Practical issues: Incentives, community engagement and costs

Magali Richard MAGALI.RICHARD@UNIV-GRENOBLE-ALPES.FR
TIMC
UMR 5525, Univ. Grenoble Alpes, CNRS
F-38700, Grenoble, France

Yuna Blum YUNA.BLUM@UNIV-RENNES1.FR
IGDR
UMR 6290, ERL U1305, Equipe Labellisée Ligue Nationale contre le Cancer, Univ Rennes, CNRS, INSERM
Rennes, France

Justin Guinney JGUINNEY@UW.EDU
Tempus AI, Inc.
Chicago, IL 60654, USA

Gustavo Stolovitsky GUSTAVO.STOLO@GMAIL.COM
DREAM Challenges
New York, NY, USA

Adrien Pavão ADRIEN.PAVAO@GMAIL.COM
Université Paris-Saclay, France

Reviewed on OpenReview: `https://openreview.net/forum?id=XXXX`

Abstract

Each organization of competitions and benchmarks involves a large number of practical problems, such as obtaining sufficient financial support or recruiting participants through appropriate incentives and community engagement. In addition to defining scientific tasks, preparing data and creating challenges, a very important practical administrative organization remains to be achieved. Indeed, cost assessment, corresponding requests for financial support and adequate publicity are key factors for successful organization of the competition. In addition, a good understanding of the incentives that lead participants to engage in a given challenge is fundamental for effective practical organization success. In this chapter, we will cover these topics and give some practical tips and examples for overcoming the "challenge" of organizing the challenges.

Keywords: practical issue, cost, publicity, management

This chapter provides a comprehensive guide to organizing successful scientific competitions, addressing both strategic and practical aspects of challenge organization. We begin by exploring participant motivations and incentives, offering insights into what drives researchers, students, and professionals to engage in scientific challenges. The chapter then delves into community building and outreach strategies, detailing effective methods for recruiting participants and disseminating challenge results within the scientific community. The final sections address the practical aspects of challenge management, including detailed breakdowns of financial and human resource requirements, along with guidance on securing funding sources. Throughout the chapter, we provide concrete examples and actionable recommendations drawn from successful competitions across various scientific domains.

The recommendations presented in this chapter stem from a multi-faceted approach to understanding competition organization. While our primary insights derive from extensive practical ex-

Figure 1: The incentives for participating in a challenge.

perience in organizing scientific competitions, we have strengthened these empirical observations through systematic analysis of documented outcomes from past competitions across various scientific domains. This analysis is complemented by structured feedback collected from both previous participants and experienced organizers, providing valuable perspectives on what contributes to competition success. Furthermore, we have aligned our practical recommendations with current research in the field, particularly regarding best practices in data handling, participant engagement, and competition design. This combination of practical experience, documented evidence, and academic research provides a robust foundation for the guidelines presented throughout this chapter.

1 Incentivizing participation

How to incentivize participants to work on complex problems is a key feature of challenge organization. In this section, we review various types of motivations (Figure 1), from a participant perspective.

1.1 Skills: Knowledge acquisition, communication, education

Traditional university programs in Artificial Intelligence are evolving rapidly, trying to meet the new needs of students, especially on their ability to work collaboratively while improving their scientific knowledge on data mining. Data challenges are mainly based on a coopetitive model, which has the advantage of responding to this dual motivation. Coopetition ((Brandenburger and Nalebuff, 2011) is an active learning pedagogical approach based on the combination of a strategy of competition, where students compete for the best result, and cooperation, where students collaborate for a mutual benefit. Coopetition-based data challenges have the advantage of simultaneously offering two types of learning. On the one hand, this gives a participant a solid methodological training on the scientific question addressed, thanks to the sharing of knowledge between professors and students, but also between the students themselves. On the other hand, these approaches allow students to acquire new skills in collaboration, communication and networking. For more details, please refer to chapter 9: Competitions and challenges in education.

Educational data challenges can be organized into teamwork, recruiting participants from different backgrounds (academic and cultural), with a scientific preparation that can range from minimal information about the challenge before starting to full preparation through a series of dedicated conferences. To meet the expectations of the students, a key factor is the will of the organizer to build a "friendly environment" which will help to boost the motivation of the students and their self-esteem, and to focus more on the process itself than on the results and objectives. Building multidisciplinary teams with different scientific expertise and focusing on real problems are important aspects in the organization of educational challenges. It is also important to provide an environment where participants can communicate with their team members, other teams, and teachers. Ultimately, setting the right reward and price is a major motivator for winning student buy-in (Abernathy and Vineyard, 2001).

Finally, organization of competitions itself can be used as a pedagogical tool. Designing such task is complex and can be, in some regards, more interesting than solving it (Pavao et al., 2019).

1.2 Hot topics: Scientific crowdsourced benchmarking

The quintessential challenge revolves around an existing quantitative standard or benchmark, and seeks to improve upon state-of-the-art. One of the more longstanding benchmark initiatives is the Critical Assessment for Structural Proteins (CASP) that asks participants to predict protein structure (folding) from protein sequence. Groups who specialize in this domain are naturally incentivized to compare their approach in the structured and objective format of a data challenge in the hope that their method out-competes other approaches and can therefore become a new standard in the field (Bender, 2016). CASP is now recognized within the protein structure community as the *de facto* forum for assessing algorithms, and is therefore as much an incentive as a mandate for formal recognition with the community. This incentive generalizes to all specialties, including image recognition (e.g. MNIST (Madry et al., 2019), ImageNet (Russakovsky et al., 2015)), gene identification and function prediction (e.g. RGASP (Steijger et al., 2013), CAFA (Radivojac et al., 2013)) or translational research in biomedicine (Saez-Rodriguez et al., 2016).

