

# Tailored Science Badges: Enabling New Forms of Research Interaction

Sebastian S. Feger  
sebastian.feger@ifi.lmu.de  
LMU Munich & CERN  
Germany & Switzerland

Jasmin Niess  
niessj@uni-bremen.de  
University of Bremen  
Germany

Paweł W. Woźniak  
p.w.wozniak@uu.nl  
Utrecht University  
the Netherland

Albrecht Schmidt  
albrecht.schmidt@um.ifi.lmu.de  
LMU Munich  
Germany

## ABSTRACT

Science faces a reproducibility crisis. There is a need to establish open science practices within the academic reputation economy. Open Science Badges address this issue by promoting and acknowledging research sharing and documentation. The generic design of currently awarded badges enabled their adoption across the sciences. Yet, their general nature makes it difficult to reflect individual practices and needs of distinct scientific fields. In this paper, we explore uses and effects of highly tailored badges in research data management. We implemented six science badges in a particle physics research preservation service. Our exploration showed that scientists were open to encouraging valuable scientific practices through *tailored science badges*. They described entirely new opportunities for interaction with research repositories. We present design implications for systems that promote reproducibility, related to meaningful criteria, repository navigation, and content discovery. Finally, we discuss the scope and uses of tailored science badges in modern science.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; *Walk-through evaluations*.

## KEYWORDS

Tailored Science Badges; Gamification; Reproducibility; Motivation; Navigation; Discovery; Visibility.

### ACM Reference Format:

Sebastian S. Feger, Paweł W. Woźniak, Jasmin Niess, and Albrecht Schmidt. 2021. Tailored Science Badges: Enabling New Forms of Research Interaction. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28-July 2, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3461778.3462067>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*DIS '21, June 28-July 2, 2021, Virtual Event, USA*

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-8476-6/21/06...\$15.00  
<https://doi.org/10.1145/3461778.3462067>

## 1 INTRODUCTION

Reproducibility is a cornerstone of modern science. It enables research validation and provides a foundation for knowledge transfer and reuse. Still, scientists in a wide variety of research domains report being unable to reproduce published work—including their own [2]. Preservation and sharing are core reproducible practices [3, 22, 44] that require substantial efforts to prepare and document research artifacts [7], including datasets, code scripts, software frameworks, and meta-data. As the academic reputation economy focuses on rewarding novel contributions, scientists often perceive following Research Data Management (RDM) practices personally unrewarding [4, 11, 14]. Consequently, encouraging reproducible contributions remains a strategic challenge. Human-Computer Interaction (HCI) can contribute to solving this issue by informing the design of interactive systems that foster reproducibility and motivate researchers to participate in RDM [16, 19].

Recent work has shown that gamification, the “use of game design elements in non-game contexts” [12], provides opportunities to address the motivation and reward challenge. In their requirements study on gamification in the context of reproducible science, Feger et al. [18] found that a wide variety of game design elements appear suitable to motivate comprehensive RDM. Based on their findings, they placed particular emphasis on the success of Open Science Badges (OSB). OSB promote and acknowledge three core open science practices: data sharing, material sharing (e.g. software, protocols, questionnaires), and analysis preregistration. Kidwell et al. [27] showed that adoption of the badges in the *Psychological Science* journal substantially increased data sharing. The general nature of OSB led to their adoption in 75 journals across a variety of scientific domains [21].

The Association for Computing Machinery (ACM) introduced a set of even more fine-grained badges that promote sharing and reproducibility in experimental computer science [6]. Their design strikes a balance between more directed support for research conducted within the ACM’s scientific scope, and the desire to remain applicable to all areas of research within ACM’s diverse scientific landscape. In this paper, we explore designs and uses of science badges that target specific practices and needs of individual scientific fields and organizations. We introduce the notion of *tailored science badges*. Those are badges *closely tailored to a target community, scientific domain, and infrastructure*. Tailored science badges are designed to acknowledge and foster practices that contribute

to establishing an open science culture. Their design and award criteria reflect the specific needs and practices of their target environment. This represents a strong conceptual difference to generic science badges. We expect tailored science badges to enable a more targeted promotion and acknowledgement of reproducible research practices.

Understanding and designing for closely directed promotion of individual research fields becomes important as scientific stakeholders increasingly develop RDM tools tailored to specific scientific domains and research organizations [10, 45]. In this paper, we report on the implementation and exploration of tailored science badges in a particle physics research preservation service. Based on related requirements research on gamification in reproducible science [18], we designed six science badges. They represent three distinct mechanisms: community votes, clear goals, and community usage. These mechanisms are designed to support our study on researchers' perceptions of diverse badge criteria.

Our paper makes three contributions: First, we report on the design and implementation of *tailored science badges* in a physics research preservation service. Second, we study researchers' assessment of the badges and key award criteria, focusing on their suitability and goal commitment. We provide extensive qualitative insights showing how tailored science badges motivate contributions and reshape service perceptions. Third, we present design implications for the implementation of tailored science badges and discuss how our findings help to frame the scope and uses of tailored science badges in modern science.

This paper is organized as follows: We first reflect on gamification design in science and position tailored science badges in the context of related research. Next, we detail the design and implementation of our tailored science badges in a particle physics preservation service. We then describe the study design and present results and findings. Finally, we discuss design implications with particular emphasis on the adoption of tailored science badges.

## 2 RELATED WORK

In this section, we first reflect on basic concepts and requirements of gamification design. Here, it should be noted that our work contributes findings from an *applied* gamification study that explores how tailored science badges can contribute to establishing an open science culture. Following Tyler and Meckler [43], we want to stress that it is not of primary relevance for our study to inform the theoretical frameworks underpinning gamification research. Accordingly, the review of underlying theoretical frameworks is brief and serves mainly as a basis for reference in the discussion.

Second, we reflect on the emerging research of gamified interaction in science, with a particular focus on the design and use of science badges. Finally, we discuss how related work informed the concrete design of our tailored science badges.

### 2.1 Gamification: Concepts and Requirements

Gamification, “the use of game design elements in non-game contexts” [12], has proven to affect the motivation of people in a wide range of different applications, from education [25] to sports [29]. One of the leading theories in gamification research [40], and motivation research in general, is self-determination theory

(SDT) [39]. SDT includes three types of motivation (amotivation, extrinsic motivation, intrinsic motivation) and six regulatory styles (non-regulation, external regulation, introjected regulation, identified regulation, integrated regulation, intrinsic regulation). Ryan & Deci [38] describe extrinsic motivation as doing an activity because of an “extrinsic reward”, such as praise or money. Intrinsic motivation is about doing something because it is personally rewarding. They postulate that the fulfilment of the psychological needs autonomy, competence and relatedness foster intrinsic motivation.

These needs can be satisfied through suitable game design elements. Most common elements are leaderboards, badges, and points [24, 40]. To date, most studies explored a combination of the most common gamification elements, which makes it challenging to identify the motivational effects of specific game design elements [24, 40]. In line with that, Mekler et al. [30] called for the exploration of individual gamification elements and their underlying psychological mechanisms. On another note, Nacke et al. [32] emphasised the need to broaden the spectrum of application of gamification elements beyond classic application areas such as education. Our work answers to these calls through the exploration of tailored science badges in a highly-skilled research environment.

