Analysing Trade-offs in Frameworks for the Design of Smart Environments

Carmelo Ardito[a], Giuseppe Desolda[a], Rosa Lanzilotti[a], Alessio Malizia[b], Maristella Matera[c]

[a]*Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro, Via Orabona, 4 – 70125 – Bari, Italy*

{carmelo.ardito, giuseppe.desolda, rosa.lanziolotti}@uniba.it

[b] *School of Creative Arts, University of Hertfordshire, Hertfordshire, United Kingdom*
a.malizia@herts.ac.uk

[c]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 32 - 20134 – Milano*
maristella.matera@polimi.it

Abstract. Smart Interactive Experiences (SIEs) are usage situations enabled by the Internet of Things that empower users to interact with the surrounding environment. The goal of our research is to define methodologies and software environments to support the design of SIEs; more specifically, we focus on design paradigms suitable for experts of given domains, who however might not be experts in technology. In this context, this paper discusses some trade-offs that we identified between six different dimensions that characterize the quality of software environments for SIE design. The trade-offs emerged from the analysis of data collected in an experimental study that compared three different design paradigms to understand in which measure each paradigm supports the creative process for SIE design. After reporting on the study procedure and the data analyses, the paper illustrates how the resulting trade-offs led us to identify alternatives for SIE design paradigms, and to structure on their basis a modular architecture of a software platform where the strengths of the three paradigms can be exploited flexibly, i.e., depending on the constraints and the requirements characterizing specific design situations.

Keywords: Internet of Things, Non-technical Domain Experts, Trigger-Action Programming, Tangible User Interfaces, Trade-offs in Design, User Study.

1 Introduction

Several interactive systems today are based on Internet of Things (IoT) technologies. IoT largely supports the development of smart objects, which are devices equipped with embedded electronics, whose functions and data can be accessed through distributed services [1]. Such devices have a great potential from the interaction point of view, as they enable the creation of tangible interactive objects users can bring with them, touch and manipulate for tackling different tasks in different application domains [2]. Through them, the use of IoT systems can be extended to creating immersive experiences where users are empowered to interact with the surrounding environment, also by means of tangible interactions, and can influence with their actions on the physical environment the state of the overall system. Such possibilities generate Smart Interactive Experiences (SIEs – pronounced *see-ehs*), i.e., usage situations, enabled by IoT systems, where the final users can determine, through their interaction, a “personalized” behaviour of the overall system. SIEs are now adopted in different fields: education [3-5], art exhibitions [6] and museums [7], therapies for intellectual disabilities [8] to name but a few.

The goal of the research discussed in this paper is to define methodologies and tools to support the design of SIEs, an aspect that is only marginally discussed in the literature [9]. The emphasis is on the synchronization among multiple objects and user actions to provide narrative threads conveying some content [10]. To define these complex usage situations, designers not only need to program the behaviour of single smart objects; rather systematic design environments and methodologies are required to guide the identification of strategies for object synchronization so that the goal of the SIE can be reached.

The methodologies and the software environments that we have developed so far relate to End-User Development (EUD) [11; 12], as they address the needs of domain experts who might not be experts in technology. In particular, our work has largely focused on supporting Cultural Heritage (CH) operators to define SIEs for museums [10]. Several studies demonstrate the importance of introducing SIEs in CH sites [7; 13-17]. However, very often the professional operators who organize visits to CH sites are not technology savvy. Therefore, it becomes important to provide them with tools that can facilitate the identification of strategies to convey the content of a CH site through the visitors' interaction with synchronized smart objects.

To support the EUD of SIEs, we initially defined visual metaphors for the definition of Event-Condition-Action (ECA) rules [18] and through user studies we assessed the effectiveness of such metaphors and identified some design implications to improve both the simplicity of rule definition and the expressiveness of visual languages for ECA rule specification [18]. During this phase of the research we also identified the necessity of introducing conceptual tools, able to support a creative process where high-level ideas are translated into rules that synchronize the behaviour of smart objects with user's actions [10]. To respond to this need, we have started investigating methodologies and enabling tools for enriching the SIE resources with semantic properties that can facilitate the programming of the resources themselves. Thus, we defined three different design paradigms, two based on Tangible User Interfaces and one on Augmented Reality, that enable domain experts to enrich SIE resources (smart objects and software services) with properties, which we call *custom attributes*, that derive from their domain knowledge. Custom attributes are meant to facilitate the definition of ECA rules, a phase that follows the initial conception of the SIE idea and serves the purpose of defining the SIE dynamics. Examples of custom attributes are the historical features of artworks exhibited in a museum (such as historical age, related events, historical movements) that can be associated to smart objects so that visitors manipulating the objects can be guided along paths that are coherent with a given historical theme. Such properties assign to the objects operational meanings that allow experts to define ECA rules by focusing on the features of the domain the SIE refers to, rather than on the technicalities of low-level events and actions exposed by the smart objects.

Contributions

In our previous work, we already proved the effectiveness of introducing *domain-oriented semantics* as a means to simplify SIE design [10]. Recently, we also conducted an experimental study to compare the three design paradigms for the definition of custom attributes, to better understand the contribution of each of them to the creative process underlying the SIE design [19]. In this article, we present the results of further analyses on the data collected during the last study, which we did not discuss previously and that focus on correlations between some dimensions selected for the evaluation. The contribution of this paper is therefore articulated along the following points:

- *EUD paradigms for SIE design*. We present three paradigms for SIE design by domain experts. Through their comparison, we try to shed light on the elements that can support non-technical designers to conceive SIEs and systematically translate high-level ideas into specifications that can facilitate and enhance the following definition of ECA rules.
- *Trade-offs for frameworks for SIE design*. Based on the results of the experimental study, we discuss some trade-offs that we identified between *UX, Workload, Engagement, Creativity, Satisfaction* and *Ecology*. These dimensions were considered for the evaluation of the three paradigms. The analysis of the collected data highlighted that the considered paradigms, even the one that was ranked as the best, are not able to maximize all such dimensions, while a combination of elements taken from each paradigm can be fruitful [19]. Considering as a basis the specific techniques devised for the three design paradigms, this article outlines some implications, related to how to combine their strengths, which can be exploited to enhance creative processes for SIE design.
- *Reference architecture*. In order to show how the presented design implications can be made concrete, and to foster the replicability of our approach, we present the architecture of a platform that is adaptable to different design situations. It allows designers to flexibly combine those elements of the three design paradigms that best suit the design situations they have to cope with. In fact, given the intrinsic flexibility of the platform, which is favoured by the decoupling of the User Interface (UI) from the other layers, different design skills, needs and goals can be easily accommodated by "plugging-in" different UIs in a cross-modality and cross-device fashion. The presented architecture proposes, therefore, a solution that can be adopted by software developers to provide flexible software frameworks.
- *Lightweight integration of resources*. The main contribution of the proposed paradigms and the enabling architecture is not to be considered at the technical level; rather, based on well established technologies for service synchronization, our approach especially aims to promote abstractions that: (1) capture and simplify the most salient technology aspects of smart objects, making them approachable by non-technical designers and also enabling the introduction into the software system of annotations deriving from the domain in which the designers operate; and (2) can be handled by lightweight architectures, making the software framework supporting SIE design easily accessible and installable in the different environments where the design has to be carried out. This lightweight paradigm could have a limited coverage with respect to the immense capability offered by IoT technologies. We, however, deem this is not a weakness with respect to the goals of our research, as we purposely tried to filter out, through a user-centred design, those aspects that can really help designers to make sense of this technology.

All these contributions are in line with the notion of *Quality of Life frameworks* that promote tools to explore innovative solutions to improve creativity, also taking into account aspects such as gaming and enjoyment [20]. Such frameworks aim to widen design spaces by fostering the exploration of flexible solutions, able to combine the strengths and reduce the weaknesses of the different involved aspects. This also corresponds to the goal of our research: the trade-off analysis and the software architecture presented in this article are original contributions that go in this direction.