Any published AI algorithm is expected to include a formal performance comparison against state-of-the-art methods. No good data-driven approach could emerge without good quality, well curated data. This task can be cumbersome and require a great deal of work to assemble and prepare benchmark datasets. Depending on the type of data, data acquisition and/or generation can be very time-consuming and costly (see cost section below). Consequently, a natural perk of a scientific data challenge is that the work involved to generate and prepare a benchmarking dataset is managed by the challenge organizers. Therefore, AI competitions offer a playground with data that are usually costly and complicated to generate. Access to high-quality datasets in machine learning remains an ongoing challenge (10.1007/s00778-022-00775-9). We believe that providing access to such datasets serves as a strong motivation for participants seeking to develop cutting-edge methodological approaches to address complex scientific problems..

Recurrent challenges also present the advantage or keeping people on a regular schedule, as they expect the challenge to come and reserve time for it. As foor a classic scientific event, it provides participants the opportunity to expand their professional network and to start new collaborations with people working in the same research field or people from different disciplines gravitating around the same topic. Finally, data challenges remain the best functioning way of implementing coopetitions: people compete and get credit for winning, then they share their solution publicly and the community can move together to the next step.

1.3 Environment and awards

One appealing aspect of the challenges is the spirit of games. This translates into a friendly yet competitive environment along with rewards. It is not unusual to gather common participants on different challenges. A passion to participate in this type of competitions can develop, along with the excitement of witnessing the evolution of the social community, particularly on commercial platforms like Kaggle. The rewards can be of various nature going from small prizes (e.g. book) to high amounts of money (e.g. 1 million dollars, Salesforce 1 Hackathon) or even positions in companies. Large awards may naturally attract more participants, but this must be balanced with the context of the challenge and the scientific problem being addressed. In other words, factors such as feedback, non-monetary recognition, and opportunities for knowledge advancement should also be considered.

1.4 Visibility, career and recruitment

Challenges are opportunities for participants to showcase their various skills to recruiters and even get a position at the end. A growing number of organizations are adopting modern hiring practices such as challenges to find best candidates. Recruiters use this tool to assess candidates' technical and behavior skills. Challenges have indeed the great advantage of evaluating many different criteria at the same time. Companies can assess technical competencies such as problem solving skills, time management and innovations. They can also assess the behavioral skills they value, such as communication, openness to diversity and leadership.

The implementation of a challenge allows recruiters to define certain expectations towards the evaluated candidates (candidates gain insight into the work culture of their future employer), while verifying if their personality corresponds to the company's fundamental values. One of the difficulties in recruitment is that many companies still follow long selection processes that waste time and interest for both candidates and recruiters. To overcome this problem, challenges can be used to evaluate candidates in a short period of time and a friendly environment, where they can demonstrate real-time expertise. It can also serve as a pre-selection process that will also save time for recruiters.

Interestingly, challenges can bring together a larger number of candidates from more diverse backgrounds than traditional recruiting. Organizers can build a portfolio of interesting candidates for present and future positions, without necessarily limiting themselves to the winners of the challenge. For instance, Kaggle, one of the leading challenges platform acting as a recruiting tool, usis a performance tracking system to evaluate participants[1]. Some companies even sell expertise from Kaggle Grandmasters[2]. Besides, challenges are also an excellent way to increase brand awareness. They can be used as a marketing tactic for big companies to reinforce their leadership in their field. Smaller companies can also increase their visibility though challenges and attract more applicants for a position.

Finally, in addition to recruiting new talent, challenges allow companies to bring innovative solutions and ideas to technical problems. Based on the clear success of challenges in the recruitment process, we can easily expect their increase in the upcoming years.

1. `https://www.kaggle.com/progression`
2. `https://h2o.ai/company/team/kaggle-grandmasters/`

Practical tips and resources to optimize incentivization

- Define your working plan and your objectives[a]
- Carefully prepare benchmarking datasets (see Chapter 3 on data preparation).
- Set up a website to collect a list of interested people[b].
- Bring together an expert steering committee
- Provide good educational material together with the challenge (i.e. a good starting kit, white paper).
- Make yourself available during the challenge to answer questions.
- Be responsive to questions on the forum.
- For a recurrent challenge, provide open-source previous winning solutions.
- Organize good publication venues (see details and examples in section 12.2 Community engagement)
- Associate with established conferences (see details and examples in section 12.2 Community engagement)
- For education challenges, you can find inspiration on existing education challenges on open-source platforms such as RAMP[c], or Codalab[d]

a. 10 tips here: http://www.chalearn.org/tips.html
b. see e.g. https://l2rpn.chalearn.org/
c. https://ramp.studio
d. https://codalab.lisn.upsaclay.fr/

2 Community engagement

Mechanisms for engaging and disseminating a competition towards a targeted community are complex and highly dependant on the scientific field. In this section, we try to review general aspects of community engagement that could help challenge organizers to properly define their strategy. See Figure 2 and Table 1 for a review of community engagement strategies and examples of recent competitions.