Various studies indicated the potential of gamification to foster intrinsic motivation in different contexts [24, 40]. However, Nicholson [33] showed that the implementation of game design elements such as points, leaderboards, and badges solely to increase performance can have detrimental effects on motivation and potentially alienates users. Consequently, we base the selection and the design of the badges used in our study on a recent requirements study by Feger et al. [18] which offers an overview of the relevant constraints.

### 2.2 Gamification in Science

Gamification has long been discussed as a design tool for professional applications [36]. While collaborative and repetitive tasks have been in the focus, highly skilled and complex environments received little attention. In the scientific context, gamification focused mainly on citizen science, i.e. trying to motivate the general public to participate in scientific processes [13]. Yet, core scientific practices also face underlying motivational issues. And while we can learn from experiences with implementations in business environments, sociotechnical frameworks of researchers are likely to differ from industry employees, necessitating dedicated research efforts [15].

In their recent work, Feger et al. [18] reported on a study of requirements for gamification design in science. In order to assess needs and constraints, they created two gamified mockups of a research preservation service that made use of a wide spectrum of game design elements. One prototype version made use of most common game elements, including points, leaderboards, and badges. The other prototype used a more informative and rational design language, providing statistics, goals, and activity overviews. They found that researchers were excited about the game elements as they can increase their visibility in the community by reflecting best practice efforts. However, the authors stressed that gamification design needs to particularly consider established scientific practice and the fair reflection of contributions. Overall, researchers found

both prototypes valuable, enjoyable, suitable, and persuasive, although individual game design elements like leaderboards were discussed more controversially. We decided to reuse those criteria in order to put the findings from our exploration of tailored science badges into perspective of these wider findings.

## 2.3 Badges for Reproducible Research

Sharing usable research artefacts is a key open and reproducible science practice [3, 22, 44]. However, effective sharing requires substantial efforts in cleaning, documenting, and preserving resources [7]. Efforts which are often not rewarded within the traditional academic reputation economy [14, 20]. Monya Baker mapped the current state of reproducibility in a large-scale survey involving 1,500 scientists working across five different domains, including physics. Her results showed that 52% of the participating scientists perceived a "significant" reproducibility crisis, another 38% perceived a "slight" crisis. More than 60% of the participants reported that they previously failed to reproduce an experiment. In Physics and Engineering, the domain in which this research is anchored, around 50% of the respondents even indicated failing to reproduce an experiment of their own at some point in time. Monya Baker identified several factors that contribute to irreproducible research. Unavailability of "method, code" and "raw data" were among the top factors, with around 80% of respondents indicating that those factors always, often, or sometimes contribute to irreproducible research. Finally, survey participants highlighted that "incentives for better practice" were among the top factors that could boost reproducibility, with more than 80% indicating that it would likely or very likely benefit reproducible research.

Badges, one of the most common elements in gamification design, relate to these findings. They are already used to promote and motivate documentation and sharing of research artifacts. Open Science Badges (OSB) have been shown to significantly increase sharing of data and materials in a psychological science journal [27]. Rowhani-Farid et al. [37] conducted a systematic review of incentives for data sharing in medical research and concluded that OSB "is the only tested incentive that motivated researchers to share data." There are three OSB, acknowledging *open data*, *open materials* and *preregistration*. The ACM introduced more fine-grained badges in their digital library, including *artifacts reusable* and *results reproduced* [6]. The general nature of those science badges allows conferences and journals across a diverse scientific landscape to adopt and award them. OSB are already issued by 67 journals in various scientific domains, from geoscience to neurochemistry [21]. While the design and implementation of tailored science badges requires significantly more effort, we expect that they will enable a more focused promotion of research practices, and reflect specific needs of individual scientific fields.

Nüst et al. [34] expected that badges can play a key role in exposing "building blocks of research." They argue that, in today's computational and data-driven science, the links between publications and underlying digital material are often not sufficiently transparent. Thus, they investigate the "concept of badges to expose, not only advertise, the building blocks of scholarship." The authors describe the implementation of a badge server and stress that further research is needed to "investigate potential effects on

willingness to publish research compendia and elaborate on trust." Our work connects to their research, as it represents, to our knowledge, the first implementation and exploration of tailored science badges in a real research preservation service.

In the study on gamified preservation service prototypes in particle physics, Feger et al. [18] discussed how their findings relate to the impact of science badges. They reasoned about the underlying factors that contribute to the success and acceptance of the badges: 1) They allow promoting valuable and accepted best practices; 2) Badges create incentives, but do not punish; 3) Badges increase visibility; 4) They acknowledge papers, not individuals; and 5) They provide accessible goals. The authors argued that the overall acceptance of badges makes them particularly suitable for scientific environments. The authors further stressed that particle physics researchers wanted to explicitly find analyses on the service that are directly *executable (reusable)*, considered *educational*, or *innovative*. Researchers further desired to navigate analyses based on *popularity*, *completeness*, and number of forks (*fundamental*). Based on their findings, we chose to implement six badges in the open source CERN Analysis Preservation (CAP) service. We considered that this approach allowed for an effective exploration of the impact of *tailored science badges*, as their design is based on extensive research in the context of particle physics and CAP [18].

## 3 TAILORED SCIENCE BADGES IMPLEMENTATION







Based on related requirements research, we designed and implemented six science badges for a particle physics research preservation service, namely CERN Analysis Preservation (CAP) [9]. CAP is an open source<sup>1</sup> service designed to support documentation, preservation, sharing and reuse in particle physics. The service maps research workflows from four large physics collaborations: ALICE, ATLAS, CMS and LHCb. Those collaborations represent the four major particle detectors installed at the Large Hadron Collider (LHC) at CERN, a key laboratory in particle physics. The service lowers the effort required to document and share research through collaboration-tailored templates and supportive mechanisms like auto-suggestion and auto-completion. But, despite the lowered effort, motivating researchers to contribute to the system remains a major concern [17]. In this section, we detail the design and implementation of the tailored science badges in CAP.

### 3.1 Design of the Badges

As discussed in the Related Work section, we decided to base the design of our tailored science badges on the findings by Feger et al. [18]. Their evaluation of a gamified physics preservation service points to six applications, uses and characteristics of preserved research that can be exposed through game design elements. We list the six badges and their descriptions in Table 1.

As part of the design process, we related the science badges to established game mechanisms. Figure 1 shows their connection to applicable gameful design elements, as listed by Tondello et al. [42]. For example, most of the badges provide the means to apply an ordinal measurement scale, which enables a weighted representation in *leaderboards*. Figure 2 (left) depicts an overview of

<sup>1</sup><https://github.com/cernanalysispreservation>

	<b>Educational</b>
Work that is particularly educational. The award is directly based on the feedback of members of your collaboration.	
	<b>Innovative</b>
Rewards work that is innovative. The award is directly based on the feedback of members of your collaboration.	
	<b>Popular</b>
Popular analyses in your collaboration. Popularity is based on the number of researchers viewing an analysis.	
	<b>Fundamental</b>
Refers to work that is fundamental: Analyses published on CAP can be cloned. Cloned research provides a foundation for future research. Frequently cloned work receives this award.	
	<b>Reusable</b>
Award goes to work that is reusable: Analyses which can be re-executed on ReAna receive this award.	
	<b>Thorough</b>
Awarded to analyses which have more than 90% of the fields documented.	