Article organization

The article is organized as follows. Section 2 discusses the rationale and background of our research. It describes our previous work and situates it with respect to some notable approaches from the literature. Section 3 then illustrates the three design paradigms compared in the experimental study; through an example scenario, it also clarifies how the domain-oriented semantics at the basis of the proposed methodology supports SIE design. Section 4 describes the study procedure and reports on the results related to the six different analysis dimensions. Section 5 discusses, for each paradigm, some trade-offs emerged from correlations between the considered evaluation dimensions. Section 6 presents some resulting design alternatives that led us to define a modular architecture of a system for SIE design so that the potentialities identified for the three paradigms can be exploited flexibly, i.e., depending on the specific design situations, the constraints to be fulfilled and the requirements that the designers aim to maximize. Finally, Section 7 draws the conclusions and outlines our future work.

2 Rationale and Background

About 30 years ago, Mark Weiser envisioned that pervasive devices and services would have become parts of our daily life [21]. IoT technologies have largely accelerated this trend till the point that many environments we live in are today augmented through an interleaving between the cyber and the physical worlds [9]. Information collected from the physical worlds by means of sensors is easily processed by the cyber world to guide the functioning of applications and services that in turn modifies the physical world itself through actuators.

This new technological landscape opens up the space for the creation of disruptive SIEs [7; 8; 10; 14; 15; 22]. This is having an unprecedented impact on the activities that can take place in environments provided with capabilities of "augmented sensing and interaction" [23]. However, there are still important issues to be solved to increase the adoption of such technologies by a larger audience of developers, including professionals who are not very much acquainted with technology.

The IoT phenomenon has been largely investigated on the technical side [24; 25]. Some approaches already try to facilitate the configuration of smart objects [15]. However, it is still hard for non-technical users to synchronize the behaviour of multiple physical and virtual (i.e., software) resources, installed in the environment or embedded in tangible objects they have to manipulate. This is required when SIEs must be defined; our research tries to fill this gap by investigating methodologies and tools supporting the SIE design by non-technical users.

Semantic enrichment of environments for SIE design

Our recent work has focused on an End-User Development (EUD) approach to facilitate SIE design to stakeholders who are domain experts in charge of designing SIEs, but do not necessarily have the required technical background. EUD is a research field that focuses on enabling people who are not professional developers to design or tailor their interactive applications [11; 12].

We identified visual metaphors [18] able to increase the simplicity and the expressiveness of languages for specifying ECA rules [26]. We also proposed a visual framework to empower non-technical SIE designers to build a semantic layer that can facilitate the definition of rules for the synchronization of multiple resources [10]. This is in line with the contribution of some other works that propose the use of ontologies to specify high-level concepts able to provide an abstract and technology-independent representation of the smart object behaviour [27-29]. The advantage of adding a semantic layer is that, by exploiting semantic terms, designers can define ECA rules by referring to the ontology concepts without worrying about technical details [27]. The semantic enrichment implies creating ontologies and associating them to smart objects; these activities require technical skills and a significant effort, still exposing the system to the risk of not covering the actual needs of the SIE designers.

Our paradigm to build a semantic layer tries to alleviate the problems observed in the other approaches. As better explained in the following section, we propose the definition, by the designers themselves, of *custom attributes*. Similar to ontology concepts (e.g., see [29]), custom attributes are meant to represent knowledge that can simplify the definition of ECA rules. However, they have a different flavour as they enable SIE designers to express the operational semantics they want to assign to the SIE resources depending on the specific usage situation they want to address. In other words, the peculiarity of custom attributes is that they are *usage-driven terms*, specific to the application domain, that help SIE designers make sense of digital resources, putting them in context with respect to the actual usage situations to be addressed.

Creativity-Support Environments

The results of a previous study demonstrated that the notion of custom attributes, as conceptual tools used in the initial phase of SIE conception, effectively aids domain experts to reason on and ideate the SIE [10]. However, that study also highlighted that the visual paradigm adopted for the definition of custom attributes was not adequate to the target users, i.e., domain experts with no technical background. Despite the satisfaction in the initial conceptual identification of custom attributes, the domain experts did not feel comfortable with the visual specification needed to store the attributes into the system. This activity revealed quite more cumbersome than the successive phase of ECA rule definition. This could be ascribed to the lack of usability of the proposed visual interface. However, the study participants also highlighted the need for additional paradigms that could stimulate creativity. They suggested the use of more natural, not necessarily visual, interaction paradigms, especially for custom attribute definition, able to take advantage of the physical nature of the objects to be synchronized and to favour the exploration of the physical environment where the SIE will be rendered. The surrounding environment, indeed, can be a source of relevant custom attributes. These results motivated us to design and evaluate the new interaction paradigms (described in the next section) that could support more effectively the *creative process* underlying SIE design [19].

Creativity is not easy to define and can be a difficult-to-measure aspect [30; 31]. The literature recognizes the impact that Creativity-Support Tools (CSTs) have in helping people with their creative processes [32]. According to [33], a CST for the computing domain is any software design environment that is used to create software artefacts. CSTs can also be considered in the larger spectrum of Creativity-Support Environments (CSEs), i.e., design environments that use different CSTs in different phases, and focus on setting enabling work environments including specialized hardware, e.g., tangible devices, and instrumented spaces fostering collaboration.

According to Shneiderman [32], CSTs should stimulate creativity based on previous knowledge, should link to associated ideas, and should also provide structured tools for exhaustive exploration. They can also support strategies for collaboration.

Shneiderman also introduces four phases a CST should be founded on:

- *Collect*: learn from previous works, available for example in libraries or online repositories.

- *Relate*: consult with peers and mentors at early, middle, and late stages.
- *Create*: explore, compose, and evaluate possible solutions.
- *Donate*: disseminate the results and contribute to enlarge repositories of knowledge based on previous experiences.

These phases and the related principles guided the design of the paradigms for custom attribute definition described in the next section.

Analysing trade-offs

The study that we conducted to compare the three design paradigms highlighted that each of them has peculiar features that amplify some qualities of the software environments for SIE design. As described in the next sections, the analysis of the experimental data led us to identify some trade-offs between the considered evaluation dimensions and put the emphasis on possible design situations where one needs to renounce to some features in order to fulfil with some constraints or also gain on other, more relevant aspects. This in the end helped us recognize that frameworks for SIE design would benefit from having flexible architectures, able to offer alternative paradigms depending on the actual situations in which the design is conducted.

This assumption is in line with some works that discuss the need for trade-offs in design. Choosing appropriate practices for a project can be hard, given the various quality dimensions that generally have to be optimized [34]. In [20], the author says that design is a process with no optimal solutions, and therefore trade-offs are essential because in many situations it is difficult, or even impossible, to identify fitting solutions without considering specific goals and constraints. These works focus especially on the quality of the final product (the SIE in our case). However, their assumptions can be easily translated at a meta-level, where quality dimensions have to refer to the design process itself, rather than to the final product for the final users. The focus of our research is indeed on meta-design (i.e., "design for designers") and on software environments through which domain and/or technology experts can customise or even create the final applications to be exploited by end users [35].

In [20], the author also introduces the notion of Quality of Life (QoL) frameworks as tools to explore innovative sociotechnical environments contributing to creativity, gaming and enjoyment. For such frameworks, it is important to identify and understand design trade-offs: in contrast to design guidelines, such frameworks indeed are supposed to widen design spaces by fostering the exploration of new approaches able to combine the strengths and reduce the weaknesses of the different involved quality dimensions [36]. Thus, the frameworks should be *permissive*: different of the majority of current design environments, they must not be pre-packaged systems, conceived for a specific context and rigidly fulfilling with pre-defined rules, checklists, and workflows. Rather, they must give to designers the autonomy to work with the solution they deem more adequate according to the design situation they have to cope with.