2.1 Organization of the challenge

The community that will engage in a specific competition will depend on several key aspects of defining the challenge. First, the organizers should define an optimal number of participants and implement the maximum number of participant (if any). Large open competitions have the advantage of ensuring visibility and optimizing scientific production (in the case of crowdsourced benchmarking for example) while smaller competitions will promote communication between participants (more adapted to challenges aiming at educational results). Then they have to determine

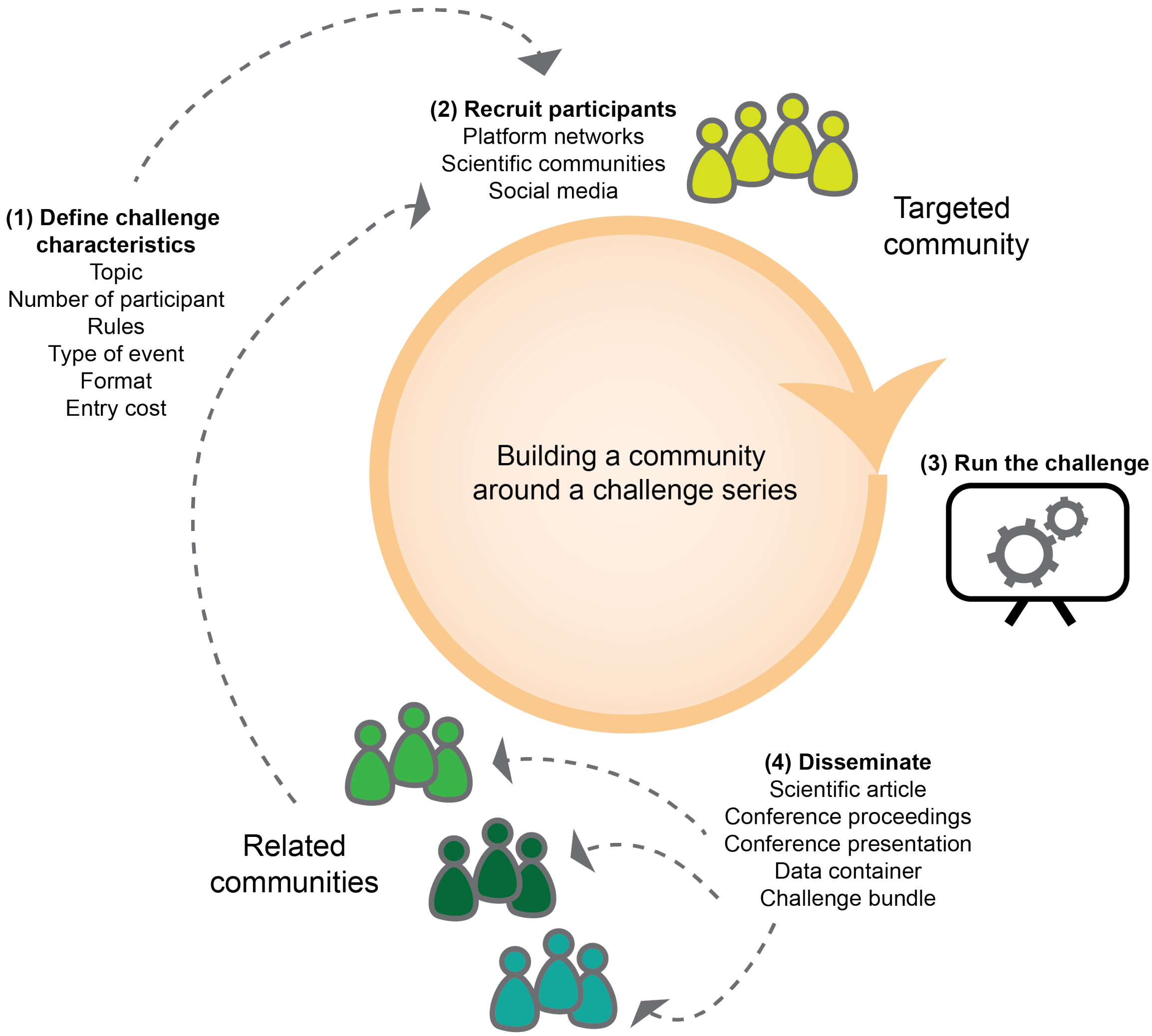

Figure 2: The process of engaging a community

an entry cost: is it easy to participate in the competition? The entry cost depends on several factors: clarity of the rules, specificity of the tasks, size of the dataset, computational resources required to run the methods... All of this will have an impact on the participants who will enter the competition and indirectly define the target audience. Finally, the organizers should established what the format of the competition will be: online events will increase the chances of getting a large pool of participants while in-person events (e.g. at dedicated schools or at scientific conferences) are more suitable for collaborative team work. Once all these parameters are specified, organizers can adapt their communication strategy accordingly and start communicating through dedicated channels, such as the scientific communities mailing list, the digital challenge platform networks and the social media.