**Table 1: The six implemented science badges and their corresponding descriptions.**

popular work on the CAP service dashboard. Based on the number of analysis views, *popular* analyses can receive one, two or three *popular* badges. In contrast, the *reusable* badge does not support comparisons, as work is either executable on ReAna<sup>2</sup> or not. These considerations are reflected in the table shown in Figure 1. *Progress feedback* is another example of a gameful design element, applicable mostly to the *reusable* and *thorough* badges. These badges are based on clear goals and the system can easily measure the progress towards those goals. Further, we designed *voting* mechanisms to promote *educational* and *innovative* work (see Figure 2, right).

We connected the science badges to gameful design elements in Figure 1 to identify and communicate unique characteristics and mechanisms of the various badges. The process supported identifying and describing three key mechanisms and criteria of the badges:

- **Community votes:** The users provide feedback and rate individual analyses. Corresponding badges are assigned based

on those votes. Every member of the research collaboration can rate any of their colleagues' analyses.

- **Community interaction:** Badges are assigned based on how the community interacts with analyses. For example, analyses that receive a large number of views are considered *popular*. Analyses that are often forked / cloned are considered *fundamental*.
- **Clear goals:** The system describes and communicates clear goals that must be met in order to receive a corresponding badge. A defining characteristic of those awards is that users can easily check progress towards those goals and know what steps must be taken to reach the award criteria.

As depicted in Figure 1, the *educational* and *innovative* badges are based on **community votes**. The *popular* and *fundamental* badges are based on **user interactions** (number of views and number of forks / clones). The *reusable* and *thorough* badges are based on **clear goals**. Here, researchers are in full control of reaching the badge criteria on their own. In the cases of *voted* and *interaction-based* badges, analysts have to trust their colleagues and the system to make a fair evaluation of their work. Still, they can expect that

<sup>2</sup>ReAna (www.reanahub.io/) is a "Reproducible research data analysis platform" that is connected to CAP.

a thorough documentation of high-quality research is likely to increase their chances to earn those badges. However, contributors have no direct control over obtaining these badges.

### 3.2 Service Implementation

The implementation of the tailored science badges impacted the design of all pages and views in CAP. A complete set of screenshots is available in the supplementary material. In the following, we describe the implementation of tailored science badges with the goal of illustrating the overall design.

**Dashboard:** We added overviews and leaderboards for each badge. Figure 2 (left) shows a selection of popular badges on the dashboard. Each list contains up to four references to analyses. The bottom element references the search page.

**Search page:** We implemented dedicated achievement facets on the search results page.

**Analysis page:** As shown in Figure 2 (right), we implemented a voting mechanism to promote educational and innovative work. Furthermore, we added a banner to analysis pages, highlighting achievements. Figure 3 shows an analysis with three awarded badges. Finally, we added a printable banner that opens when one of the badges is selected. This banner is designed to export key information about the analysis, including title, authors and abstract. Also, it lists all awarded badges.

We added tooltips to all badges on all pages. The tooltips described the badges, as well as their award criteria.

## 4 METHOD

We invited 11 researchers to explore the system in order to establish an empirical understanding of the impact of tailored science badges on the researchers' motivation and ability to navigate and discover research repositories. Here, we describe the recruitment of participants, the structure of the sessions and the data analysis.

### 4.1 Participants

We recruited 11 research physicists working at CERN. None of the research analysts had ever participated in any previous study related to gamification in the scientific context. The participants' ages ranged from 29 to 48 years old (mean = 35 years, SD = 6.6 years). We assured participants that we would not disclose the age of individual research analysts, in order to protect their anonymity. The 11 interviewees were all male. This partially reflects CERN's employment structure: according to the latest available report, the 2019 personnel statistics, between 80% and 93% (depending on the contract type) of research physicists working at CERN were male [8]. All interviewees were employed by CERN or by an institute that was collaborating with CERN. As all interviews were conducted during regular working hours, they became part of an analyst's regular work day. Thus, participants received no extra remuneration for the study participation.

Table 2 provides an overview of the 11 participants. We recruited physics data analysts with a diverse set of experiences and roles within the LHC collaborations. In order to create a most complete understanding of perceptions, requirements, needs and impact of tailored science badges in particle physics research preservation, we made sure to recruit both early-career and senior researchers.

Reference	Affiliation	Experience
P1	ATLAS	Postdoc
P2	ATLAS	Postdoc
P3	LHCb	Upper Management
P4	ATLAS	Postdoc
P5	FCC	Postdoc
P6	CMS	Convener
P7	CMS	Convener
P8	CMS	Postdoc
P9	CMS	Postdoc
P10	ATLAS	Upper Management
P11	ATLAS	Postdoc

**Table 2: We recruited researchers with a diverse set of experiences and roles. Most participants were affiliated with the two largest collaborations ATLAS and CMS. Two were affiliated with much smaller experiments (LHCb and FCC).**

We recruited two conveners. Although conveners have a project management role within a collaboration, they are often involved in technical analysis work. In addition, we recruited two active or former members of the upper management of two of the collaborations. We asked those two participants to rate only a subset of the questionnaire, as they are unlikely to preserve analyses themselves. However, we consider their participation to be a key data source, as they provide an administrative perspective from which related work has not profited.







None of the interviewees had any hierarchical connection to any of the authors. And none of the participants had previously taken part in research conducted by any author of this paper. The participants reflect the cultural diversity at CERN. We did not list the nationalities of individual participants, as this might allow to identify some of the researchers based on the information already provided in Table 2. However, we can report the nationalities involved in alphabetical order: Austrian, English, German, Japanese, Portuguese, Spanish, and Swiss. We conducted all interviews in English, which all interviewees spoke fluently. English is the predominant language in research at CERN.

### 4.2 Protocol

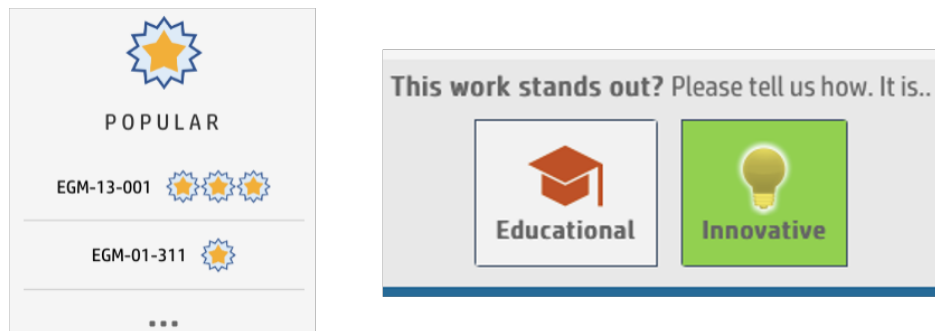
In this section, we describe the structure of the user sessions. The complete material, including the interview protocol and questionnaire, are available as supplementary material.

First, we introduced the participants to CAP. As they had not used CAP before, we asked them to explore the current production version **without badges**. In particular, they reviewed some of the available analyses, the analysis description template and the search page. We then asked them to respond to a questionnaire designed to explore: *value*, *enjoyment*, *suitability*, and *persuasiveness* of this service. For those subscales, we reused the questionnaire items from the study by Feger et al. [18]. The value and enjoyment subscales are based on the Intrinsic Motivation Inventory (IMI).

