METRICS Metrological evaluation and testing of robots in international competitions

Deliverable title	D2.2: Good practice guide for the organization of evaluation campaigns
Deliverable lead	LNE
Related task(s)	T2.1 Evaluation framework definition
Author(s)	Agnes Delaborde (LNE) and METRICS WP2 Partners
Dissemination Level	Public
Related work package	WP2 : Methodology
Submission date	June 30th, 2020
Grant Agreement #	871252
Start date of project	1st January, 2020
Duration	36 months
Abstract	The present deliverable aims at defining the organizational procedures that must be taken into account in the competitions. The document is organized in two sections: fairness and safety. This includes for example guidelines for the fair treatment of participants, for a good control of legal liability, or procedures for the safe conduct of competitions.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Versioning and Contribution History

VERSION	DATE	MODIFIED BY	MODIFICATION REASONS
V1	June 2020	Agnes Delaborde	First Version

List of Abbreviations and Acronyms

ABBREVIATION	MEANING

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

1 Objectives

The present deliverable aims at defining the organizational procedures that must be taken into account in the organization of the competitions. In contrast to deliverable D2.1 ("Preliminary methodological framework"), which dealt with good metrological practice in competitions, this deliverable focuses on good organisational practice, which are intended to protect the competitors, the organizers or even the public.

In this regard, the document is organized in two sections: fairness and safety. This includes for example guidelines for the fair treatment of participants, for a good control of legal liability, or procedures for the safe conduct of competitions.

2 Fairness

2.1 Introduction

The notion of fairness is addressed in the Deliverable D.2.1 Common evaluation framework, in the light of metrological considerations. In particular, the D.2.1 explains the importance of resorting to dry-run so as to calibrate the evaluation tools, or to ensure that the test beds are accessible to a large number.

The practice guide also cover these aspects, but the focus is rather set on the participants' perspective than on metrological rigour. This section looks at how to ensure an optimally fair treatment of the participants.

2.2 Simultaneity

In order to ensure the fairness of the competitions, the organizers shall ensure a simultaneity of the evaluation. This will guarantee that participants are evaluated fairly, without giving the opportunity to some of them to adapt their systems in function of the other's results. This notion of "simultaneity" can be spread across the one or two days of the competition.

The tolerance level about what may be considered "simultaneous" must be discussed on a case-by-case basis. It is unlikely that evaluations can be carried out for all teams at the same time, particularly when the evaluation involves the use of physical and human resources. A test bed may not be used by several teams at the same time, and the organising committee should have sufficient human resources to evaluate all teams at the same time.

One option for maximizing the simultaneity of evaluations may be to have several teams compete on the same field, if the field is divided into sections (one section to evaluate one task, another section for another task, etc.), and rotate the teams. If this is not possible (e.g. the whole environment must be used for each assessment), then the teams must pass one after the other without delay between teams. This means that the timing of the evaluations must be organized without any gaps.

As noted in the 3.3 (Testing environments) of the D.2.1 Common evaluation framework, "any outdoor setting will never be completely repeatable: clouds change in the sky, waves and tides modify the visibility underwater, etc.". This lack of repeatability, in addition to its impact on the metrological rigour of the evaluation, has an impact on the fairness of the evaluation between competitors: it is not conceivable that one participant will have to compete in pouring rain, while another participant will suffer from maximum sunshine. In this regards, the organizers shall define thresholds and limits in several parameters that are considered as influencing the performance of the devices. Outside of this acceptability range, the organizers shall define remedial strategies to have teams compete in reasonably similar conditions.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

2.3 A priori ignorance

This aspect only concerns evaluation tasks where the intent is to assess the robustness of the algorithm to novelty: for example, a change in colors or shape of the object to grasp, or an environment with a different map, or unexpected obstacles.

Evaluated systems may have a learning capacity and as a consequence, they should not have a priori knowledge of the testing environment (testbeds and testing datasets) used for the evaluation in order to avoid measurement bias. This remark remains valid for systems that do not have learning skills, since developers can influence the design of their systems if they have a priori information about testbeds and data.

In addition, the participants may be offered to perform the tasks several times, with time in between to adapt their algorithms (e.g. to estimate their ability to improve their performance past the first discovery). In this context, the participants must be able to benefit from a substantially equivalent adaptation time between all the teams. For example, the organization must not allow that there is not one team benefiting from just one hour of adaptation, when another would have a full night.

A fair treatment thus implies that participants are given the same amount of information at every stage of the evaluations, with an equivalent amount of time for each phase, including the adaptation times.