2.2 Ensuring diversity and inclusion

A crucial aspect of organizing scientific competitions is ensuring Equity, Diversity, and Inclusion (EDI). Challenge organizers must proactively work to make their competitions accessible and attractive to participants from diverse backgrounds. This includes considering participants who may be traditionally underrepresented due to their gender, race, socioeconomic status, or neurodivergence. Practical measures include offering flexible participation options, such as remote participation possibilities and adjustable deadlines to accommodate different time zones and work constraints. Financial accessibility should be addressed through measures like reduced registration fees for students and participants from low-income countries, or travel grants for in-person events. Ideally, the competition platform and documentation should be designed with accessibility in mind, ensuring compatibility with screen readers and providing materials in multiple formats. Additionally, organizers should establish clear codes of conduct and communication guidelines that promote respectful interactions and create an inclusive environment. The selection of challenge topics, datasets, and evaluation metrics should also be examined for potential biases that might disadvantage certain groups. Building a diverse organizing committee can help identify and address potential barriers to participation early in the planning process. Regular feedback from participants about accessibility and inclusion can help refine these measures over time.

2.3 Challenge output dissemination

The dissemination of the data challenge can take several formats (complementary and not exhaustive) and should match the following question: how would it serve the targeted community?

Participatory benchmarking competitions generally result in scientific publications (see examples (Creason et al., 2021; HADACA consortium et al., 2020; Marbach et al., 2012; Eicher et al., 2019; Marot et al., 2021; Le et al., 2019)) which will be of use to the community. Offering authorship to competing teams, along with participation in manuscript design and writing, is also a strong incentive that will provide international visibility and recognition to participants. Organizers might try to connect with high-profile journal editors ahead of the challenge organization to discuss the possibility of publishing the competition outcome. Depending on the scientific field of the competition, publications can take various form, such as scientific articles, contributions to special issues, conference proceedings, or even books. Best performing teams can also be offered the ability to present their solution in international scientific conferences (e.g., since 2008 all best performing teams in the DREAM Challenges present at the yearly "RECOMB/ISCB RSGDREAM" conference). In addition to an article describing the results of the competition, a challenge built on the data to modeler model (Guinney and Saez-Rodriguez, 2018) could also result in publishing the benchmark dataset along with a container providing a reproducible and continuous benchmark (e.g. a dedicated docker container). Competition data can then be re-used by research scientists as gold standard for new computational methods that will be developed in the future. Challenge organizers may also consider giving open access to their challenge design and templates, especially regarding educational challenges, so that these competition can be massively disseminated to various universities at no cost.

Challenge output and dissemination strategy differ a lot according to the competition organizers and environments. Academic competitions massively rely on the open science framework, encouraging participants to submit their code under an open source license (ex: L2RPN, DREAM challenges). On the opposite, private companies are often motivated by solving an theoretical and methodological obstacle in order to further develop private commercial solutions that will be put on the market. Such organizers may be more inclined to follow a 'private output' model where participant surrender intellectual property of their findings in exchange for earning money prizes.

COMMUNITY ENGAGEMENT					
Name	Field	Year	Platform	Number of participants	Dissemination
TrackML Particle Tracking Challenge	Physics	2018	Kaggle	739 participants	IEEE WCCI competition (Rio de Janeiro, Jul 2018) and NIPS competition (Montreal, Dec 2018)
LAP series	Computer Vision	2013-22	CodaLab	more than 300 teams	Springer Series on Challenges in Machine Learning, ECCV, IEEE TPAMI, JMLR, IJCV, PAA, CVPR
Tumor Deconvolution	Health	2019-20	DREAM	38 teams	2019 RECOMB/ISCB Regulatory and Systems Genomics, Nature Communications
AutoDL series (6 competitions so far)	Automated ML	2019-21	CodaLab	more than 300 teams	ECML/PKDD, ACML, NeurIPS, IJCNN, WAIC, IEEE TPAMI
Digital Mammography	Health	2017	DREAM	126 teams	RECOMB/ISCB Regulatory and Systems Genomics, JAMA Netw Open.
L2RPN	Energy	2020	CodaLab	more than 300 participants	NeurIPS, ArXiv
Challenge AI for industry	Aeronotic	2021	CodaLab	10 teams	
HADACA series (3 competitions so far)	Life sciences	2018-24	CodaLab	150 participants	BMC bioinformatics, JOBIM

Table 1: Table of communities engagement

As a complement, a non-exhaustive list of conferences that have call for competitions, or can offer workshops and/or proceedings, as well as journals that can welcome competition result publication :

- *Conferences and workshops*: ESANN, ICMLA, WCCI (IJCNN, CEC), ECML/PKDD (Discovery challenges), KDD (KDD cup), CVPR, ECCV, ECML/PKDD, ICPR, ICDAR, IEEE international conference on big data, IEEE International Conference on Automatic Face and Gesture Recognition (FG), ACM SIGIR Forum, NeurIPS dataset and benchmark track, NeurIPS competition track, Workshops @ NeurIPS, ICML, AAAI, CVPR, ICCV, Workshop on Semantic Evaluation, etc.

- *Book series*: CiML Springer series, etc.

- *Journals and pre-prints*: International Journal of Forecasting, International Journal of Information Retrieval Research (IJIRR), IEEE Journal of Biomedical and Health Informatics, IEEE Access, Machine Vision and Applications, IEEE TPAMI, Nature methods, Nature com-

3 Costs, human labor and resources

Depending on the model chosen by the organizers, various costs will be associated with a competition organization. To mitigate the problem of financing a competition, diverse sponsors, private companies or academic institutions can be involved. See Figure 3 and Table 3 for a review of costs and resources associated with recent competitions. Complementary to this section, "Chapter 2: Challenge Design Roadmap" offers guidelines and case studies for developing a robust plan for challenge design.