Next, we switched to our version of CAP **with badges**. The participants were directed to the dashboard. As depicted in Figure 2 (left), they saw an overview of preserved analyses that were

						
	Educational	Innovative	Popular	Fundamental	Reusable	Thorough
Social comparison or pressure					X	X
Leaderboards	X	X	X	X		X
Teams					X	X
Friend invite					X	X
Humanity hero	X			X	X	X
Challenges					X	X
Levels or progression						X
Meaning or purpose	X	X		X	X	X
Progress feedback					X	X
Knowledge sharing	X	X	X	X	X	X
Voting mechanisms	X	X				
Key Mechanism	Voted (Out-of-Control)		Interaction-Based (Out-of-Control)		Clear goals (In-Control)	

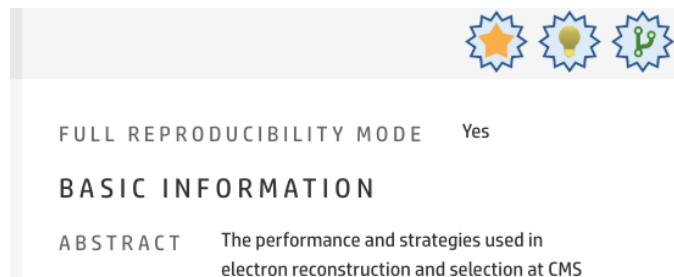
**Figure 1:** We related the six badges to gameful design elements, as listed by Tondello et al. [42]. We further identified three key reward mechanisms: *community voted*, *interaction-based*, and *clear goals*. Our discussions amongst the authors and with scientific colleagues showed that especially the assignments corresponding to the *clear goals* badges (Reusable and Thorough) need further explanations: We related their design to the gameful design elements "Teams", "Social comparison or pressure", and "Friend invite" because we found that fulfilling those clear goals for particle physics analyses can usually only be done through integration of all analysis members. Those badges further relate to "Challenges", "Levels or progression", and "Progress feedback" because of the fact that analyses are assessed according to a set of clearly defined binary rules which are transparently communicated to the analysis authors.



**Figure 2:** Left: An overview of popular analyses on the service dashboard. Analyses can be awarded one to three *popular* badges. Right: The educational and innovative badges are awarded based on community votes.

published by their colleagues and had been awarded a badge. Here, it should be noted that we populated the database with a set of actual physics analyses. Next, the participant received a notification, referring to an analysis of their own that had just received the *popular* badge. As the participants had not used CAP before, we asked them to imagine pre-populated physics analyses as their own. Participants were invited to open the analysis and comment on the different badge-related mechanisms on the analysis page (e.g. the exportable badge banner or the badge preview, as depicted in Figure 3). Here, we asked about the value of badges awarded for their own analyses and on analyses preserved by their colleagues.

Back on the dashboard, another notification appeared. As depicted in Figure 4, the participants were informed about the upcoming introduction of a new badge: the *thorough* badge. Analysts were asked to get more information about this badge by following the link. On the referenced page, the criteria for the thorough badge was described: more than 90% of the analysis fields have to be documented. Two analyses managed by the participant were listed as close to reaching this goal. We then asked about the value of thoroughness in research preservation and the importance of such a badge in navigating the research repository. Finally, we invited the participants to use and review the vote mechanisms (Figure 2,



**Figure 3: The badge banner indicates that the analysis has been awarded popular, innovative and fundamental badges.**

right) and the badge-related search facets. We asked corresponding questions and concluded the practical exercises on this CAP version.

We invited the researchers to answer the same questionnaire as before, assessing the value, enjoyment, suitability, and persuasiveness of this service. Finally, we asked the analysts to rate the suitability, trust, and goal commitment of each of the six badges:

- To assess **suitability**, we reused a slightly adapted statement from Feger et al. [18]: *The [title] badge is NOT suitable for a research preservation service (R)*.
- We used the following statements regarding **trust** in innovative and educational badges: *I trust the RESEARCH COMMUNITY to make a fair assessment of [innovative / educational] work.* Trust statements for the other badges were constructed as follows: *I trust that THE SYSTEM will calculate and award the [title] badge fairly.* The two participants from upper management (see Table 2) completed only those first two scales.
- To assess **goal commitment**, we employed the five-item goal commitment scale by Klein et al. [28].

### 4.3 Qualitative Data Analysis

We recorded a total of 6.2 hours of audio during the sessions. We transcribed the recordings non-verbatim and used Atlas.ti data analysis software to analyze and code the transcriptions. We performed Thematic Analysis [5] to identify themes. Two authors performed open coding of the first two interviews. They discussed and merged their codes and assigned them to code groups. This code tree was used in coding the remaining transcriptions. In total, we created 153 codes. We further discussed the resulting code groups and adapted and merged some of them. Fourteen code groups resulted from this highly iterative and collaborative process. Out of those, we constructed three high-level themes: EFFECTS, CONTENT INTERACTION, and CRITERIA. The theme EFFECTS, for example, is based on the code groups "Visibility", "Career", "Feedback", and "Motivation". Additional information about the structure of the various code groups is available in the Atlas.ti code group report included in the supplementary materials.

## 5 RESULTS

In this section, we first report on the analysis of the questionnaire responses. Next, we report findings from our extensive qualitative

analysis based on three themes: EFFECTS, CONTENT INTERACTION, and CRITERIA.

### 5.1 Questionnaire Responses

Our analysis showed that differences in terms of *suitability* and *commitment* between the badges were most pronounced. Reflecting the exploratory nature of our study, we focus on presenting most pronounced observations in this section. Given our study's qualitative focus and the hereby reflected participant count, we do not report inferential statistics. The complete set of box plots and data is available as supplementary material.

Our analysis of the questionnaire responses showed that participating physicists found both the *reusable* and *thorough* badges more *suitable* than all other badges (Figure 5). This means that the badges in the *clear goals (in-control)* group were considered more suitable than those based on different key mechanisms. The color schemes in Figure 5 and Figure 6 relate to the underlying core mechanisms: community votes (green), community interaction (blue), and clear goals (yellow).

The comparison of rated *goal commitment* showed pronounced differences in participants' commitment towards the *reusable* and *thorough* badges, as compared to all but the *educational* badge (see Figure 6). Noted differences in *goal commitment* and *suitability* between badges in the clear-goals (in-control) group and badges based on different key mechanisms provide valuable insight, especially in light of the extensive qualitative data that we report next.

Our analysis of *trust* towards the badges showed no pronounced effects. Comparison between the two service versions also showed no strong differences. All questionnaire responses are included in the supplementary materials.

### 5.2 Qualitative Findings

Our qualitative data analysis provides further insights into researchers' assessment of the badges regarding the uses and the value of tailored science badges, and the constraints and requirements of their implementation. We present those findings using the three themes: EFFECTS, CONTENT INTERACTION, and CRITERIA.

**5.2.1 Effects.** Researchers perceived the suitability of individual badges differently. Still, they rated the service as suitable, persuasive, and valuable. Based on our qualitative data analysis, this is mainly due to the badges' positive effects. Most participants referred to an increase in *visibility*. Both for research analyses and researchers: *"I mean if it shows up on the main page, people will have to look for it, I guess. Top analysis more people will have to look for it, I guess. Which makes sense, I suppose."* (P9)

*"So, fundamental is I think getting exactly at that. Because then you have some master student who forks it and they do a lot of work on their masters thesis and never publish it in a peer-reviewed journal and never gets cited. But it's still work. It's still interesting science. And that would capture that."* (P4)

In addition, the badges were likely to provide an opportunity for smaller groups or smaller experiments to get visibility: *"I am thinking more to smaller experiments. Because they are completely invisible. So, yes, this could be nice."* (P2)

P7 further discussed multiplication effects enabled by the increased visibility: *"It would give me some insurance that my analysis*



Figure 4: A scripted notification informs the participant about the introduction of a new badge: the *thorough* badge.