The frameworks for SIE design we focus on are strictly related to the QoL framework, as they are tools that aim to stimulate creativity to guide the design of innovative smart environments. In line with the main assumptions of the works commented above, the studies we conducted also highlighted a difficulty in identifying a design paradigm that best fits the different needs and qualities of a SIE design process. Thus, we exploited the results of a trade-off analysis, reported in this article, to understand how to design a "permissive" framework, where elements of different design paradigms can be mixed to accommodate varying design contexts. Besides the identification of some implications for paradigms for SIE design, the flexible architecture that we defined for the resulting tool is an original contribution towards achieving permissive frameworks supporting SIE design.

3 Design paradigm description

In this section, we describe the prototype systems that implement three new paradigms for SIE design. An example of SIE, i.e., a game played by visitors at a museum, also shows the role of custom attributes as a conceptual tool that aids domain experts to reason on and design the SIE [37]. A video demonstrating the use of the three systems is available at the following link https://goo.gl/K2s3DS

Example scenario

John is a professional guide who wants to offer pupils a game to explore the “smart” rooms of an archaeological museum, which currently hosts an exhibition on the archaeological investigation process and used tools. The displayed artefacts are equipped with smart tags, i.e., QR codes or RFID tags, which visitors can scan to obtain additional information. The game goal is to identify artefacts with a specific characteristic, for example, those related to a specific phase of the archaeological investigation process (e.g., collection of historical sources, excavation, and stratigraphic reconstruction). To play the game, each pupil is provided with a smart magnifying glass that, thanks to an embedded RFID reader, reads by proximity the properties assigned to a tool, for example, its usage phase. Using an app available in a tablet, John sets a quest for the players: *Find tools whose “Usage phase = stratigraphic reconstruction”*. Pupils explore the museum, identify the tools corresponding to John’s request and put the magnifying glass close to it. If the tool is actually used during the stratigraphic reconstruction, a video describing the main characteristics of the tool is shown on the display of the magnifying glass. Points are given as a reward. The game continues with John asking other questions and setting new quests.

John has to manage a number of smart objects, i.e., the archaeologists’ tools and magnifying glasses that are part of the exhibition. For each of them, he must define behaviours by specifying ECA rules.

In order to simplify the process of synchronizing the behaviour of all such smart objects, we propose to empower non-technical SIE designers to build, on top of the SIE smart objects, a semantic layer by defining *custom attributes*. This allows SIE designers to adopt in rule specification a language closer to their own domain-expert language and would introduce abstractions that favour generalization. For each smart object, John defines properties (which, more technically, are attributes of the object) that can express the meaning and the role of the object according to the game dynamics. The variables later used in the rules are exactly the attributes previously defined by the SIE designers.

In the example scenario, each magnifying glass is used to identify the visitor who carries it during the game. Thus, one possible attribute for the magnifying glass is “Owner”: for each magnifying glass it will hold a value indicating the player identifier (e.g., Player_1, Player_2, etc.). Similarly, it is possible to enrich the tools with attributes such as “Usage phase” (with values: *source collection*, *excavation*, and *stratigraphic reconstruction*), “Exposition room” (with values such as *source collection room 1*, *excavation room 1*, *stratigraphic reconstruction room*) to indicate the museum room where the tool is exhibited, “Video” (with values indicating names of video files) to specify the video to be shown on the magnifying glass display when the retrieved tool is the right one. John “freely” defines these attributes and their values depending on the goal he wants to pursue through the game, without any constraint (syntactic or semantics) on the type of properties to be specified. Therefore, we name them *custom attributes*.

After defining custom attributes, John specifies the ECA rules controlling the behaviours of the smart objects. He uses a visual paradigm, similar to the one proposed in the popular IFTTT platform [38], which is more usable for non-technical users. An example of a rule, which for brevity we represent here in a formal syntax, is:

$Rule_i$: “IF a magnifying glass is close to a tool WHERE tool.Usage_Phase = quest.Usage_Phase THEN magnifying glass shows the tool.video_file”.

Without custom attributes, several rules would be defined for each device, such as: “If the *magnifyingGlass_012* is put close to the *tool_032*, and the current quest is *Usage_phase = stratigraphic reconstruction*, then the *magnifyingGlass_012* shows the video <video01.mp4>”. This rule would be replicated for each exhibited tool and for each magnifying glass. Thanks to the custom attributes, the single $Rule_i$ addresses an entire class of smart objects with the same behaviour.

Description of the three design paradigms

By taking into account the results of a design workshop where 28 participants, arranged in groups of 5 or 6, were asked to reason on possible solutions to support the creative process for SIE design [37], we implemented three systems, i.e., *Tangible*, *Explorative* and *Tactile*, which are based on Tangible User Interfaces, Augmented Reality, and a mix of tangible and multitouch interaction paradigms, respectively. The focus on these interaction paradigms is motivated by the advantages they offer in relation to the CST phases outlined by Shneiderman in [39]. As described in the remaining of this section, each of the corresponding systems supports one or more CST phases. All the three systems intrinsically support the *Donate* phase: by their nature, they favour the establishment of domain knowledge repositories, based on the definition of custom attributes, which can be exploited for future design sessions and by other designers.

Tangible system

The *Tangible* system implements Tangible User Interfaces, which couple digital information with everyday physical objects and environments to augment the real physical world. Manipulation of physical artefacts improves tangible thinking, that is the ability to think by means of the manipulation of objects augmented with digital information [40]. It can thus support the *Create* phase, where new knowledge has to be identified and represented in form of custom attributes.

The *Tangible* system revolves around the idea of letting designers manipulate tangible objects, to exploit the capability of tangible interaction to stimulate creative thinking [41; 42]. Two kinds of tangibles are used: i) the *smart objects* to be used during the final SIE; ii) other tangibles that SIE designers manipulate for defining custom attributes; we call them *tangible attributes*.

The participants of the design workshop identified three main types of tangible attributes, *textual*, *numerical*, and *locational*, and three corresponding objects. The most desirable proposals resulted in a *pen* for *textual* attributes, *dice* for *numerical* attributes, and a *compass* for locational attributes. The idea that most largely emerged for custom attribute definition consisted in exploiting the co-proximity of tangible attributes and smart objects, and specifying attribute name and value by using post-it notes attached to the tangible attribute.

Going back to the scenario described above, Figure 1 illustrates how John would define custom attributes for his serious game. He puts on a table some tools players will use during the SIE, i.e., a trowel (on the right) and a pick. Then, he puts pertinent tangible attributes close to the tools. For example, he puts the pen close to the trowel to indicate his intention to define a textual attribute; then he attaches a post-it to the pen to specify the name and value of the attribute, for example, *"Usage_phase = excavation"*. He repeats the same actions for each custom attribute he wants to define. In the end, he uses a mobile app to take a picture of all the elements on the table. The recognized elements are automatically converted into the definition of custom attributes (<*attribute name = value*> pairs). Thus, the trowel in the system is enriched with the custom attributes <*Usage_phase = excavation*> (textual), <*Points = 1*> (numerical), and <*Exposition room = Excavation room 1*> (locational). Similarly, the pick is characterized by the attributes <*Usage_phase = excavation*>, <*Points = 3*> and <*Exposition room = excavation room2*>.

Once the custom attributes are in place, John proceeds with the creation of ECA rules, by using a visual interface such as the one proposed in [18].