2.4 Impartiality

The evaluation must be carried out by a "trusted third party". This trusted third party will have metrology expertise applied to the evaluated systems, to develop an evaluation protocol common to all participants. In addition, the third party must guarantee that there are no conflict of interest between the organization team and the competitors.

It is not appropriate to impose a ban on the members of a laboratory that is part of the organization committee. Indeed, for example, the committee may include large laboratories whose research units are also developing the type of system being evaluated. In this likely case, it is recommended that the organizers from the competing laboratory are not part of the evaluation committee: no participation in any voting, and no refereeing on the field. It is also recommended, naturally, that the organizers do not communicate information outside the scope of the public evaluation rule book.

2.5 Affordability

This concerns:

- Cost of the devices:
  - If the device has to be built by the competing team, it shall not include elements of a cost that would be unaffordable by the participants. In concrete terms, this means that when the organisers set the technical characteristics expected of the robots, or the features of the tasks to be performed, they shall make sure that this will not compulsorily require the purchase of expensive elements such as costly sensors or effectors. A quick analysis of the elements available on the market, or on the feasibility of a realization by the team itself, must be carried out.
  - If the competition relies on standard platforms that must be used and adapted by the competitors (a specific robot, sensors, etc.), the organizers must at most allow the competitors to benefit from preferential rates, or ensure that the competitors have the possibility of carrying out their developments in their laboratory on other platforms with similar characteristics, and that they can on the day of the competition use a robot placed at their disposal, with sufficient time for the adaptation of their algorithms.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

- Accessibility of the competition – It is expected that, in general conditions, the competitors provide for their own travel expense and potential registration fees. The organizers must however look for the best way of making them affordable to a large number: encouraging locations where accommodation and catering are available in different price ranges (e.g. avoid tourist places in high season); adjust the fees according to the rates usually charged for this type of event.

2.6 Strategy for open dissemination and participation

Organizers shall ensure that the right to competitiveness of industries and research organization are guaranteed in the participants. In this regard, the call for participation must be broadcast as largely as possible, and needs to reach as wide an audience as possible. The organizers must set up a communication plan that defines all the networks that must be reached, including notably mailing lists for national and European research groups, industrial networks, etc.

The rules for participation must be reasonably open to any type of participant: industrial, academics, students, etc. The notion of "reasonable" may, however, be conditioned by certain factors such as the presence of principals sponsoring the competition and wishing to encourage technological innovation in a product. However, this should not condition the nature of the participants, so as not to restrict competitiveness and innovation; rather, the rules of participation may focus on the level of TRL aimed at, thus allowing any participant capable of reaching such a level of maturity to take part.

3 Safety

3.1 Introduction

The section covers all aspects linked to the direct management of safety during the competitions, but also covers more broadly the issue of the legal aspects that may arise during competitions. Here, the objective is to ensure that the competition organizers cater to the protection of all the actors of their competitions. Additionally, the organizers shall also make sure that their own legal liability is adequately covered by taking certain reasonable measures.

3.2 Risk assessment

A proper risk assessment should be carried out in order to prepare the competition in the best safety conditions. One must note that the liability of the organizers may be engaged if a damage occurs that could be avoided through a proper risk reduction strategy. This risk assessment concerns the organization and conduct of the competition itself, and considers the damage that may occur due to the competition activities. The conclusions of the risk assessment shall guide the decision to implement risk reduction strategies.

The risk assessment should follow the procedure detailed below.

3.2.1 Identification of the phases and actions of the competition

The organizers shall proceed to the identification of the phases and actions of the competition that involve actions or events that may cause damage. These phases and actions must be tightly linked to the competition. Three phases have been identified:

- **The preparatory phase** – When the competing teams prepare for the test (this may also be between tests). This implies on the part of the organisers that they may provide stands, tables, areas for the teams to prepare for the competition, which must be safe. In this location, the competing teams may assemble and test their device.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

- **The execution of the test** – When the competing teams are actually performing the test actions on the test bed.
- **Teams repacking** – When the competing teams are clearing the preparation area at the end of the evaluation.

In order to know if the considered action is in the scope of the risk assessment, one may ask this question: "*If the competition had not taken place at that moment and in that location, would that damage still have occurred?*" If the answer is "no", then the action should give rise to a risk analysis. The Table 1 provides some examples of competition actions and how they can be addressed.