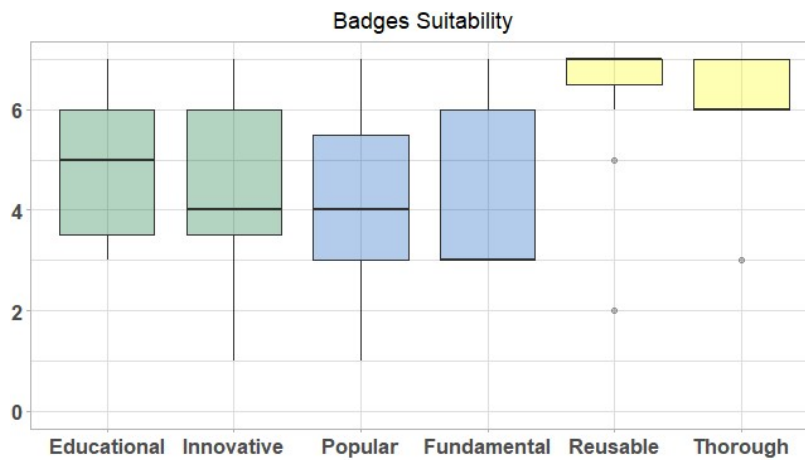


Figure 5: Box plot for badge *suitability*. Based on a 7-point Likert scale. Differences in suitability between the *reusable* and *thorough* badges and all other badges are most pronounced (Mean/SD): Educational (4.8/1.5); Innovative (4.5/2.0); Popular (4.1/2.0); Fundamental (4.5/1.8); Reusable (6.3/1.6); Thorough (6.2/1.2).

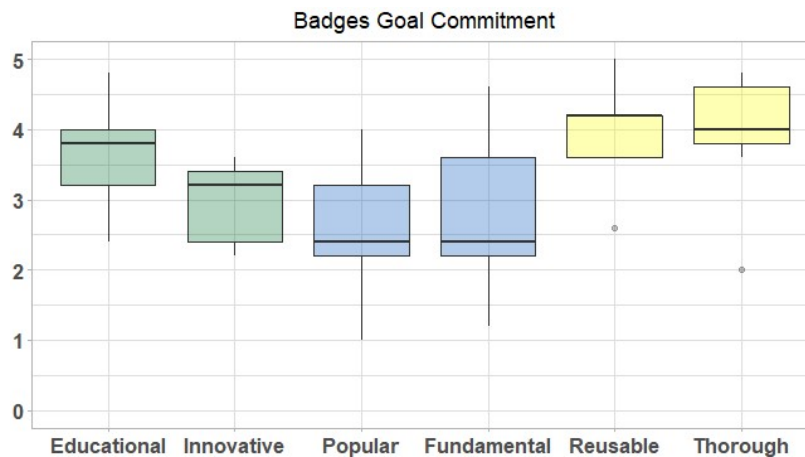


Figure 6: Box plot for goal commitment towards the six badges. Based on a 5-point Likert scale. Differences between the *clear goals (in-control)* group badges (*reusable* and *thorough*) and the *innovative*, *popular*, and *fundamental* badges are most pronounced (Mean/SD): Educational (3.6/0.8); Innovative (3.0/0.5); Popular (2.5/1.1); Fundamental (2.9/1.1); Reusable (4.0/0.7); Thorough (4.0/0.7).



is interesting. Would probably also tell others that this is interesting and it would make it more likely for others to actually look at the analysis. Again, boosting the popularity. Yeah, so I mean it would be nice if you got this if this was available.”

Ultimately, analysts expected that the increased visibility impacts career opportunities. P6 imagined that researchers would add the exportable badges banner to their CVs. P7 thought about an official mechanism where the number of awarded badges are considered as criteria in the promotion of employees. The convener discussed this as an approach to improve the transparency within the organization, as current processes were considered intransparent.

Related to visibility and career opportunities, researchers discussed the role of presentations. They imagined that the exportable badge banner would be a valuable resource in presentations. P4 asked to provide badges tailored to preserved presentations: “When people make talks, the whole point is you are presenting yourself. This is different than a publication which has a thousand authors and isn’t actually attached to you. You know it’s your publication, whereas on a talk it’s a name, maybe on behalf of, but you are giving the talk.”

Finally, most participants discussed feedback as an important driver enabled by the badges: “So, my very first reaction compared to the first version is much more positive. Specifically, the notification I think it’s good. So that you get positive feedback. [...] And there is some abstract later gain, but you often don’t get notified normally. And so if you get this notification I think it’s very useful.” (P1)

P5 asked for the possibility to provide short comments as part of the vote mechanism: “That actually seems interesting. Because I wonder if I can then look at the discussion and can learn a bit more and get more views on more opinions of this analysis. So, this is... If there is actually a discussion there to be viewed, this seems kind of like an interesting thing.”

**5.2.2 Content Interaction.** Most participants described badges as a tool that enabled new forms of interaction with preserved content — and with research work in general. Foremost, they provided a mechanism to navigate large research repositories: “I like this too. (Educational) Exactly the same as the innovative tag. [...] Yeah this one is I think, it’s good to have a few of the analyses of the big pool stand out in certain aspects.” (P8)

“I think the biggest problem at the moment, it’s just that we are beyond 900 papers [...] you basically try to look into the details of the individual analyses, you know the thoroughness badge would probably be very good to have.” (P7)

As the participants noted, the badges were likely to provoke browsing the service and aiding in discovery that would currently rely on unstructured forms of direct communication. In this context, P2 referred to structured and collaboration-wide feedback provided by the vote mechanism: “Now, I put my editorial hat on. I like the concept. [...] Feedback of people that’s the kind of things that you hear around coffee discussions. Oh yeah, go to this analysis. It’s nicely done. It’s nicely documented. You can start from there and learn from it. But, it’s never written anywhere. So, that is useful.”

P1, P4, P5 and P8 referred to mechanisms of serendipitous discovery that were likely to result from researchers browsing the content promoted by the badges: “You don’t really find their work and it’s difficult to discover like this. Unless you work with them and you know where they put their stuff. Can be very nice to kind of like

discover analysis and like that to get an overview of stuff like that.” (P5)

Most researchers discussed re-execution of preserved analyses as a desirable goal. In this context, convener P7 discussed the re-usable badges as a mechanism to filter noise: “Because most of them there is no information that goes beyond the very basics. I could at least filter all the noise. That’s something important.”

P4 expanded upon the notion of improved navigation: “I think the main thing is attaching the badges is important. Because it gives you a different way to query the database. [...] I think the main things is you attached new information that the current way we archive science doesn’t afford.”

**5.2.3 Criteria.** Researchers rated the individual badges differently. They extensively commented on requirements for designing tailored science badges. The initial contact with the gamified service was a critical moment, as the initial reactions of the following two researchers show: “I see the gamification already, there. So, I am not sure about the achievements being used. [...] So, I don’t know what popular is...” (P6)

“So, you basically rate the analysis or somehow like this right... Ok, I mean then the question would probably be how you rate something. Or what is more interesting than others.” (P11)

These quotes refer to two major challenges that surfaced during the exploration of the service by every participant: the need to understand the rules of the badges and inter-badge comparison.