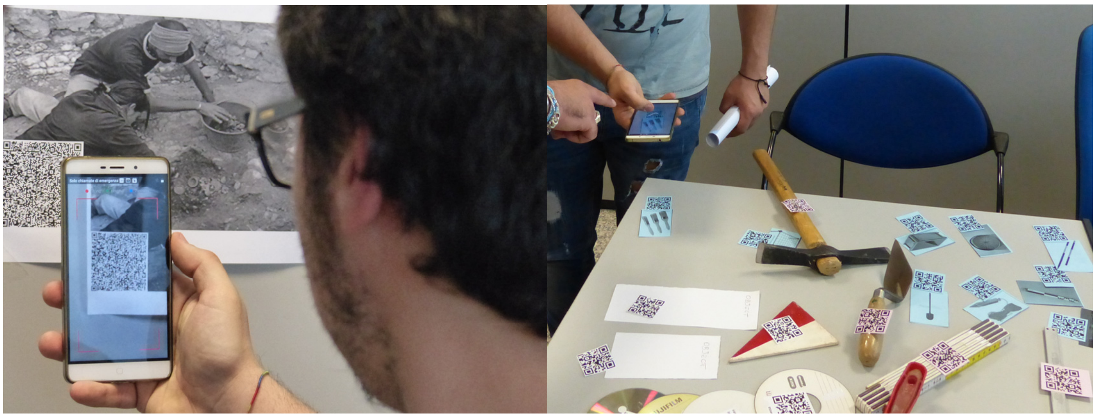
Figure 1. Tangible system: tangible attributes, post-it notes and a mobile phone are used to define custom attributes for a smart trowel and a smart pick.

Explorative system

The *Explorative* system is based on the Augmented Reality (AR) paradigm, which incorporates hand gestures and voice commands instead of traditional user command tools such as a touchscreen, mouse, or keyboard [43]. Thus it is suitable for exploratory search [32] and walking [44], which were shown to be effective to stimulate creativity. For these reasons, it can be appropriate in the Collect and Create phases.

The *Explorative* system promotes the interactive exploration of the real world. The idea is that codified properties of real objects in the surrounding environment can suggest custom attributes for SIE smart objects; such real objects are called *source objects*. For example, referring again to John's scenario, the exhibited documents and tools that are equipped with QR-codes and RFID tags can be source objects. In an AR fashion, a mobile app is used to augment source objects, framed with the camera of the mobile device, with a virtual layering of properties that may suggest custom attributes.

These properties can be "copied and pasted" onto the SIE smart objects, as shown in the example in Figure 2. Specifically, John walks in the museum rooms devoted to the exhibition, looking for interesting source objects. He notices a picture that shows two archaeologists during a digging activity, thus he uses the mobile app to scan its QR-code and visualize its description and properties (Figure 2a). From a pop-up menu listing the picture properties, he selects two of them (the first one indicating a location, <*Exposition room* = *excavation room 2*>, the second one indicating a textual information, <*Usage_phase* = *excavation*>). Once the attributes are selected, the app allows John to edit their name or values, or to remove them. He can also add new attributes from scratch. To associate the selected attributes to the trowel smart object, John scans the brush QR-code (see Figure 2b). In the end, John creates the ECA rules that define the smart object behaviour.



(a) (b)

Figure 2. *Explorative* system: a smartphone is used to (a) explore the environment to capture attributes from source objects; (b) associate them to SIE smart objects.

Tactile system

The *Tactile* system is a hybrid solution that mixes the use of tangible objects and tactile interaction with a horizontal interactive display that acts as a digital workspace enabling the association of custom attributes with smart objects. Multitouch tabletops specifically engage multiple users to interact with physical and virtual objects at the same time, and privilege natural and intuitive social interactions. Most of the computer-mediated approaches for creativity support single-user interactions, thus failing to account for collaboration in group-based, face-to-face scenarios [45]. To overcome this problem, we considered tabletops and tangible objects interaction, which proved suitable for collaborative processes needed in the *Relate* phase [41; 42; 45].

The tangible attributes used in the *Tangible* system, i.e., the pen, the dice and the compass, are used to represent types of custom attributes. The association occurs by putting a smart object close to the tangible attributes on the tabletop. Instead of using posts-it, attribute names and values are specified by means of menus displayed on the digital workspace.

In the example of Figure 3, John acts in front of the tabletop to assign attributes to smart objects. He starts by putting on the surface a smart object (e.g., a DVD on the left of Figure 3). A proximity area (i.e., a rounded halo) appears around the object to indicate that tangible attributes can be placed inside it. John puts a pen inside the halo, thus a pop-up appears on the interactive surface, asking him to define the attribute name and value, i.e., <*Usage_phase* = *excavation*>. John goes on by defining further attributes. Attributes names and values are specified using the virtual keyboard shown on the screen. As with the other systems, in the end, John creates the ECA rules that define the behaviour of the smart object.

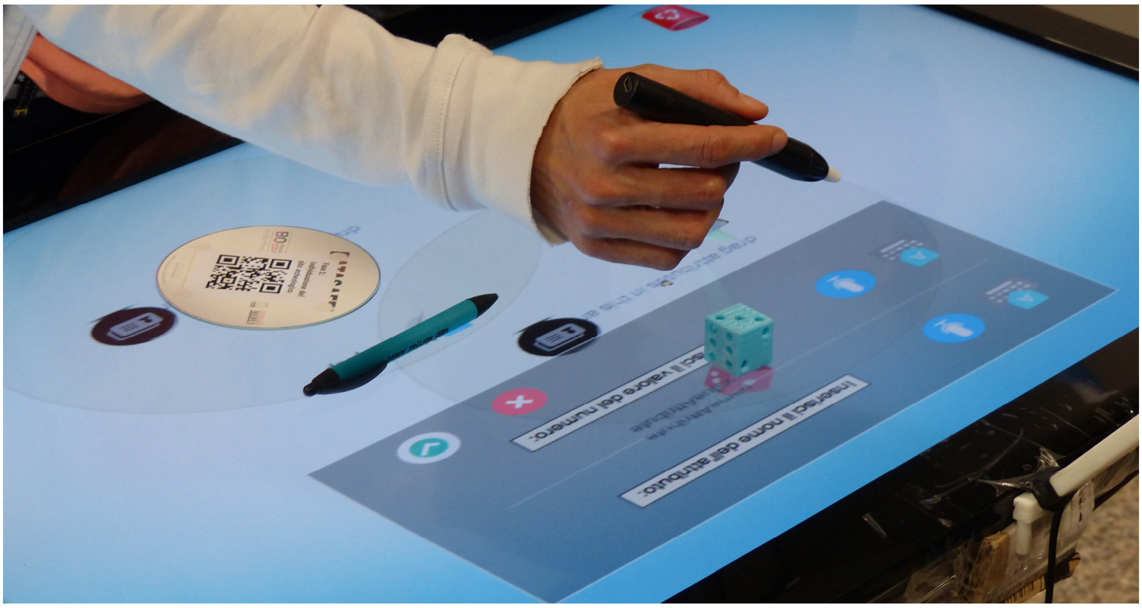


Figure 3. ***Tactile*** **system: user puts smart objects and tangible attributes on the display and specifies attribute names and values by means of a virtual keyboard or speech transcription.**

4 Experimental study

As ground for this experiment, we chose the Cultural Heritage (CH) domain due to the recent and growing interest of guides and curators in the adoption of smart objects as artefacts integrated into museums and CH sites to support the fruition of content [14; 15]. The data presented and analysed in this article were collected during a comparative study, whose main goal was to better understand the contribution of each design paradigm to the creative processes of SIE design. As reported in [19], the analysis of *Creativity* and *Satisfaction* led us to identify interesting implications for the design of creativity-support environments. During the study, we also collected other data that in this article are analysed to investigate how the three different paradigms affect some aspects related to *UX*, *Workload*, *Creativity*, *Engagement*, *Satisfaction* and *Ecology*. More specifically, through these new analyses, we aim to identify trade-offs that could suggest how to calibrate different elements of the design paradigms to maximize the quality of the design environment according to specific design situations.

