of Open Access in term of scientific workflow acceleration. Therefore, this work will not only evaluate impact via standards metrics such as number of citations or the impact factor of the journals, but also in term of

¹ Except human effort already included in the calculation of the unit costs.

scientific lifecycle (Figure 1 Scientific life cycle) efficiency.

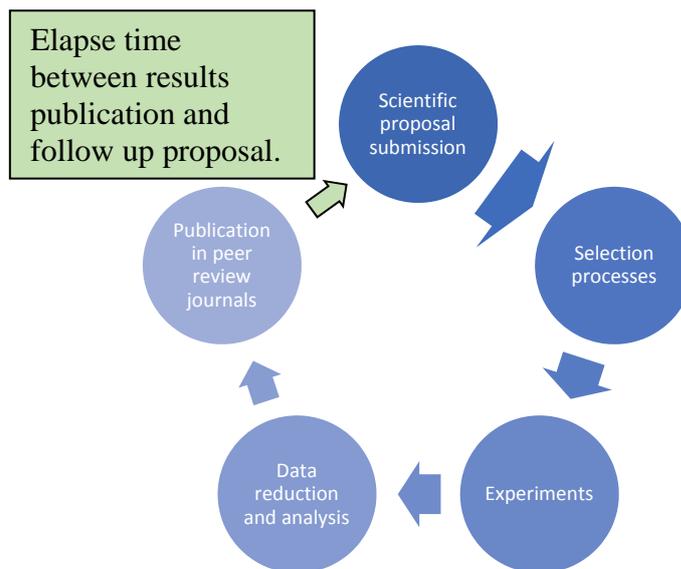


Figure 1 Scientific life cycle

Description of work

Task 1: Building the corpus of publication.

This task aims at harvesting the ILL and all neutron scattering related publications for building the corpus for the investigations of this work package. For publication retrieval, in order to be as complete as possible we will use a mix of open access repositories and databases as well as our subscription to commercial publisher database (web of science) and publishers' subscription so as to get access to full text articles.

For each publication we need to identify:

- The bibliographic references (title, year, journal, DOI ...)
- The authors and their affiliation
- The full text article
- Scientific keywords, which with the journal will enable the classification in scientific area
- The references to other publication or data (through the use of data DOI)
- The type of article (journal article, proceeding, book, thesis)
- In case of journal article, information about licensing model (Open Access model with embargo time, subscription model).

We intend to cover a period of 10 years, which is going to represent between 40 000 to 60 000 documents to be analysed.

Task 2: Set up of the extraction tools.

Depending on the publisher database services, information retrieval is more or less comprehensive, for this reason it is necessary to complete articles information through an automatic analyse of the full text article. In this task we will add to our repository information not always provided by the publishers' services (for example: affiliations, keywords...) and also more domain specific information which could permit to increase the

matching score between an experiment and a publication (for example: instrument, formula and formula consistency...). Specialized libraries like GROBID, CERMINE, ... will be used and analysis workflow developed.

Task 3: Matching publications and ILL related experiments.

In order to produce some of the necessary measures of impact, like for instance the delay between experiment and publication, we need to identify the data and consequently the experiments that are at the origin of the publications. For such purpose data DOIs are the obvious solution, but until their usage are widely spread in the community, we need to set up semi-automatic process to perform this matching. We therefore propose in this task to develop and set up a matching between publications and experiments: for each publication we will identify the most probable experiments based on common information, mainly: team members, sample characteristics, instrument references and bibliographic references. When multiple experiment candidates exist we will contact corresponding authors to validate matching. The matching publication and experiments will be stored in a database alongside information of publication authors/laboratories matching ILL users/affiliations.

Task 4: Analysis tools and visual maps

We will setup, integrate and develop when necessary, business intelligence software in order to provide data visualization tools and interactive dashboards for the project.

The use of such tools should allow easy exploitation of the collected and aggregated data and eventually provide insight such as:

- The laboratories and universities, and therefore the affiliated countries, which produce scientific articles based on data produced at ILL, in collaboration with ILL users but without currently being LL scientific partners.
- Emerging use of neutron scattering techniques in scientific fields (worldwide) and geographical map of different scientific community using neutron scattering.
- Research strengths (or weakness) at ILL in view to reinforce experiments (and publications) in leading edge research, or to put new research areas forward.
- Evolution of the impact of the different neutron scattering techniques.
- Citation and use of ILL data.
- Analyse research trends for new collaborations.
- Time between a proposal submission and the related publication.
- The time between the date of publication of an article and its first citation by another publication
- The delay between the publication of an article and its citation in a scientific proposal submitted at the ILL.

Deliverables (brief description and month of delivery)

Table 3.1 b: List of work packages

Work package No	Work Package Title	Lead Participant No	Lead Participant Short Name	Person-Months	Start Month	End month
				Total months		

Table 3.1 c: List of Deliverables²

Deliverable (number)	Deliverable name	Work package number	Short name of lead participant	Type	Dissemination level	Delivery date (in months)
3.1	Report on the publication corpus and completion workflow	3	ILL	R	PU	3
3.2	Report on the matching techniques and its results	3	ILL	R	PU	12
3.3	Report and demonstration of the analysis tools.	3	ILL	Other	PU/CO	24

² If your action is taking part in the Pilot on Open Research Data, you must include a data management plan as a distinct deliverable within the first 6 months of the project. This deliverable will evolve during the lifetime of the project in order to present the status of the project's reflections on data management. A template for such a plan is available on the Participant Portal (Guide on Data Management).

KEY

Deliverable numbers in order of delivery dates. Please use the numbering convention <WP number>.<number of deliverable within that WP>.

For example, deliverable 4.2 would be the second deliverable from work package 4.

Type:

Use one of the following codes:

R: Document, report (excluding the periodic or final report)

DEC: Websites, patents filing, market studies, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

Dissemination level:

Use one of the following codes:

PU = Public, fully open, e.g. web

CO = Confidential, restricted under conditions set out in Model Grant Agreement

CI = Classified, information as referred to in Commission Decision 2001/844/EC.

Delivery date

Measured in months from the project start date (month 1)