The tooltips placed on the badges proved effective in communicating the individual mechanisms and rules of the badges. Based on those descriptions, the physicists stressed that sophisticated protection mechanisms for most of the badges would need to be implemented. And that this implementation should have been communicated to the users in order to establish trust. The use of those protections is twofold. First, they protect from *unintended* side effects: “If it’s forks, then it could have the nasty effect that— if there is a problem with some particular analysis, people try to fork it several times. [...] Fork again. But still doesn’t work. You see? Oh it’s very reusable! We forked a lot. No, it’s not. But you can get around that. But you would need to put protections at the number of unique forks by unique people.” (P3)

Second, they protect from any attempt to game the system. P3 referred to such concerns related to the thorough badge: “If it is automatically calculated by the computer, it would tend to encourage people to just add some meaningless words everywhere. Or some minimal, just to have something in all fields. While a documented analysis is something different.”

Physicists also referred to adoption within their collaboration as criterion for their use of certain elements. For example, P5 commented on the exportable badge banner: “It depends a lot on how like collaboration or colleagues would use it. I think I wouldn’t go ahead using this kind of thing, because people would kind of wonder, why is this like a popular analysis. And who gives out badges and stuff.”

Participants further discussed the role of the administration in awarding some badges. In particular, regarding the vote mechanisms, the browsing of the service might not be sufficient to provide strong and reliable data for the community feedback: “So, here my

question kind of is: *When does this happen that I am on this page of a different analysis and think 'Ok, I want to vote on this.'*" (P2)

Instead, P2, P3 and P7 imagined that feedback for the innovative and educational badges could be based on the *"decision of some sort of experts."* (P2) Although here, as P2 continued to state, *"the main worry is that these experts then be overloaded and then the quality of their work may not be that high."*

Most participants reflected on the differences in complexity of the individual badges. Foremost, they distinguished between the complexity in terms of awarding them: *"Like the popular. Definitely, it's just counting. This is easy I would say. And educational and innovative. I mean this is how other people see the analysis. Ok, that's also fine. And then fundamental, reusable, yes, there I have a bit more doubts I would say. This is a fair thing and it would work."* (P10)

This reflects a common observation we made during the sessions: Participants tried to imagine examples of analyses that might qualify for individual badges. Reviewing examples proved to be a crucial step in being able to evaluate the usefulness and suitability of a badge. This was especially true for the reusable badge that aimed for a goal for which, to date, only few particle physics analyses qualify. Here, several participants explicitly asked for a finer granularity. P7 provides examples of more accessible steps towards the reusable badge: *"So, I think there could actually be smaller steps towards this, so you know basically your code is available via the portal or something. It's like the first thing. Then it also compiles. [...] There should be more granularity there."*

The discussions regarding complexity also relate to common scientific challenges. P10 had concerns related to the fundamental badge, as *"basically (...) all what we are doing is fundamental."* And P5 wondered about the meaning of the innovative badge, as *"research is supposed to be innovative by definition."*

## 6 DISCUSSION

We discuss findings from our exploration around tailored science badges that we implemented in a particle physics research preservation service. Foremost, we note that researchers and practitioners need to ensure that the design of tailored science badges fit research practices of the target domain, support specific requirements for open science, and integrate into community tools. We expect that adapted versions of the six badges implemented in CAP, or a subset thereof, will prove suitable in other fields. Further, we are confident that a variety of additional tailored science badges will result from field-specific investigations.

In this section, we first discuss how the scope of tailored science badges differs from other generic game design elements in science. Next, we discuss design implications for the implementation and adoption of tailored science badges. Finally, we stress how tailored science badges expand the design goal from *motivating* practices, to *supporting* research practices and enabling new forms of content *interaction*. We expect and wish that our findings and discussions will spark a debate at DIS and within the SIGCHI community on meaningful implementations and adoption of science badges.

### 6.1 Scope of Tailored Science Badges

This study presented empirical findings on the design and exploration of game design elements that are specifically tailored to a

science tool and research community. With this tailored design approach, the badges target a different scope than Open Science Badges (OSB) [21] and ACM badges [1]. While OSB and ACM badges can be easily adopted by a wide variety of journals and conferences, **tailored science badges enable a more focused support of scientific practices. They also differ in terms of underlying mechanisms.** OSB badges are awarded based on the review of committees and experts. The same mechanism applies for most of the ACM badges. However, ACM foresees a form of community interaction related to the *Results Reproduced* badge: The reward can be claimed once other researchers report that they successfully reproduced findings from an ACM publication.

Our findings showed that participants were concerned about overloading committees or experts with tasks of reviewing content and awarding badges. This might impact the quality of the reviews. Here, we particularly profited from the assessments of two members of the upper management of the particle physics collaborations. Notably, researchers recorded no pronounced differences in *trust* towards the six badges. Provided that the badges are based on strong protection mechanisms, **researchers stressed that they trust the system and their research community to make fair assessments — independently of the underlying mechanism.** This finding provides a different perspective on reward mechanisms among the more general science badges like OSB and ACM badges.

### 6.2 Adoption

We observed that **the initial contact with the gamified service is crucial** in the process of assessing the value of science badges. While most physicists directly commented that the CAP Badges version is more attractive and appealing than the current non-gamified version of CAP, most researchers immediately started to compare and reason about the individual badges. They often stopped at the first badge that was not clear to them or one they found troubling. At this point, they showed initial concern for the badge implementation in general. It is reasonable to imagine that many researchers at this stage would lose interest in the badges or even the service if they had no motive to further reason about the badges. In this study, we explicitly asked the participants to further explore the service and to review the mechanisms of the individual badges, at which point the initially concerned researchers stressed that they considered most of the badges useful. In conclusion, **future systems should guide scientists who are experiencing a gamified research service for the first time through the initial exploration process.** This guidance might be provided through notifications that inform about the introduction of a badge in frequently used tools or through tips displayed during the first use.

Participants found the tooltips that appeared once they hovered the mouse over badges useful. While the information was helpful in communicating the basic concept and reward mechanism of a badge, researchers often started thinking about good examples of analyses that would qualify for a particular badge. Occasionally, this proved difficult, as some of the badges promoted mechanisms that were not yet applicable for the majority of the research work conducted within the collaborations. This was particularly true for the reusable badge. Thus, **providing strong examples and justifications in the badge descriptions can foster understanding**

**and assessment of badges.** In addition, researchers repeatedly asked for strong protection mechanisms to prevent deliberate or accidental manipulation. **Service and tool designers do not only need to implement protection mechanisms, but also communicate their implementation to the users.** We argue that communicating badge motivation, strong examples, and protective mechanisms are essential to justify and explain “the provenance of badges (i.e. who awarded it, to what, using which criteria), (which) would be crucial in a scholarly setting to establish trust” [34]. Given those challenges, **the adoption process might benefit from initially introducing badges that are awarded based on clear goals that the users are fully in control of reaching by themselves.** The participants rated commitment towards and suitability of those badges significantly higher. Based on our qualitative findings, we argue that this is primarily explained by the fact that those badges are not prone to many of the challenges described by the CRITERIA theme. In particular, those achievements do not depend on actions of their scientific peers. Rather, they provide a clear sense of progress and communicate required steps needed to fulfil requirements, which in turn favors goal commitment.

Our findings showed that individual badges can be controversial. All researchers mentioned concerns related to the implementation of at least one badge. However, their **perception of the overall service was informed by the most suitable and useful elements.** Still, we need to stress that this might not necessarily be the case if a tool implemented game design elements that provoked most serious concerns. Thus, we want to stress that user-centered research is a necessary requirement for the design of gamified science tools.

