

Deliverables

<i>Deliverable Number</i>	D2.9
<i>Deliverable Title</i>	Integrated User statistics overview
<i>Lead Beneficiary</i>	ELETTRA
<i>Type</i>	Report
<i>Dissemination Level</i>	Public
<i>Due date of delivery</i>	Month 34

Background

The motivation for proposing a deliverable on User statistics stems from previous work to standardize certain information across all European synchrotron and free electron laser sources, during the course of the past nine years of collaborative projects.

During the FP7 CALIPSO project, the consortium succeeded in developing standardized facility and beamline datasheets, which were refined and improved in the framework of project EUCALL under HORIZON 2020, to also include part of the optical lasers. The catalogue is available at the wayforlight.eu¹ portal and is fully searchable and exportable to other websites or platforms.

The original aim of this deliverable was to develop a common set of definitions for reporting usage to the European Commission (hereinafter EC), as well as to the funding agencies of the facilities. This is obviously no simple task, as every facility uses its own definitions and criteria to comply with internal or national regulations, constraints imposed by funding bodies and practical needs.

When the present deliverable was first conceived, the LEAPS initiative² was just about to be launched. The CALIPSOplus beneficiaries all cooperate within the LEAPS initiative; one of the goals of LEAPS Working Group 5 “User services and impact” is the development of common metrics for impact assessment of the facilities.

Within CALIPSOplus, it was decided to combine efforts with LEAPS and, starting from a set of parameters used for the first LEAPS metrics survey performed in 2017, feedback was provided to improve their definitions. The present deliverable starts from the list of

¹ www.wayforlight.eu

² www.leaps-initiative.eu

parameters used by LEAPS in summer 2019, with the addition of proprietary research use of facilities and of some additional proposed changes.

List of parameters

The following parameters were considered to be most relevant and in need of a precise and commonly agreed definition:

- Users
- User visits
- Community members
- Number of proposals
- Number of hours delivered (peer-reviewed)
- Number of hours delivered (purchased)
- Number of publications
- Number of beamlines that can be operated in parallel
- Number of endstations to which beamtime can be requested independently

The “Number of hours” of purchased beamtime and the inclusion of proprietary research in the User and User visits parameters are the only additional parameters, compared to those used in the recent LEAPS survey.

Shared definitions

The definitions of these parameters, as agreed amongst the facilities in June 2019 and now modified to include the aforementioned proprietary research use parameters, are detailed below. There was no unanimous agreement on any of the parameters, however every facility committed to indicate any difficulties in deriving a specific quantity e.g. Remote Users. This information will then be used both to improve the definitions for future data collection campaigns and for changing the way data are produced and categorized at the facilities. The time interval considered in the LEAPS survey was five years, but we propose that all these parameters are collected for a time span of one year and then simply summed up, since this allows more precise following of trends, for example.

Parameter	Definition
Users	All Users that are both: - Academic Users or Industrial Users AND - Active Users
Academic Users	<ul style="list-style-type: none"> • Everyone who participates in an experiment which is based on a peer-reviewed proposal and competing for the beamtime provided for external Users is an “Academic User” and will be counted. This includes national, transnational, EU access Users as well as internal and industrial Users applying via peer-review access routes competing for the same beamtime contingency as external Users.

	In-house research access routes are excluded
Industrial Users	<ul style="list-style-type: none"> • Everyone who participates in an experiment purchased as proprietary research beamtime
Active Users	<ul style="list-style-type: none"> • Users that participated in at least 1 experiment session on-site OR • Users that participated in at least 1 experiment session remotely
Remote Users	<ul style="list-style-type: none"> • Those who were given access to the data until the end of the related proposal's lifetime (depending on how the User portal works) in addition to the on-site Users^(*) • Those who have registered for a remote experiment session • Those who mail-in samples (registered for data access in whatever way from the start until the end of experiment)
User visits	<ul style="list-style-type: none"> • Number of visits Active Users as defined above • of which Remote Users visits (if figure available)
Community members	<ul style="list-style-type: none"> • All Users (on-site, remote, academic, industrial... → irrespective of the access route) <p style="text-align: center;">PLUS</p> <ul style="list-style-type: none"> • All persons listed on any submitted proposal

Table 1: Proposed definition for first set of User Metrics parameters.

The number of Users and User visits were always collected according to **Country of Affiliation** and not the nationality of the User.

(*) There was a debate about whether “Remote Users” should also include those Users who do not participate in an experiment either on-site or remotely but who get access to the data sets produced during an experiment. Many facilities cannot provide these data, and since Users are able to transfer data to any colleague without the facility knowing, the figures given would be imprecise (an underestimation). For the above reasons, it is proposed to exclude this definition from the agreed set of User metrics parameters. This means a proposed change to the definition of “Remote Users” and “Active Users” above.

In previous the previous Framework Programmes FP6 and FP7, as well as in the current programme Horizon 2020, reporting on Transnational Access includes several additional parameters among them:

- nationality of the users
- gender (M/F) of the users

The information on nationality was used for estimating mid-term brain drain and fluxes in the European Research Area, while gender statistics were exploited to monitor the evolution of gender balance over time as well as positive effects of TNA support programmes. To give

an example, supporting two users per experiment was found to result in an increase of the fraction of both young and women researchers.

At present, nationality and gender are parameters which are not shared for integrated User statistics. The nationality of a user needs to be recorded by the facilities in addition to the country of affiliation to enable government sanctions list screening. In general, facilities also collect gender data, but some of them do not keep gender/nationality data for more than three years, meaning that long-term statistics on these parameters would not be available.