Participants and Design

A total of 18 students at the last year of the Bachelors' or Master's degree in archaeology (13 females) participated in the study that was recognized as part of their curriculum activities and rewarded with additional credits. Their mean age was 23.9 years (SD = 6.65, min = 20, max = 44). Slightly more than half of the participants (i.e., 10) already attended at least one excavation campaign and spent a good amount of time in professional activities. Only 5 of them organized guided tours in museums or in archaeological parks. Regarding their experience in Information Technology (IT), it emerged that they had a moderate experience in IT and in using mobile devices and a medium familiarity in interacting with smart objects and interactive displays.

The participants knew each other well, because they attended the same university courses, which usually include a few students (about 20 people), or were used to participate in the same professional and social activities.

A within-subject design was performed, with the system as an independent variable and three within-subject factors, i.e., *Tangible*, *Explorative*, *Tactile*. The participants were organized in 9 groups of two. Participants were allowed to express a preference for the partner with whom they would have liked to undertake the experiment, although we tried to include in as many groups as possible a participant with previous experience as a guide.

Procedure

The procedure of the study consisted of 3 main phases. Three HCI experts were involved: one acted as moderator in the first phase and in the third phase; the other two acted in the second phase, one as a facilitator and the other as an observer. The experimental study lasted three days, i.e., 3 groups were observed each day. All 9 groups got the same design brief.

In the first phase, the group conceived the SIE. After the HCI expert introduction about the study purpose, the group signed a consent form and filled in the questionnaire to collect demographic data. Then, the participants were asked to act as curators of a museum and to arrange an exhibition titled "How do archaeologists work?" to disseminate the scientific value of the archaeological investigation. Indeed, they were asked to define an interactive visit by using objects (e.g., a book reporting clues to identify the excavation site, aero-reconnaissance photos of potential excavation sites, several digging tools), conceived by the participants as smart. The participants were provided with a scenario including details (e.g., the smart objects to be adopted) to help them in elaborating ideas and in shaping up their thoughts. A brainstorming was organized to promote the generation of ideas before their implementation. It was inspired by [46-48] and was structured in the following four steps:

1. Each member of the group proposes 3-4 rough ideas of SIE (5 minutes)
2. Each member illustrated ideas to his/her partner and the partner provided feedback (5 minutes)
3. Each member separately refined his/her ideas thanks to the partner's feedback (10 minutes)
4. The group discussed the new ideas in order to select the best one or to create a new one by merging some of their ideas (20 minutes).

The participants used a flipchart to sketch their ideas. The final idea with indications of possible smart objects, custom attributes and smart object behaviour were transcribed on a blank sheet. This first phase lasted about 1 hour for every group.

The second phase was related to the SIE implementation by using each of the three systems. To avoid carry-over effect, the systems ordering was counterbalanced according to a Latin Square design. The phase started with a demo of the first paradigm by showing examples of custom attributes creation. Then, the participants had about 5 minutes to get familiar with the specification of some custom attributes, with the possibility to ask the facilitator for help. After the training, the group started to define the custom attributes identified in the first phase; they were also free to introduce further attributes conceived during the system usage. Then, they had to use such custom attributes to define ECA rules governing the smart-object behaviour. To simplify and lighten the ECA rule creation, whose validity was already assessed in previous experiments [18], the participants were asked to write down the ECA rules on a paper sheet, where empty schemas of rules were reported. This phase lasted 15 minutes. In the end, they filled in an online questionnaire about the system they used. Before repeating the same procedure with the next system, the group was invited to relax for 5 minutes.

At the end of this phase, a paper questionnaire was administered to compare user satisfaction with the three systems. It lasted about 90 minutes.

At the end of each day, once all the three groups had completed the second phase, a focus group was conducted with all the 6 participants. Topics like the experience in using the systems, in working in a group, in creating SIEs were discussed.

The experimental study took place in quiet university rooms. In order to create an environment familiar to the participants, the rooms were enriched by placing on desks and shelves material typical of archaeologist's offices: books, objects, tools used in the archaeological investigation process, pictures of excavation campaigns tagged with QR codes. The three rooms were identically equipped. In each of them, the apparatus for one of the three systems was previously installed.

Data Collection

Both quantitative and qualitative data were collected through 1) the reports of the SIEs participants created during the first conception phase, 2) the notes taken by the observer on significant behaviours or externalized comments of the participants during the three phases, 3) the answers to the questionnaires the participants filled in during the study. All the interactions and focus group discussions were audio-video recorded.

Initially, the participants filled in a questionnaire for collecting demographic data and their competences on IT, especially on using smartphones, smart objects and interactive displays.

A second questionnaire, organized in 5 sections, was used to evaluate each system during the second phase. The first section included the AttrakDiff questionnaire consisting in 28 seven-step items whose poles are opposite adjectives (e.g. "confusing - clear", "unusual - ordinary", "good - bad"). It is based on a theoretical work model illustrating how the pragmatic and hedonic qualities influence the subjective perception of attractiveness giving rise to consequent behaviour and emotions. In particular, the following system dimensions are evaluated: *i) Pragmatic Quality (PQ)*: describes the usability of a system and indicates how successfully users are in achieving their goals using the system; *ii) Hedonic Quality - Stimulation (HQ-S)*: indicates to what extent the system support those needs in terms of novel, interesting, and stimulating functions, contents and interaction- and presentation-styles; *iii) Hedonic Quality - Identity (HQ-I)*: specifies to what extent the system allows user to identify with it; *iv) Attractiveness (ATT)*: describes a global value of the system based on the quality perception.

The second section included the Creativity Support Index (CSI), a psychometric survey to evaluate the ability of a tool in supporting users engaged in creative works and which aspects of creativity support may need attention [33]. The CSI measures 6 dimensions of creativity support: *Exploration*, *Expressiveness*, *Immersion*, *Enjoyment*, *Effort*, and *Collaboration*.

The third section proposed the NASA-TLX questionnaire, used as "Raw TLX" [49]. It is a 6-item survey that rates perceived workload in using a system through 6 subjective dimensions, i.e., *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Performance*, *Effort* and *Frustration*, which are rated within a 100-points range with 5-point steps (lower is better). These ratings were combined to calculate the overall NASA-TLX workload index [50].

The fourth section presented the new UES (User Engagement Scale) short-form, derived from the UES long form. It is a 12-item survey used to measure the user engagement, a quality characterized by the depth of a user's investment when interacting with a digital system [51], which typically results in positive outcomes [52]. This tool measures user engagement by summarizing an index that ranges from 0 to 5. It also provides detailed information about four dimensions of user engagement, i.e., *Focused Attention* (FA), *Perceived Usability* (PU), *Aesthetic Appeal* (AE) and *Reward* (RW).

The last section had two open questions about what participants liked and disliked about the system.

The third questionnaire was administered at the end of the second phase, i.e., when the participants had used all the three systems. It evaluated the participant's satisfaction asking them to rank the three systems based on their *Utility*, *Completeness* and *Ease of use* (from 1 to 3, 1 is the best), and to vote for the best system.

It is worth noticing that the results of the analysis of the qualitative data collected during the first conception phase, the notes taken by the observer during the three phases, of the CSI and of the third questionnaire are reported in [19]. In this article, we concentrate on the data collected through the second questionnaire. In addition, a further analysis was performed also considering the CSI and the data coming from the third questionnaire in order to identify possible trade-offs among the different analysis dimensions in relation to the design of smart environments.