Related to adoption, we find that **the design of tailored game elements promoting scientific practices needs to explore mechanisms to reflect achievements outside the original application context.** Research data management tools that are tailored to organizations, institutes, or scientific fields are likely to restrict access to the corresponding research community. While this is not an issue for scientists who stay within the original research area, it becomes challenging for those who change their academic framework or move to industry. Thus, designers should **consider the implementation of exportable formats, as well as forms of communicating achievements that are comprehensible outside the original research context.** The exportable badge banner in our study proved to be a good example and starting point.

### 6.3 Science Interaction

Gamification is commonly used to motivate actions and practices [25, 29]. Leaderboards and badges usually do not hide the intention to push users to perform better, complete certain activities, and compete against each other. Yet, most participants in our study did not explicitly discuss motivation. Instead, they discussed the EFFECTS and uses that the implementation of tailored science badges enabled, focusing on how they can enrich scientific practice on a regular basis. This notion is also reflected in the statement of one of our participants: “[...] *the main thing is you attached new information that the current way we archive science doesn't afford.*” Uses related to the impact on content discovery and repository navigation emerged as part of the CONTENT INTERACTION theme.

Improved content interaction profits those who want to find and use information within the research repository. But, the participants stressed that this also provided a strong incentive to contribute to the preservation service and to follow certain practices which will likely result in more visibility within the research collaboration. Given that participants discussed increased visibility as a driver in the career development, **we argue that the tailored science badges provide an implicit form of motivation that is tied to new forms of interaction with preserved research.** That way, they also differ from the more generic OSB and ACM badges. OSB badges appear on corresponding publications. However, adopting journals are not mandated to implement facets within their digital libraries. To date, ACM only added one badge (Artifact Badge) as search criteria in the *Advanced Search* of their digital library<sup>3</sup>. **We recommend that designers and adopters of science badges – tailored and general – explore means to systematically make the sum of additional meta-data collected on research artefacts accessible to the research community.** Doing so will benefit the scientific community well beyond individual authors who promote their achievements.

While *acknowledging* and *motivating* open science practices remain key design rationales in the implementation and adoption of tailored science badges, we find that researchers’ perceptions of tailored science badges shifted from *motivational drivers* towards *tools that provide new forms of interaction with preserved research.* Further exploration of the relationship between meaningful forms of content interaction and *implicit* motivation might pave new ways for design in gamification, which could be closely connected to the exploration of new application contexts [32].

### 6.4 Limitations and Future Work

We aim to foster the replicability of our work and provide a base for future research in the context of tailored science badges and gamification in science. Thus, we make several of the study resources available as supplementary material. Those include the study protocol, Atlas.ti code group report, the questionnaire, questionnaire responses, plots, and screenshots from the service implementation. Regarding the questionnaire items, we note that some are based on established scales (e.g. the goal commitment scale by Klein et al. [28]), others have been introduced or adapted in related work (e.g. suitability and value statements by Feger et al. [18]), but have not been validated in the science badges context. For systematic future explorations, validation of instruments in this domain and context will be important.

We presented findings from the first implementation and exploration of tailored science badges in a fully functional particle physics research preservation service. Given the novelty of tailored science badges in general, and their first appearance on CAP in the exploration sessions, we want to note that the novelty effect likely impacted researchers’ perceptions. In this context, we want to stress that future large-scale evaluations need to consider and adjust for novelty effects. Further, implementing the badges in this open source preservation service is a limitation of the study, as Feger et al. [18] previously presented findings on gamification design in this environment. However, their findings were limited to

<sup>3</sup><https://dl.acm.org/advsearch.cfm?coll=DL&dl=ACM>

the design and evaluation of two gamified preservation service *mockups*. Our work represents the first, to our knowledge, study of a fully functional implementation of game design elements in a research tool. The study represents a necessary second step (after initial conceptualisation presented in related work) in the systematic development of gamified research data management tools that must precede long-term evaluations in production environments. This is largely due to the fact that such an evaluation would need to involve researchers whom we have to convince about the value of research data management tools in the first place.

The introduction of a strict requirement for data reproducibility and the establishment of universal standards are currently matters of intense political discussion in academia [23, 31]. Consequently, researchers are not acquainted with specific tools for reproducible science. This fact largely affected our study design where participants were first introduced to an RDM system and then presented with a gamified variant. While we recognise this study format could have resulted in order effects, alternative designs were not possible due to ecological constraints.

Our study design was further motivated by the strong impact of even small infrastructure changes specific to physics research in particular, and research culture in general [17, 18, 26, 35, 41]. As service designers, we must not risk deploying gamified services into the scientific cyberinfrastructure without having an empirical understanding of their effects. We cannot risk alienating researchers who commit to open science practices. Based on our findings, we envision opportunities for future work to explore and evaluate tailored science badges in long-term studies across a larger sample. In particle physics and beyond. It would be particularly interesting to map commonalities and differences between requirements for gamification in general, and tailored science badges in particular, between distinct fields of science.

## 7 CONCLUSION

This paper presented a systematic study on the design and exploration of *tailored science badges* in a particle physics research preservation service. We presented findings from our exploration with 11 researchers. The participants were postdocs, group leaders, and members of the upper management of the physics collaborations. Our findings showed that the badges enable new forms of research discovery and navigation within research repositories. We presented researchers' perceptions, as well as the discussed uses, requirements, and needs related to the design of tailored science badges in three themes: EFFECTS, CONTENT INTERACTION, and CRITERIA. Based on our findings, we related the mechanisms and uses of tailored science badges to the wider concept of gamification in science. In particular, we discussed how the design rationales behind tailored science badges differs from generic science badges. Finally, we presented design implications for the implementation and adoption of tailored game design elements.