An example of challenges costs: the L2RPN challenge / NeurIPS 2020

- **Research field**: Energy and environment.
- **Challenge platform**: Codalab[a].
- **Duration of the challenge**: 4 months.
- **Number of participants**: 300.
- **Data generation, access and curation: costs and resources description** : 70,000 euros.
- **Challenge engineering: costs and resources description**: 120,000 euros.
- **Challenge design, scientific expertise: costs and resources description**: 170,000 euros.
- **Prices, travel, conference organization (approximate evaluation of costs)**: 30,000 euros.
- **Challenge governance (cost evaluation of legal, ethics and data privacy costs)**: none.
- **Dissemination**: RTE, Google Research, University College of London, EPRI, IQT Labs. Chalearn.
- **Sponsors**: PMLR[b] & ChaLearn[c]

a. https://competitions.codalab.org/competitions/25426
b. https://arxiv.org/abs/2103.03104
c. https://l2rpn.chalearn.org/

3.1 On overview of the requirements and associated costs

PLATFORM AND REGISTRATION SYSTEM

Several digital platforms can support challenge organization (see chapter 5 for different models of challenge platforms). Defining the platform should be a starting point of challenge organization, as open-source projects such as CodaLab or commercial challenge platforms such as Kaggle will provide very different resources (technical support, engineering manpower dedicated to the compe-

Figure 3: Costs of data challenge organization. Pictures adapted from open work on freepik: macrovector, alvaro_cabrera, visnezh & vectorjuice.

tition...) and associated costs. Please refer to Chapter 5 for more details on the different services provided by each platform.

Data generation, access and curation

High-quality, well-curated data is fundamental to competition success. Recent research, particularly the work of Mougan et al. (2023), has provided comprehensive frameworks for data handling in scientific competitions. A high-quality dataset requires careful planning across multiple dimensions: from initial requirement analysis that clearly establishes the dataset's purpose, through implementation considerations such as sample size and data balance, to thorough documentation and annotation. Additionally, a robust data management plan is essential to ensure data integrity and accessibility throughout the competition (for detailed guidelines on dataset development, see Chapter 4). This structured approach to data preparation helps ensure that the competition's scientific objectives can be effectively addressed while providing participants with reliable resources for developing their solutions. General cost evaluation of data generation is complicated because it is highly variable depending on the scientific discipline involved. Data generation has always a cost, but this cost can bee supported by different players of the competition (sponsors, private companies, organization committee, care providers, health insurance, etc). This costs also depends on the data type, size and accessibility. Good quality data also relies on the willing of organizers to work in synchronisation with the global efforts for technical standardization and ethic responsible data sharing, e.g. Global Alliance for Genomics and Health or FAIR principles for data management and stewardship (Wilkinson et al., 2016; Cabili et al., 2018).

Governance and legal costs

Competition governance strategy should also include legal counseling costs, that will ensure that the data storage and sharing concept complies with national and international legal requirements. In particular, usage of identifiable personal data (such as patient clinical data) is a complex and significant legal and data protection challenge (Nicol et al., 2019). Moreover, rules for awarding prizes and travel grants should be clearly defined, this includes definitions of:

- jury's composition (committee of experts)
- criteria of evaluation (e.g. relevance, usefulness, novelty, etc.)
- challenge submission process
- intellectual properties
- exclusion and appeal procedures
- control of the use of funds and goods, including prices
- privacy policy
- errors, frauds and breaches of rules mitigation plan

Computation and storage

The digital data challenge platforms rely on cloud computing services to run and evaluate models. Access to these services can be externalized (such as Google Could Platform, Openstack, IBM Cloud or Amazon web services) or provided internally using the computing infrastructure of the challenge organizers. Depending on the competitions, the problem to solve and the type of data, the required resources vary a lot. For instance, in the case of code submission, it is important to estimate well the number of participants, and sometimes to limit the entries by setting a hard threshold. Indeed, code submission offers many advantages (controlled environments, confidential data, good sharing of the resources among participants, etc.) but is computationally very demanding. Thus, the organizers must accurately estimate the computation time of the expected methods as well as the type of computing units to use ((Ellrott et al., 2019)), knowing that donation of cloud units from Google, Azure and Amazon are relatively easy to obtain. Some platform such as Codalab can be coupled with such cloud services, via the use of compute workers. Finally, they need to decide accordingly whether they wish to offer computational services (allowing code submission) or ask participants to provide their own computational resources (only allowing the submission of results).

SCIENTIFIC EXPERTISE, CHALLENGE DESIGN AND ENGINEERING

Bringing together an expert steering committee is a key factor to ensure that the issue raised by the competition corresponds to the needs of the community, and that the data will be used correctly to ask the right question. These two points are essential to ensure community engagement and the quality of the competition. Code development is also an important factor to consider. In certain specific situations, building a dedicated application or a realistic environment to simulate the various tasks of a competition can demand significant effort, including extensive research and substantial engineering manpower prior to the competition. For instance, L2RPN competition series required the generation of a dedicated framework and the generation of synthetic data with several people working on the project for over a year (cost of ∼200k€). Once the competition is completed, manpower is also needed to analyse the results, summarize, and disseminate the challenge outcomes.