One-way repeated measures ANOVAs (all Greenhouse–Geisser corrected) with posthoc pairwise comparisons (Bonferroni corrected) were adopted to analyse CSI, NASA-TLX, AttrakDiff and UES results and some efficiency measures, such as the number of smart objects and custom attributes involved in the created SIEs. Friedman test was adopted to analyse differences in systems ranking, with Wilcoxon signed-rank test used as posthoc pairwise comparisons. A Pearson product-moment correlation was used to determine, for each system, the relationship between CSI scores with number of custom attributes/number of smart objects. A p-value <0.10 was considered as a threshold for statistically significant results for all the previous tests.

Analysis and results

Support to creative design of SIEs

By using the CSI questionnaire, we measured the perception that designers had about the creativity support. The three systems obtained a CSI score close to 80/100, which means very good support for creative design of SIEs (*Explorative* $\bar{x}$ = 80.25, SD = 11.56; *Tangible* $\bar{x}$ = 78.79, SD = 13.55; *Tactile* $\bar{x}$ = 78.25, SD = 13.47, see Figure 4), without significant differences ($F(1.994, 33.901) = .178$, $p = .837$, partial $\eta^2 = .010$).

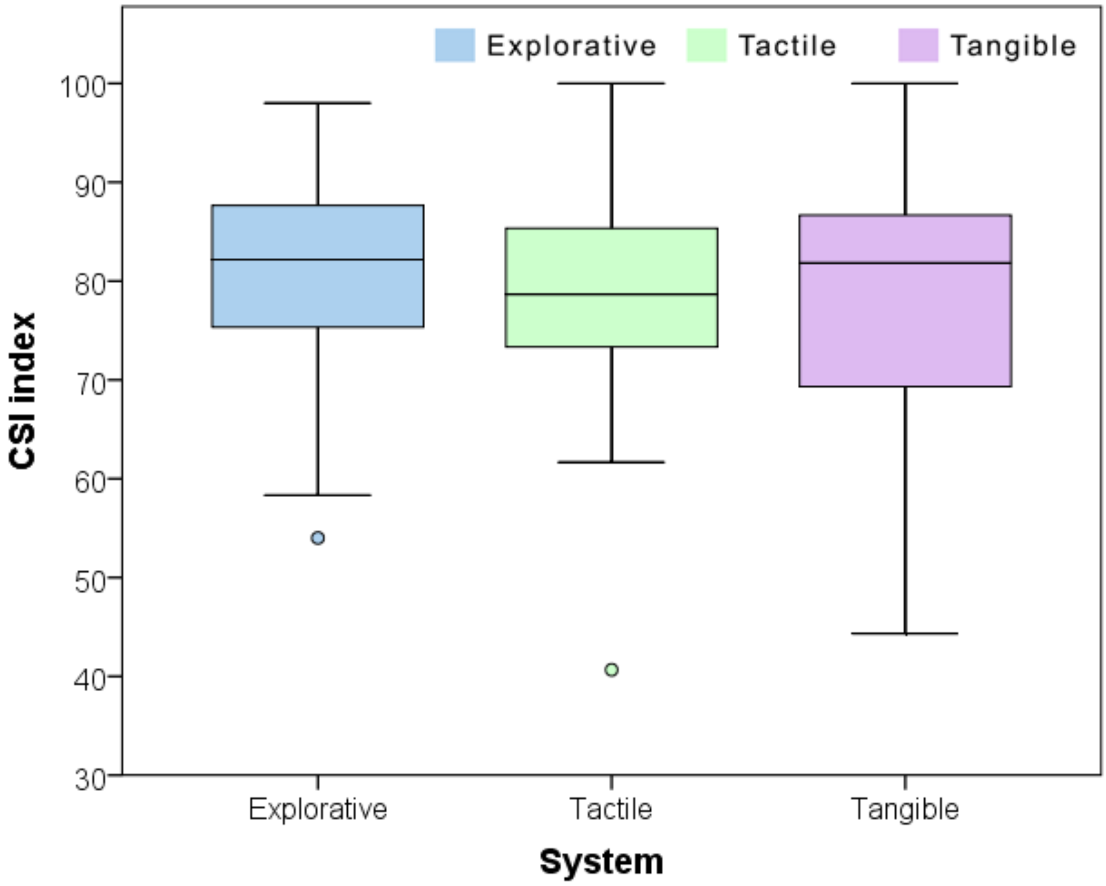


Figure 4. CSI scores for each system. Higher score is better.

The mean and the standard deviation of the CSI dimensions for each system were reported in Table 1 and depicted in Figure 5.

	Explorative		**Tactile**		**Tangible**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Collaboration	55.67	18.87	48.83	24.81	51.78	27.81
Effort	51.39	26.04	49.78	27.73	53.72	27.20
Enjoyment	24.17	24.05	31.00	23.81	30.00	26.66
Exploration	61.83	24.28	54.61	19.29	50.28	22.63
Expressiveness	30.50	18.71	32.50	21.66	29.50	20.19
Immersion	17.22	20.94	18.06	15.37	21.11	19.79

Table 1. Mean and the standard deviation of the CSI dimensions for the three systems.

For the Exploration dimension, a significant difference emerged (F(1.941, 33.000) = 2.744, p = .080, partial η^2 = .139). However, post-hoc test was not able to detect specific differences in the pairwise comparison; the only notable result is that the *Explorative* system score resulted higher than the one of the *Tangible* system (+11.556 points, SE = 5.228, p = 0.123), but with a p-value slightly greater than the 0.1 threshold. Thus, we consider this as a positive trend in favour of the *Explorative* system. No differences emerged for the other dimensions, i.e., Collaboration (F(1.964, 33.394) = .757, p = .475, partial η^2 = .043), Effort (F(1.826, 31.045) = .253, p = .758, partial η^2 = .015), Enjoyment (F(1.995, 33.918) = .918, p = .409, partial η^2 = .051), Expressiveness (F(1.487, 25.277) = .323, p = .663, partial η^2 = .019), and Immersion (F(1.831, 31.124) = .539, p = .573, partial η^2 = .031).