## REFERENCES

- [1] ACM. 2018. Artifact Review and Badging. Website. (April 2018). <https://www.acm.org/publications/policies/artifact-review-badging> Retrieved September 10, 2018.
- [2] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. *Nature News* 533, 7604 (2016), 452.
- [3] Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danis Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems* 29, 2 (2013), 599–611. <https://doi.org/10.1016/j.future.2011.08.004>
- [4] C. Glenn Begley and Lee M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483, 7391 (2012), 531–3. <https://doi.org/10.1038/483531a> arXiv:9907372v1 [arXiv:cond-mat]
- [5] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool Publishers, 51–60. <https://doi.org/10.2200/S00706ED1V01Y201602HCI034>
- [6] Ronald F Boisvert. 2016. Incentivizing reproducibility. *Commun. ACM* 59, 10 (2016), 5–5.
- [7] Christine L Borgman. 2007. *Scholarship in the digital age: information, infrastructure, and the internet*. MIT Press, Cambridge, MA.
- [8] CERN. 2019. CERN Annual Personnel Statistics 2019. (2019). <https://cds.cern.ch/record/2719035>
- [9] Xiaoli Chen, Sünje Dallmeier-Tiessen, Anxhela Dani, Robin Dasler, Javier Delgado Fernández, Pamfilos Fokianos, Patricia Herterich, and Tibor Simko. 2016. CERN analysis preservation: a novel digital library service to enable reusable and reproducible research. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 347–356.
- [10] Xiaoli Chen, Sünje Dallmeier-Tiessen, Robin Dasler, Sebastian Feger, Pamfilos Fokianos, Jose Benito Gonzalez, Harri Hirvonsalo, Dinos Kousidis, Artemis Lavasa, Salvatore Mele, et al. 2018. Open is not enough. *Nature Physics* (2018), 1.
- [11] Open Science Collaboration. 2012. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science* 7, 6 (2012), 657–660. <https://doi.org/10.1177/1745691612462588>
- [12] Sebastian Deterding, Rilla Khaled, Lennart Nacke, and Dan Dixon. 2011. Gamification: toward a definition. *Chi 2011* (2011), 12–15. <https://doi.org/978-1-4503-0268-5/11/0> arXiv:9781450302685
- [13] Alexandra Eveleigh, Charlene Jennett, Stuart Lynn, and Anna L Cox. 2013. "I want to be a Captain! I want to be a Captain!": Gamification in the Old Weather Citizen Science Project. In *Proceedings of the first international conference on gameful design, research, and applications*. ACM, 79–82.
- [14] Benedikt Fecher, Sascha Friesike, Marcel Hebing, and Stephanie Linek. 2017. A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications* 3, 1 (2017), 1–10.
- [15] Sebastian Feger, Sünje Dallmeier-Tiessen, Paweł Woźniak, and Albrecht Schmidt. 2018. Just Not The Usual Workplace: Meaningful Gamification in Science. *Mensch und Computer 2018-Workshopband* (2018).
- [16] Sebastian Stefan Feger. 2020. Interactive Tools for Reproducible Science—Understanding, Supporting, and Motivating Reproducible Science Practices. *arXiv preprint arXiv:2012.02570* (2020).
- [17] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Albrecht Schmidt, and Paweł W. Woźniak. 2019. Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '19* (2019). <https://doi.org/10.1145/3290605.3300685>
- [18] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. Gamification in Science: A Study of Requirements in the Context of Reproducible Research. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '19* (2019). <https://doi.org/10.1145/3290605.3300690>
- [19] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. The Role of HCI in Reproducible Science: Understanding, Supporting and Motivating Core Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI EA '19). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3290607.3312905>
- [20] Sebastian S Feger, Paweł W Woźniak, Lars Lischke, and Albrecht Schmidt. 2020. 'Yes, I comply!' Motivations and Practices around Research Data Management and Reuse across Scientific Fields. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26.
- [21] Center for Open Science (COS). 2021. Open Science Badges. Website. (2021). <https://cos.io/our-services/open-science-badges> Retrieved February 3, 2021.
- [22] FORCE11. 2014. The FAIR data principles. Website. Retrieved August 8, 2017 from <https://www.force11.org/group/fairgroup/fairprinciples>.
- [23] Sascha Friesike, Bastian Widenmayer, Oliver Gassmann, and Thomas Schildhauer. 2015. Opening science: towards an agenda of open science in academia and industry. *The Journal of Technology Transfer* 40, 4 (2015), 581–601. <https://doi.org/10.1007/s10961-014-9375-6>
- [24] Juho Hamari, Jonna Koivisto, Harri Sarsa, et al. 2014. Does Gamification Work?—A Literature Review of Empirical Studies on Gamification.. In *HICSS*, Vol. 14. 3025–3034.
- [25] María-Blanca Ibáñez, Angela Di-Serio, and Carlos Delgado-Kloos. 2014. Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on learning technologies* 7, 3 (2014), 291–301.

- [26] Radu Jianu and David Laidlaw. 2012. An Evaluation of How Small User Interface Changes Can Improve Scientists' Analytic Strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). ACM, New York, NY, USA, 2953–2962. <https://doi.org/10.1145/2207676.2208704>
- [27] Mallory C Kidwell, Ljiljana B Lazarević, Erica Baranski, Tom E Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, et al. 2016. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology* 14, 5 (2016), e1002456.
- [28] Howard J Klein, Michael J Wesson, John R Hollenbeck, Patrick M Wright, and Richard P DeShon. 2001. The assessment of goal commitment: A measurement model meta-analysis. *Organizational behavior and human decision processes* 85, 1 (2001), 32–55.
- [29] Kristina Knaving, Pawel W Woźniak, Jasmin Niess, Romina Poguntke, Morten Fjeld, and Staffan Björk. 2018. Understanding grassroots sports gamification in the wild. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. ACM, 102–113. <https://doi.org/10.1145/3240167.3240220>
- [30] Elisa D Mekler, Florian Brühlmann, Alexandre N Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71 (2017), 525–534.
- [31] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1, 1 (2017), 0021. <https://doi.org/10.1038/s41562-016-0021>
- [32] Lennart E Nacke and Christoph Sebastian Deterding. 2017. The maturing of gamification research. *Computers in Human Behaviour* (2017), 450–454.
- [33] Scott Nicholson. 2015. A recipe for meaningful gamification. In *Gamification in education and business*. Springer, 1–20. [https://doi.org/10.1007/978-3-319-10208-5\\_1](https://doi.org/10.1007/978-3-319-10208-5_1)
- [34] Daniel Nüst, Lukas Lohoff, Lasse Einfeldt, Nimrod Gavish, Marlena Götza, Shahzeib Tariq Jaswal, Salman Khalid, Laura Meierkort, Matthias Mohr, Clara Rendel, et al. 2019. Guerrilla Badges for Reproducible Geospatial Data Science. *AGILE 2019* (2019). <https://doi.org/10.31223/osf.io/xtsqh>
- [35] Gerard Oleksik, Natasa Milic-Frayling, and Rachel Jones. 2012. Beyond Data Sharing: Artifact Ecology of a Collaborative Nanophotonics Research Centre. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) (CSCW '12). ACM, New York, NY, USA, 1165–1174. <https://doi.org/10.1145/2145204.2145376>
- [36] Florin Opreescu, Christian Jones, and Mary Katsikitis. 2014. I PLAY AT WORK - ten principles for transforming work processes through gamification. *Frontiers in psychology* 5 (2014), 14.
- [37] Anisa Rowhani-Farid, Michelle Allen, and Adrian G Barnett. 2017. What incentives increase data sharing in health and medical research? A systematic review. *Research integrity and peer review* 2, 1 (2017), 4.
- [38] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.
- [39] Richard M Ryan and Edward L Deci. 2019. Brick by brick: The origins, development, and future of self-determination theory. *Advances in motivation science* 6 (2019), 111–156.
- [40] Katie Seaborn and Deborah I Fels. 2015. Gamification in theory and action: A survey. *International Journal of human-computer studies* 74 (2015), 14–31.
- [41] Andrea K. Thomer, Michael B. Twidale, Jinlong Guo, and Matthew J. Yoder. 2016. Co-designing Scientific Software: Hackathons for Participatory Interface Design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) (CHI EA '16). ACM, New York, NY, USA, 3219–3226. <https://doi.org/10.1145/2851581.2892549>
- [42] Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. 2017. Elements of Gameful Design Emerging from User Preferences. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17* (2017), 129–142. <https://doi.org/10.1145/3116595.3116627>
- [43] April Tyack and Elisa D Mekler. 2020. Self-determination theory in HCI games research: Current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [44] 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. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (2016), 160018. <https://doi.org/10.1038/sdata.2016.18>
- [45] Daniel J Worden. 2017. Emerging Technologies for Data Research: Implications for Bias, Curation, and Reproducible Results. In *Human Capital and Assets in the Networked World*. <https://doi.org/doi:10.1108/978-1-78714-827-720171003>