PRIZES, TRAVEL AND CONFERENCE ORGANIZATION

Reward costs should be included in the challenge budget. Prizes can be an important incentive to recruit participant (see section 1). In case of in-person events, travel and conference organization costs should be considered. This can include speakers invitations, participation to the venue costs and travel grants for students. Competitions can be short (one week) or long (over several months), held remotely or in person, and may or may not be associated with an international conference (see “Part II : The best of challenges and benchmarks" for more examples of academic and industry competitions). All these elements must be taken into account when preparing the budget. For example, the HADACA challenges (Health Data Challenges) take place in the form of a one-week winter school in the French Alps, with around fifty participants. The total cost of organizing the event (including accommodation and meals) was €30,000 for the 2024 edition (HADACA3[3]). Example of costs to organized a one day workshop can be found in Table 2).

	Expense type	Estimated cost (EUR)
1	Invited speakers registration (4x$250)	1,000
2	Organizer travel expenses (3x$2000)	6,000
3	Lunch (catering) for 40*$50	2,000
4	Dinner for invited speakers, winners, organizers (20*$50)	1,000

Table 2: Conference or workshop organization for a total budget of 10,000 euros.

3.2 Person power

Person power is crucial in competition organization and should not be underestimated. While 3 provides an average estimation of person power required to organize a challenge, accurately estimating human resource needs remains one of the most challenging aspects of competition planning. These requirements often evolve throughout the competition lifecycle, with varying demands across different phases - from initial planning to final evaluation. Resource needs can fluctuate based on

3. HADACA3 website : https://hadaca3.sciencesconf.org/

unexpected technical challenges, participant engagement levels, or administrative complexities. A proven strategy to address this uncertainty is to establish a robust technical committee from the outset, comprising members with diverse expertise. This committee should include not only scientific experts but also professionals skilled in administrative tasks, accounting, publicity and communication, software development, data analysis, and reporting. Such diversity in expertise helps ensure that the competition can adapt to evolving demands while maintaining high standards across all aspects of organization. This distributed approach to human resources also provides redundancy and flexibility, allowing the organizing team to better handle peak workloads and unexpected challenges.

3.3 Resources: sponsors and grant agencies

As the global cost of competition organization grows along with the complexity of the data and tasks, proposal and grant writing to find money is essential. By leveraging institutional support and sponsors, organizers will achieve good quality challenges and ensure community participation. More and more universities and national funding agencies[4] or scientific societies[5] support competition organization. Building partnership with private companies[6] and involving collaborators in scientific consortium is also likely to be very helpful to reduce the financial barriers in organizing challenges.

4 Conclusion

Organizing a competition necessitates the dedication of a scientific committee, substantial time, and financial resources. It is imperative not to underestimate the level of commitment required for the successful execution of such events. However, potential organizers should not be discouraged. On the contrary, organizing a competition is a highly rewarding experience, and we encourage any aspiring organizer to undertake it. It's worth noting that competitions represent just one approach to collaborative science. Recent initiatives demonstrate the diversity of possible formats: from large-scale collaborative projects like BLOOM by BigScience [7], which brought together hundreds of researchers to create an open multilingual language model, to the development of innovative evaluation frameworks for language models. Furthermore, while this chapter has primarily focused on traditional competition formats, emerging approaches such as dynamic benchmarking offer promising alternatives to static competitions. These dynamic formats enable iterative data collection and model development, though they require specific design considerations to be implemented effectively."

This chapter is designed as a practical guide, and given the large number of competitions already held, newcomers to the field will find abundant examples to draw inspiration and ideas from. The recommendations and guidelines presented in this chapter are intended to serve as a theoretical framework, not as rigid constraints. The innovative nature of this field extends to the format and design of the competitions themselves, fostering a continuous environment of creativity and development.

---

4. For instance the University College of London, the National Research Agency in France, the ETH in Switzerland, or the EIT Health in Europe
5. National Science Foundation in the United States, the IEEE Computational Intelligenece Society, or the International Neural Network Society
6. Non-exhaustive list of potential sponsors: Google, Microsoft, Orange, Kaggle, Health discovery corporation
7. BLOOM: https://huggingface.co/bigscience/bloom

H]

Task	Description	Hours
1	**Finding/reviewing data.**	50
2	**Formatting data.** Preprocess and format the data to simplify the task of participants, obfuscate the origin, anonymize.	100
3	**Assessment.** Define a task and evaluation metrics. Define and implement methods of scoring the results and comparing them.	50
4	**Baseline software; starting kit.** Implement a simple example performing the tasks of the challenge. Prepare useful software libraries, make examples.	100
5	**Result formats and software interfaces.** Define the formats in which the results should be returned by the systems and how experimentation will be conducted during the challenge.	50
6	**Benchmark protocol.** Define the rules of the competition and determine the sequence of events.	50
7	**Web portal.** Implement on challenge platform the benchmark protocol allowing on-line submissions and displaying results on a leaderboard.	25
8	**Guidelines to participants.** Write the competition rules, document the formats and the scoring methods, write FAQs.	50
9	**Beta testing.** Organize and conduct tests of the challenge.	25
10	**Run the challenge.** Answer participants, attend to the platform (2h/week).	100
11	**Prepare the workshops.** Write proposals. Look for tutorial speakers. Select speakers. Create a schedule. Advertise.	50
12	**Competition result analysis.** Compile the results. Produce graphs. Derive conclusions.	50
13	**Reports.** Write reports on the benchmark design, the datasets, and the results of the competition.	100
14	**On-line result dissemination.** Make available online the competition result analyses, fact sheets of the competitors's methods, and the workshop slides.	50
15	**Prepare workshop proceedings.** Solicit papers, organize the review process, and edit the papers.	100
16	**Distribute prizes and awards.**	10

Table 3: Evaluation of person power to organize a challenge (varies from challenge to challenge, should be estimated by the organizing team)

References

Tammy V. Abernathy and Richard N. Vineyard. Academic Competitions in Science: What Are the Rewards for Students? *The Clearing House*, 74(5):269–276, 2001. ISSN 0009-8655. URL https://www.jstor.org/stable/30189679. Publisher: Taylor & Francis, Ltd.