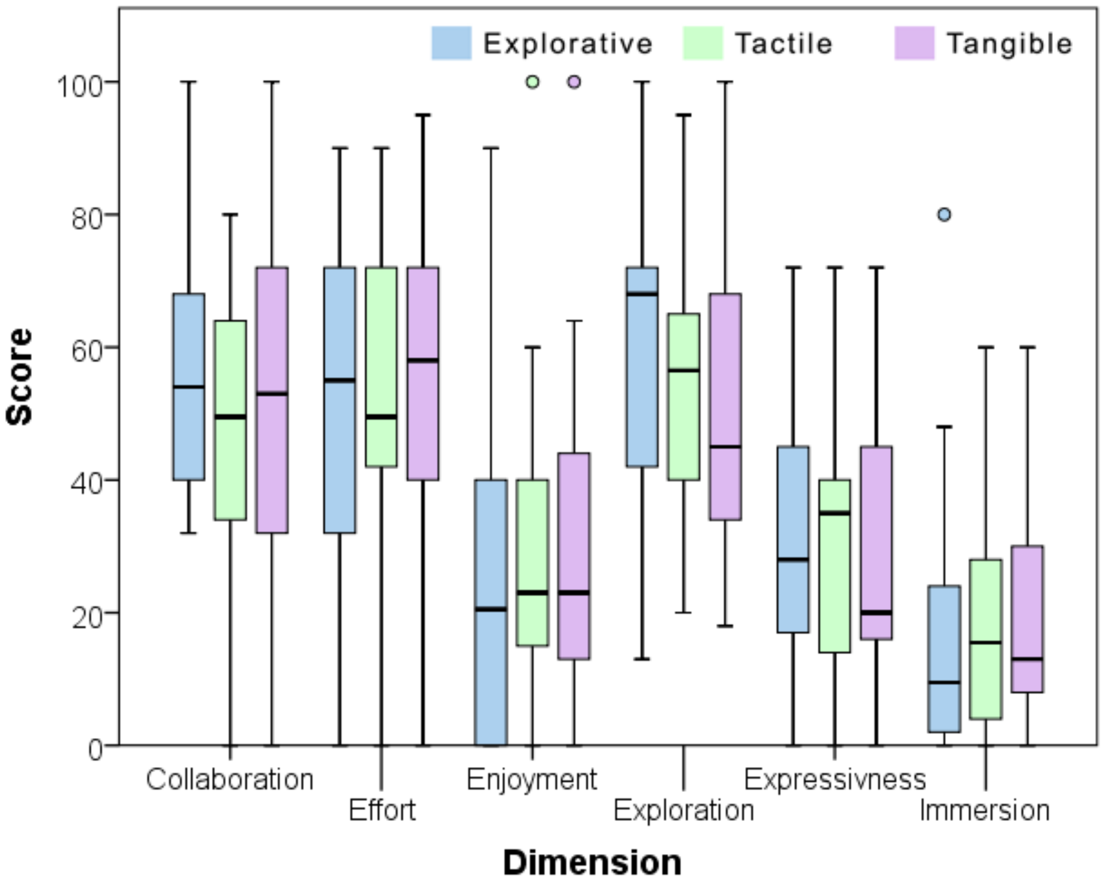


Figure 5. Scores for each CSI dimension of the three systems. Higher score is better.

Inspired by [44], a further analysis was carried out. It considered two variables as possible indicators capable of objectively representing how creative an SIE design process is: *i)* the number of useful custom attributes and *ii)* the number of useful smart objects. In both cases, 'useful' indicates those attributes and ECA rules that participants actually included in their final SIE. Indeed, almost 10% of CAs and ECA rules conceived during the ideation phases were not used in the SIEs, thus resulting useless. Groups specified a larger set of attributes when using the *Explorative* system (*Explorative* $\bar{x}$ = 6.88 SD = 3.78, *Tactile* $\bar{x}$ = 4.22 SD = 1.98, *Tangible* $\bar{x}$ = 4.55 SD = 1.66) with significant statistical differences (*F*(1.277, 91.333) = 3.328, p = .094, partial η^2 = .294). In particular, the *Explorative* score resulted higher than the one of the *Tactile* (+2.667 attributes, SE = .986, p = 0.081). No difference emerged in the other pairwise comparisons.

The number of smart objects considered in the SIE was similar while using the three systems (*Explorative* $\bar{x}$ = 3.22 SD = 1.86, *Tactile* $\bar{x}$ = 2.55 SD = .73, *Tangible* $\bar{x}$ = 2.55 SD = .73), without significant differences (*F*(1.181, 9.449) = 1.067, *p* = .342, partial η^2 = .118).

We also investigate existing correlation between these values CSI results and number of custom attributes(CAs)/smart objects involved in the SIEs. A significant correlation emerged for the *Tangible* system between the CSI score and number of smart objects (r = .394, p = .05). In the rest of the cases, there were not statistically significant correlations for the *Explorative* (CA: r = .268, p = .141; smart objects: r = .286, p = .125), *Tactile* (CA: r = .175, p = .244; smart objects: r = .050, p = .422) and *Tangible* (CA: r = .251, p = .157) systems.

User eXperience (UX)

An overview of the AttrakDiff results is represented by the portfolio diagram shown in Figure 6, which summarizes the hedonic (HQ) and pragmatic (PQ) qualities of the three systems according to the respective confidence rectangles. The bigger the confidence rectangle, the less the certainty on the region it belongs to. It is evident that the performance of the three systems is quite similar and very good. The systems have a high HQ and PQ and are classified as *desirable* products, a very promising UX. A slight difference occurs between the *Explorative* and *Tangible* systems: the first one has a lower PQ but higher HQ, meaning that its usability is slightly lower than the other systems but the users felt anyway playful sensations while interacting.

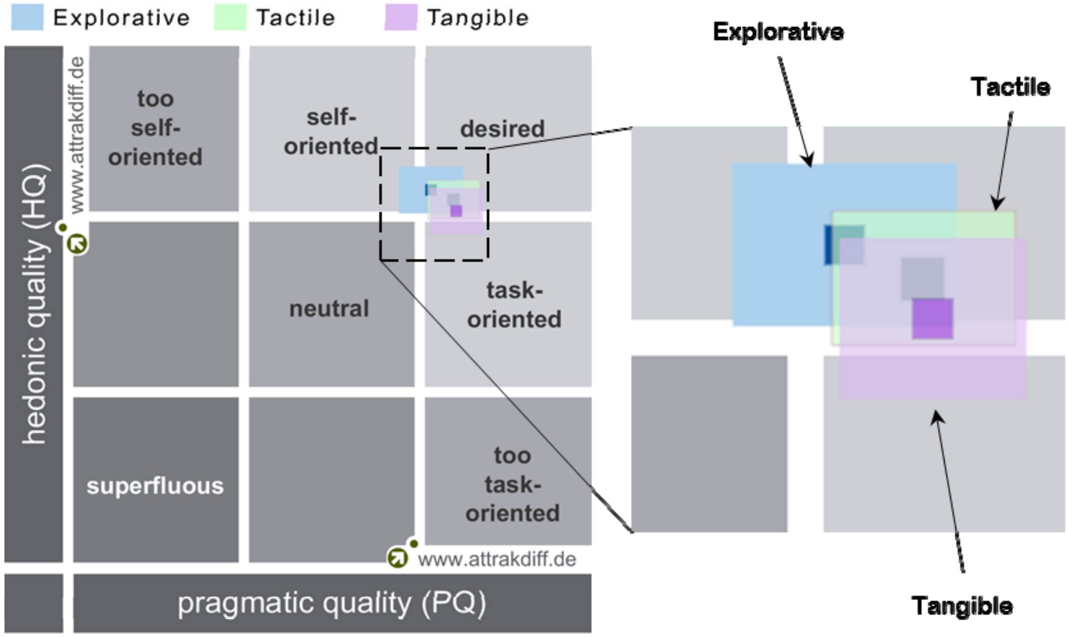


Figure 6. Portfolio diagram depicting AttrakDiff results of the three systems.

	Explorative		Tactile		Tangible	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Attractiveness (ATT)	5.91	.71	6.15	.61	5.94	.56
Hedonic Quality - Identity (HQ-I)	5.21	.67	5.02	.59	5.10	.65
Hedonic Quality - Stimulation (HQ-S)	5.48	.63	5.42	.58	5.16	.75
Pragmatic Quality (PQ)	5.15	.77	5.40	.66	5.46	.68

Table 2. Mean and the standard deviation of the AttrakDiff dimensions for the three systems.

Details about the mean and the standard deviation of the AttrakDiff dimensions for each system were reported in Table 2 and depicted in the diagram of Figure 7. The ANOVA test revealed that there are no statistically significant differences in PQ (F(1.711, 29.090) = 1.226, p = .306, partial η2 = .067), HQ-I (F(1.998, 33.796) =.354, p = .703, partial η2 = .020), HQ-S (F(1.75, 29.754) = 1.166, p = .320, partial η2 = .064) and ATT (F(1.71, 29.078) = .838, p = .426, partial η2 = .047).

