Room 2 - Brainstorm Challenge Call & ERA-NET

Brainstorm Questions

What type of Challenge format

What make a topic convenient for a challenge call

What requirements for successful evaluation campaigns?

Part I - Solo Brainstorm - 10 Minutes

Practice Area

Part II - Group Brainstorm - 20 Minutes

Take on "RoCKIn's Legacy" and stick to the tried-and-tested (and effective) methodology of having Task Benchmarks and Functionality Benchmarks, where the FBMs involve functionalities used by the TBMs (not necessarily ALL of them)

the same challengers across the years help for a better accompaniment of the competitors and a better adaptation of the evaluation protocol to the technological solutions

Relationship: organizers are consulted in the competitors' selection, to assess the measurability and comparability of the solutions, organizers are conflict solvers for the competitors on technical matters, and intermediary with the funders for other types of conflicts in the competitors' consortium. Organizers report periodically to the funders about the progress and highlight the traceability and fairness of the evaluation.

the first year is a dry-run (prepare and test the evaluation tools), then the official evaluation campaign starts

The key point is actually the funding of the participating teams.
If there is little or no funding, except maybe some from sponsors (such as in the European Robotics League) teams mainly come from universities, and involve mostly students. This means that approaches are often cutting-edge but implementation is generally not. It is VERY difficult to involve companies. On the other side, it is easy to foster collaboration between them.
If teams are paid (such as in ROSE) for their participation it is much easier to involve companies, but approaches are likely to be more conservative.
Collaboration is difficult to get since companies are thinking of how they can transform their experience into a product and do not want to disclose too much, or help their competitors.
If payment takes the form of (large) prizes to the best teams, approaches tend to be cutting edge again, but collaboration between teams is ruled out since competition is fierce.

Addition to the observation of the left: if budget allows it, organising parallel tracks of the challenge may be a way to foster "cross-pollination" (especially between research and industry). However, this makes organisation even more complex.

first year evaluation is based on comparison of submitted tools architecture , and the second year evaluation could be based on tool performamces

Challenges that are dedicated on topics where evaluation of performance is in itself a topic of importance (economic reason, regulation, etc.). E.g. with explainability of artificial intelligence: the assessment of XAI performance is a research subject. An evaluation campaign allows to refine the assessment strategy (metrics, indicators, etc.) and to compare their usability among different solutions.

The evaluation campaign is linked to the topic definition, since the topic must cover solutions that are: available (maturity is high enough to have solutions that produce relevant enough results, and there must be an existing community of developers/manufacturers of the solution to ensure that competitors will join - such as SMEs and academic labs) ; and measurable (in some cases the subjectivity and/or complexity of the systems is not fully suited to evaluation, plus some results may only be observed with a general and long-term observation, hence the need of a topic selection that only covers aspects of the system that can be rigorously assessed in the lifetime of the Challenge)

An effective approach is to try and propose an early version of the challenge to stakeholders, involving several alternatives, and select among these according to the feedback on the early version. This ensures that the challenge is of interest to the stakeholders and relevant.

Topic definition must be performed with potential evaluators/organizers, and with representatives of the users of the solutions (to express their needs), and with developers/manufacturers (to understand the potential limitations)

Resources: Budget depends in particular on the type of infrastructure involved: physical testbenches or not. In the case of a Challenge on XAI, only data would be required. In the case of the existing Challenge ROSE in agriculture, plots of land are required. Concerning time, organization requires a first "testing" of the evaluation tools, that is formalized : the first occurrence of the evaluation is a dry-run, results are not relevant for competitors but highlight the capacity of the organizers to perform a relevant evaluation in the following official campaign.

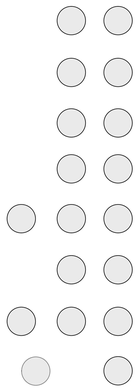
a key issue to consider (not only in terms of costs) is long-term sustainability of the challenge. A multi-year challenge has more appeal and relevance, but requires continued availability of funds. If these are provided by a project, it becomes crucial to create links with sponsors, which can also be public bodies. This must be an integral part of the design of the competition.

Specifications : organizers must prove their independence and absence of conflict of interest. Must prove their ability for quality management, confidentiality.

To facilitate organization: availability (in the organizers' consortium or provided by the funder) of experts on the topic. Evaluation campaign organizers may not be expert of the topic itself, they can be experts in metrology and in performing inter-laboratories comparisons. See for example the French ANR Challenge ROSE project, where LNE (metrology and campaign organizer) co-organizes with INRAE (technologies for agriculture and agronomy)

the complexity of organising a challenge should not be underestimated. This is especially true if teams are not funded, so control over them (and their actual participation) is low or null.

Part III - Vote



Organisation of a challenge is in itself a challenge, especially if participants are not under control of the organisers (such as when they are not funded)