Eric Bender. Challenges: Crowdsourced solutions. *Nature*, 533(7602):S62–S64, May 2016. ISSN 1476-4687. doi: 10.1038/533S62a. URL https://www.nature.com/articles/533S62a. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7602 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Drug discovery and development Subject_term_id: drug-discovery-and-development.

Adam M. Brandenburger and Barry J. Nalebuff. *Co-Opetition*. Crown, July 2011. ISBN 978-0-307-79054-5.

Moran N. Cabili, Knox Carey, Stephanie O. M. Dyke, Anthony J. Brookes, Marc Fiume, Francis Jeanson, Giselle Kerry, Alex Lash, Heidi Sofia, Dylan Spalding, Anne-Marie Tasse, Susheel Varma, and Ravi Pandya. Simplifying research access to genomics and health data with Library Cards. *Scientific Data*, 5(1):180039, March 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.39. URL https://www.nature.com/articles/sdata201839. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Medical research;Research data Subject_term_id: medical-research;research-data.

Allison Creason, David Haan, Kristen Dang, Kami E. Chiotti, Matthew Inkman, Andrew Lamb, Thomas Yu, Yin Hu, Thea C. Norman, Alex Buchanan, Marijke J. van Baren, Ryan Spangler, M. Rick Rollins, Paul T. Spellman, Dmitri Rozanov, Jin Zhang, Christopher A. Maher, Cristian Caloian, John D. Watson, Sebastian Uhrig, Brian J. Haas, Miten Jain, Mark Akeson, Mehmet Eren Ahsen, Gustavo Stolovitzky, Justin Guinney, Paul C. Boutros, Joshua M. Stuart, Kyle Ellrott, Hongjiu Zhang, Yifan Wang, Yuanfang Guan, Cu Nguyen, Christopher Sugai, Alokkumar Jha, Jing Woei Li, and Alexander Dobin. A community challenge to evaluate RNA-seq, fusion detection, and isoform quantification methods for cancer discovery. *Cell Systems*, page S2405471221002076, June 2021. ISSN 24054712. doi: 10.1016/j.cels.2021.05.021. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471221002076.

Tara Eicher, Andrew Patt, Esko Kautto, Raghu Machiraju, Ewy Mathé, and Yan Zhang. Challenges in proteogenomics: a comparison of analysis methods with the case study of the DREAM proteogenomics sub-challenge. *BMC bioinformatics*, 20(Suppl 24):669, December 2019. ISSN 1471-2105. doi: 10.1186/s12859-019-3253-z.

Kyle Ellrott, Alex Buchanan, Allison Creason, Michael Mason, Thomas Schaffter, Bruce Hoff, James Eddy, John M. Chilton, Thomas Yu, Joshua M. Stuart, Julio Saez-Rodriguez, Gustavo Stolovitzky, Paul C. Boutros, and Justin Guinney. Reproducible biomedical benchmarking in the cloud: lessons from crowd-sourced data challenges. *Genome Biology*, 20(1):195, September 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1794-0. URL https://doi.org/10.1186/s13059-019-1794-0.

Justin Guinney and Julio Saez-Rodriguez. Alternative models for sharing confidential biomedical data. *Nature Biotechnology*, 36(5):391–392, May 2018. ISSN 1546-1696. doi: 10.1038/nbt.4128.

URL http://www.nature.com/articles/nbt.4128. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5 Primary_atype: Correspondence Publisher: Nature Publishing Group Subject_term: Policy;Research data Subject_term_id: policy;research-data.

HADACA consortium, Clémentine Decamps, Florian Privé, Raphael Bacher, Daniel Jost, Arthur Waguet, Eugene Andres Houseman, Eugene Lurie, Pavlo Lutsik, Aleksandar Milosavljevic, Michael Scherer, Michael G. B. Blum, and Magali Richard. Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software. *BMC Bioinformatics*, 21(1):16, December 2020. ISSN 1471-2105. doi: 10.1186/s12859-019-3307-2. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3307-2.

E. P. V. Le, Y. Wang, Y. Huang, S. Hickman, and F. J. Gilbert. Artificial intelligence in breast imaging. *Clinical Radiology*, 74(5):357–366, May 2019. ISSN 0009-9260, 1365-229X. doi: 10.1016/j.crad.2019.02.006. URL https://www.clinicalradiologyonline.net/article/S0009-9260(19)30116-3/abstract. Publisher: Elsevier.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards Deep Learning Models Resistant to Adversarial Attacks. *arXiv:1706.06083 [cs, stat]*, September 2019. URL http://arxiv.org/abs/1706.06083. arXiv: 1706.06083.