Parameter	Definition
Number of proposals	<ul style="list-style-type: none"> Number of total proposals received through the peer-reviewed calls including all types of proposals (regular, long-term, block allocation group, rapid access etc.) Thereof, the number of accepted proposals (have been granted beamtime)
Number of beamtime hours <ul style="list-style-type: none"> requested peer reviewed delivered peer reviewed delivered purchased 	Beamtime hours related to number of proposals above: <ul style="list-style-type: none"> Beamtime requested (per time interval) in number of total hours of beamtime through the peer reviewed calls Beamtime delivered (per time interval) in number of total hours and in addition <ul style="list-style-type: none"> Beamtime hours delivered through proprietary research programmes
Number of publications(**)	<ul style="list-style-type: none"> Scientific articles or reviews in journals listed in the “Master Journal List” of Clarivate (ISI) or in the “SCOPUS List of Titles” of Elsevier or in the Directory of open access journals AND <ul style="list-style-type: none"> Must include data collected at the facility OR <ul style="list-style-type: none"> Must describe instrumentation (including computing tools) developed at the Facility
Number of beamlines that can be operated in parallel (***)	<ul style="list-style-type: none"> Number of beamlines that can be operated in parallel – should coincide with the number of experiments that can be executed at the same time in the same shift using synchrotron / FEL light
Number of endstations to which beamtime can be requested independently	<ul style="list-style-type: none"> Number of endstations (or branchline) to which beamtime can be requested independently, e.g. through choice in the drop-down menu of web-based User Office when submitting a proposal for an experiment

Table 2: Second set of User Metrics parameters and respective definitions.

A parameter which was not considered in the last LEAPS survey is the “Number of experiments / experiment sessions”. For those facilities not offering a significant amount of long-term or BAG proposals this figure almost coincides with the “Number of proposals”, but for other facilities these two figures can be very different. It might therefore be interesting to update the Metrics in this direction and include it in future data collection campaigns, since it correlates more closely with the number of User visits as well as the number of publications, as well as indicating duration of experiments and changing trends in this duration (faster experiments lead to more experiments in the same available beamtime).

It must be noted that the amount of proprietary beamtime can be composed of beamtime purchased both by academic groups and by industry. However, not all facilities can easily differentiate between those categories.

(**) In the LEAPS survey carried out during the summer of 2019, the publications metric included papers published by the facility staff, even if they do not include data collected at the facility. It may be reasonably argued that such papers should count as facility publications since the researcher is funded by the facility, but in the end the majority of contributors felt that since the aim of these particular metrics is to report on facility usage by the community, this point should be excluded from the definition.

Several discussions took place before coming to an agreement on the last two parameters, i.e. the number of beamlines that can be operated in parallel and the number of endstations or branchlines at which beamtime can be requested independently. In the first LEAPS data collection in 2017, only the first one was considered. Although it reflects the number of experiments that can be performed at the same time, this parameter alone cannot describe a facility. This is particularly true when considering both synchrotron sources and free electron lasers: at present, several operating FELs can run one lasing line at a time and therefore perform only one experiment at a time. This is due to the complex fine-tuning of the whole accelerating and undulator sections of the FEL. However they offer access to a range of different endstations for users and this is hidden if the last parameter is not given. To give an example, an FEL such as FERMI in Italy has one beamline working at a time but possesses six different experimental stations, thus offering a wide range of experimental techniques and sample conditions.

At FLASH, another FEL example, two experiments can be run in parallel, while it offers six operating beamlines. Of those, only two beamlines have a permanent endstation and the remaining have open ports. In the FLASH proposal submission process, proposers can choose between beamlines with specific endstations or beam quality (focused/non-monochromatized; unfocused/non-monochromatized; monochromatized beam).

As a consequence, the last parameter was added and the corresponding data were first collected during the summer of 2019. As explained by the FLASH example above, these parameters do not perfectly fit the case of FELs. Subsequent improvements will be proposed in the future.

(***) A minor amendment in the definition of “Number of beamlines that can be operated in parallel” agreed in June 2019 was requested by EuXFEL. At this facility three beamlines, i.e. three scientific instruments out of six, can deliver beam in parallel in the same shift. However, six experiments can run on all six instruments in the same 24 hours because a switch happens at the start of every 12-hour shift. Adding a duration reference to ‘in parallel’ would make the definition clearer.

Discussion as to whether these two statistics are sufficient to fully describe the portfolio of a facility is ongoing.

Usage of data

The reason for this exercise is to obtain a meaningful, reproducible and solid set of User data, enabling a more complete overview of the use and needs of the European synchrotron and free electron laser sources.

Care should be taken when comparing metrics between single facilities, since many other parameters related to techniques, local requirements and constraints can play an important role and are not reflected in the reported metrics.

As an example, a beamline devoted to macromolecular crystallography (MX) is inherently more publication-intensive than a microscopy beamline since MX experiments require less beamtime, i.e. compared to other beamlines, in the same number of shifts, more experiments can be done.

The data are intended to be used to follow trends across the European user community, to show interactions, evidenced for example in joint publication charts, or to showcase the European Research Area with respect to the situation in the U.S.A. or in Asia.

Next steps

The present deliverable is public, therefore once finalized and accepted by the EC Project Officer, it is planned to share it with the communities of other analytical facilities facing very similar issues, such as Laserlab Europe, as well as with the neutron source community. In addition, the document shall be shared with the RIPaths consortium, a project funded under H2020 and dealing with impact metrics for Research Infrastructures.

The results constituting this deliverable will be made available to the LEAPS consortium, recommending their adoption for the next data collection. In addition, it is suggested that the quantity of access provided to proprietary research programmes should be included as a complementary figure.