Table 1: Identification of the competition actions that should lead to a risk analysis (example)

Considered phase	**Linked to the competition**	**Organizer's potential liability**	**Proceed to a risk analysis?**
Travel of the challenging teams to the competition site. *Comment : Organizers cannot be held responsible for such an outside event.*	Yes	No	→ No
The teams tune their robot onsite the day before the test. *Comment : Organizers must offer safe conditions onsite.*	Yes	Yes	→ Yes

3.2.2 Defining risk reduction strategies

For each "risky" action considered, the organizers are expected to define:

1. **The source of the damage (who/what causes the damage?)**
   - The competitors
   - The robot
   - The public
   - Any other entity
2. **The target of the damage (who/what is harmed?)**
   - On individuals:
     - On the competing team itself
     - On the organizing team
     - On the public
     - On external individuals
   - On goods:
     - On the competing team's own equipment
     - On the other competing team's equipment
     - On the test bed
     - On the building hosting the competition
     - On other structures or equipment
3. **The likelihood of the damage occurrence (will it happen?)**
   - Low (1) - It is not really likely to happen

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

- Average (2) - It can happen
- High (3) - It will almost certainly happen

4. **The criticality of the damage (what is the impact?)**
   - Minor (1) - Does not require any renovation work, nor specific medical care
   - Average (2) - Will require renovation work or medical care
   - Major (3) - Irreversible destruction or major injuries/death

The calculation of the values of *likelihood* x *criticality* can provide an estimation of the strength of the damage. This strength may condition the nature of the risk reduction strategy to engage:

- **By warning** - Mainly for weak damages, or in addition to other risk reduction strategies for stronger damages. This consists in warning of the potential danger, for example by placing signs.
- **By additional protection** - Mainly for damages that cannot be prevented "By design", but are still strong enough to require an imperative risk reduction. For example, placing barriers to prevent public from coming too close to robots.
- **By design** - Mainly to prevent strong damages. This means that the configuration of the settings itself will almost never make it possible to trigger the possible damage. An extreme example can be: if one considers that a barrier can be climbed over and is not enough (for example if the accident will surely be lethal), one can decide to perform the evaluation in a separated room and to broadcast it remotely on a screen to the public.

3.3 Applicable regulative requirements

The good practice guide does not require that organizers make a proper assessment of the compliance of the competitors' robots. However, the evaluation tasks must not encourage the recourse to implementation strategies that may be in violation of the law and/or cause a damage. It is recommended that the organizers are aware of applicable European safety directives, among which, for example:

- Directive on Machinery (2006/42/EC) that sets the essential health and safety requirements for the placing on the market of machinery (including robots);
- Directive RoHS (2002/95/CE) (Restriction of hazardous substances in electrical and electronic equipments) that restricts the use of hazardous substances;
- Directive on Electromagnetic compatibility (2014/30/EU) (EMC) that rules on the limitation of electromagnetic emissions in order to ensure the absence of disturbance, etc.

The organizers should also consider regulation such as the General Data Protection Regulation (EU 2016/679) (GDPR) that sets the rules for data protection and privacy. Additionally, they shall make sure that the competition settings do not violate local regulations, for example concerning the autorized aerial areas for drones.

The organizers shall also note that there are regulations for public demonstration, exhibition and trade fairs that apply to machines, in particular concerning the EC marking. A regular infringement seen at robotics fairs, due to a lack of knowledge on the part of the organisers and designers, concerns the presentation of prototypes which do not meet the characteristics of machines which can be placed on the market in Europe. The Directive on Machinery however cites: "*For trade fairs, exhibitions and such like, it should be possible to exhibit machinery which does not satisfy the requirements of this Directive. However, interested parties should be properly informed that the machinery does not conform and cannot be purchased in that condition.*" The applicability of this statement should be considered by the organizers; if appropriate, the organizers must make sure that disclaimers are displayed.

Concretely, this means that the organizers must have carried out a quick search for regulations that could potentially be applicable in the field of the systems expected in the competition. As it is not possible to

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

anticipate all the implementation choices made by the participants, and as the legal liability of the organizers should only be engaged to the extent of their actual control capabilities, the organizers can simply require from the participants a commitment to comply with the regulations in force, optionally providing examples of applicable regulations they have identified, or obvious design flaws that could be detrimental to safety or law, in order to guide the participants.

In any case, the organizers must remind each participating team that the latter assumes its own share of legal liability.

The disqualification of a team whose robot may present a serious offence that may endanger the safety of the organizers, other teams, the public, and property, must be considered by the organizing committee during the creation of the evaluation plan, and must be stated in the written materials provided to the participants.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252