Daniel Marbach, James C. Costello, Robert Küffner, Nicole M. Vega, Robert J. Prill, Diogo M. Camacho, Kyle R. Allison, DREAM5 Consortium, Manolis Kellis, James J. Collins, and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, July 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2016.

Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aïdan O'Sullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, and Camilo Romero. Learning to run a Power Network Challenge: a Retrospective Analysis. *arXiv:2103.03104 [cs, eess]*, March 2021. URL http://arxiv.org/abs/2103.03104. arXiv: 2103.03104.

Dianne Nicol, Lisa Eckstein, Heidi Beate Bentzen, Pascal Borry, Mike Burgess, Wylie Burke, Don Chalmers, Mildred Cho, Edward Dove, Stephanie Fullerton, Ryuchi Ida, Kazuto Kato, Jane Kaye, Barbara Koenig, Spero Manson, Kimberlyn McGrail, Eric Meslin, Kieran O'Doherty, Barbara Prainsack, Mahsa Shabani, Holly Tabor, Adrian Thorogood, and Jantina de Vries. Consent insufficient for data release. *Science*, 364(6439):445–446, May 2019. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aax0892. URL https://science-sciencemag-org.insb.bib.cnrs.fr/content/364/6439/445. Publisher: American Association for the Advancement of Science Section: Letters.

Adrien Pavao, Diviyan Kalainathan, Lisheng Sun-Hosoya, Kristen Bennett, and Isabelle Guyon. Design and Analysis of Experiments: A Challenge Approach in Teaching. *CiML Workshop, NeurIPS*, page 3, December 2019. URL http://ciml.chalearn.org/ciml2019/accepted/Pavao.pdf?attredirects=0&d=1.

Predrag Radivojac, Wyatt T. Clark, Tal Ronnen Oron, Alexandra M. Schnoes, Tobias Wittkop, Artem Sokolov, Kiley Graim, Christopher Funk, Karin Verspoor, Asa Ben-Hur, Gaurav Pandey, Jeffrey M. Yunes, Ameet S. Talwalkar, Susanna Repo, Michael L. Souza, Damiano Piovesan, Rita Casadio, Zheng Wang, Jianlin Cheng, Hai Fang, Julian Gough, Patrik Koskinen, Petri Törönen, Jussi Nokso-Koivisto, Liisa Holm, Domenico Cozzetto, Daniel W. A. Buchan, Kevin Bryson, David T. Jones, Bhakti Limaye, Harshal Inamdar, Avik Datta, Sunitha K. Manjari, Rajendra Joshi, Meghana Chitale, Daisuke Kihara, Andreas M. Lisewski, Serkan Erdin, Eric Venner, Olivier Lichtarge, Robert Rentzsch, Haixuan Yang, Alfonso E. Romero, Prajwal Bhat, Alberto Paccanaro, Tobias Hamp, Rebecca Kaßner, Stefan Seemayer, Esmeralda Vicedo, Christian Schaefer, Dominik Achten, Florian Auer, Ariane Boehm, Tatjana Braun, Maximilian Hecht, Mark Heron, Peter Hönigschmid, Thomas A. Hopf, Stefanie Kaufmann, Michael Kiening, Denis Krompass, Cedric Landerer, Yannick Mahlich, Manfred Roos, Jari Björne, Tapio Salakoski, Andrew Wong, Hagit Shatkay, Fanny Gatzmann, Ingolf Sommer, Mark N. Wass, Michael J. E. Sternberg, Nives Škunca, Fran Supek, Matko Bošnjak, Panče Panov, Sašo Džeroski, Tomislav Šmuc, Yiannis A. I. Kourmpetis, Aalt D. J. van Dijk, Cajo J. F. ter Braak, Yuanpeng Zhou, Qingtian Gong, Xinran Dong, Weidong Tian, Marco Falda, Paolo Fontana, Enrico Lavezzo, Barbara Di Camillo, Stefano Toppo, Liang Lan, Nemanja Djuric, Yuhong Guo, Slobodan Vucetic, Amos Bairoch, Michal Linial, Patricia C. Babbitt, Steven E. Brenner, Christine Orengo, Burkhard Rost, Sean D. Mooney, and Iddo Friedberg. A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3):221–227, March 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2340. URL `https://www.nature.com/articles/nmeth.2340`. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 3 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Bioinformatics;Protein function predictions Subject_term_id: bioinformatics;protein-function-predictions.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, December 2015. ISSN 1573-1405. doi: 10.1007/s11263-015-0816-y. URL `https://doi.org/10.1007/s11263-015-0816-y`.

Julio Saez-Rodriguez, James C Costello, Stephen H Friend, Michael R Kellen, Lara Mangravite, Pablo Meyer, Thea Norman, and Gustavo Stolovitzky. Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nature Reviews Genetics*, 17(8):470–486, 2016.

Tamara Steijger, Josep F. Abril, Pär G. Engström, Felix Kokocinski, Tim J. Hubbard, Roderic Guigó, Jennifer Harrow, and Paul Bertone. Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, 10(12):1177–1184, December 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2714. URL `https://www.nature.com/articles/nmeth.2714`. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Genome informatics Subject_term_id: genome-informatics.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George

Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, March 2016. ISSN 2052-4463. doi: 10.1038/sdata.2016.18. URL http://www.nature.com/articles/sdata201618. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Publication characteristics;Research data Subject_term_id: publication-characteristics;research-data.