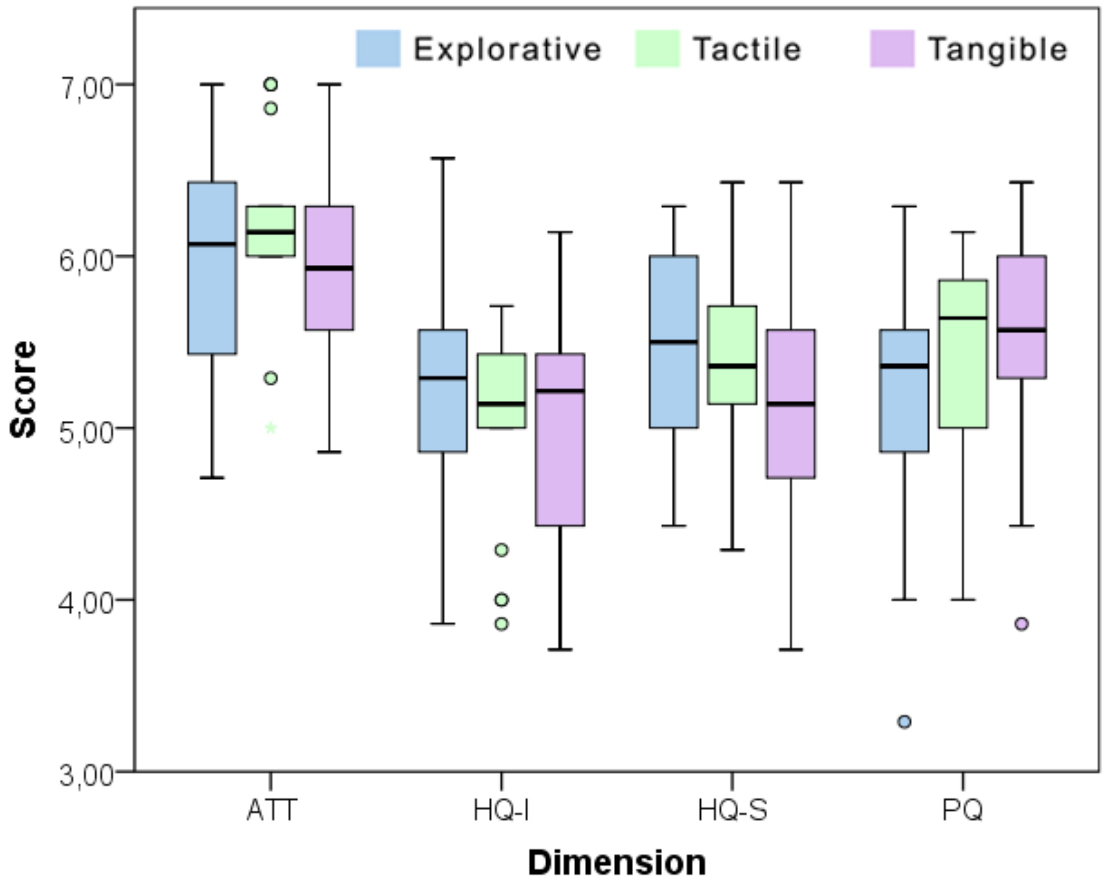


Figure 7. Average values of the three systems detailed for PQ, HQ-I, HQ-S and ATT dimensions (Y scale ranges from 1 to 7, we reduced the Y scale to improve the graph readability). Higher score is better.

User Engagement

UES short-form provided indications about the systems' user engagement. Figure 8 shows the UES indexes of the three systems (*Explorative* $\overline{x}$ = 4.02, SD =.57; *Tangible* $\overline{x}$ = 4.11, SD =.54; *Tactile* $\overline{x}$ =4.18, SD =.48). There are not statistically significant differences between them ($F(1.796, 30.524) = 1.595$, $p = .220$, partial $\eta 2 = .086$).

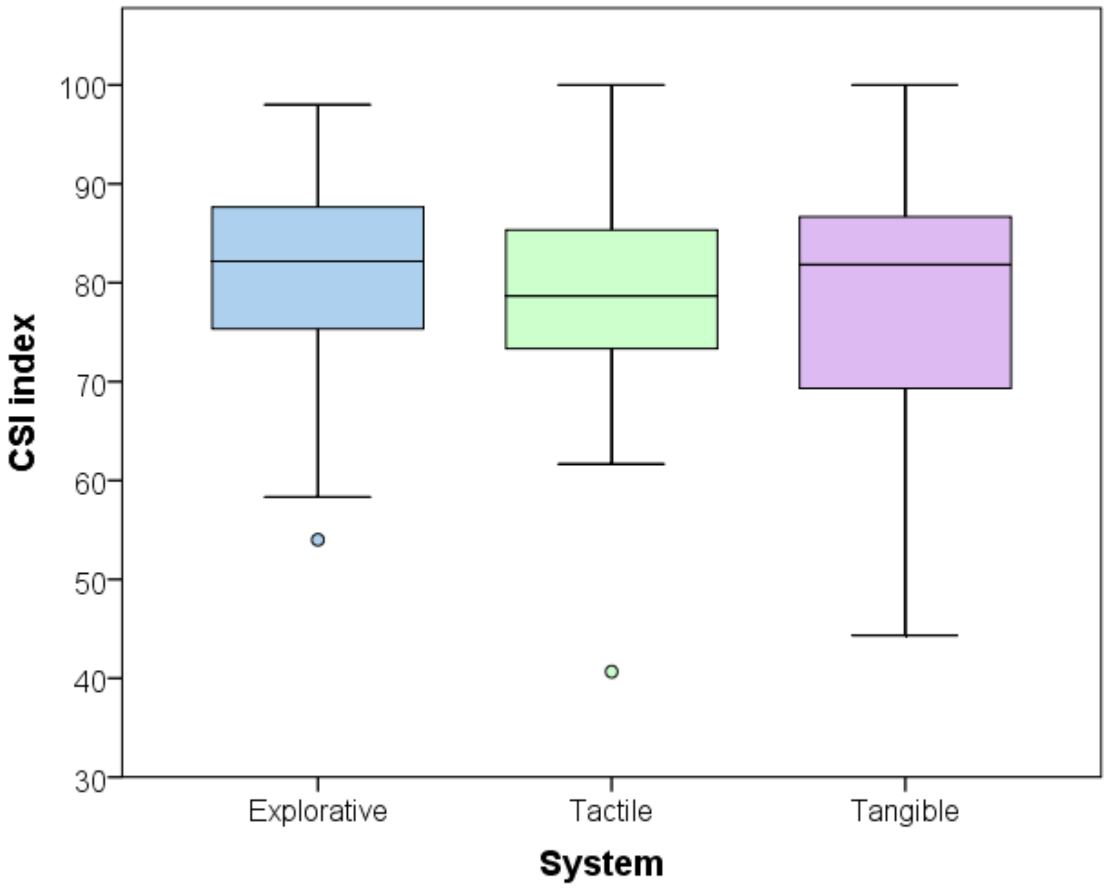


Figure 8. The boxplot chart depicting UES score results of the three systems. Higher score is better.

A more detailed analysis was also carried out to investigate systems differences with respect to the UES dimensions. The mean and the standard deviation of the UES dimensions for each system were reported in Table 3.

	Explorative		**Tactile**		**Tangible**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Aesthetic Appeal (AE)	4.17	.75	4.46	.54	4.15	.68
Focused Attention (FA)	3.46	.41	3.37	.53	3.33	.58
Perceived Usability (PU)	3.89	.65	5.42	.58	3.91	.49
Reward (RW)	3.59	.96	3.96	.73	5.46	.68

Table 3. Mean and the standard deviation of the UES dimensions for the three systems.

Figure 9 depicts the UES scores of each system calculated on each dimension. The ANOVA test demonstrated that there are no statistically significant differences between the three systems in term of AE ($F(1.483, 25.216) = 1.840$, $p = .186$, partial $\eta2 = .198$), FA ($F(1.907, 32.427) = .238$, $p = .779$, partial $\eta2 = .014$), PU ($F(1.336, 23.216) = 2.096$, $p = .157$, partial $\eta2 = .110$) and RW ($F(1.270, 21.593) = 1.539$, $p = .234$, partial $\eta2 = .083$).

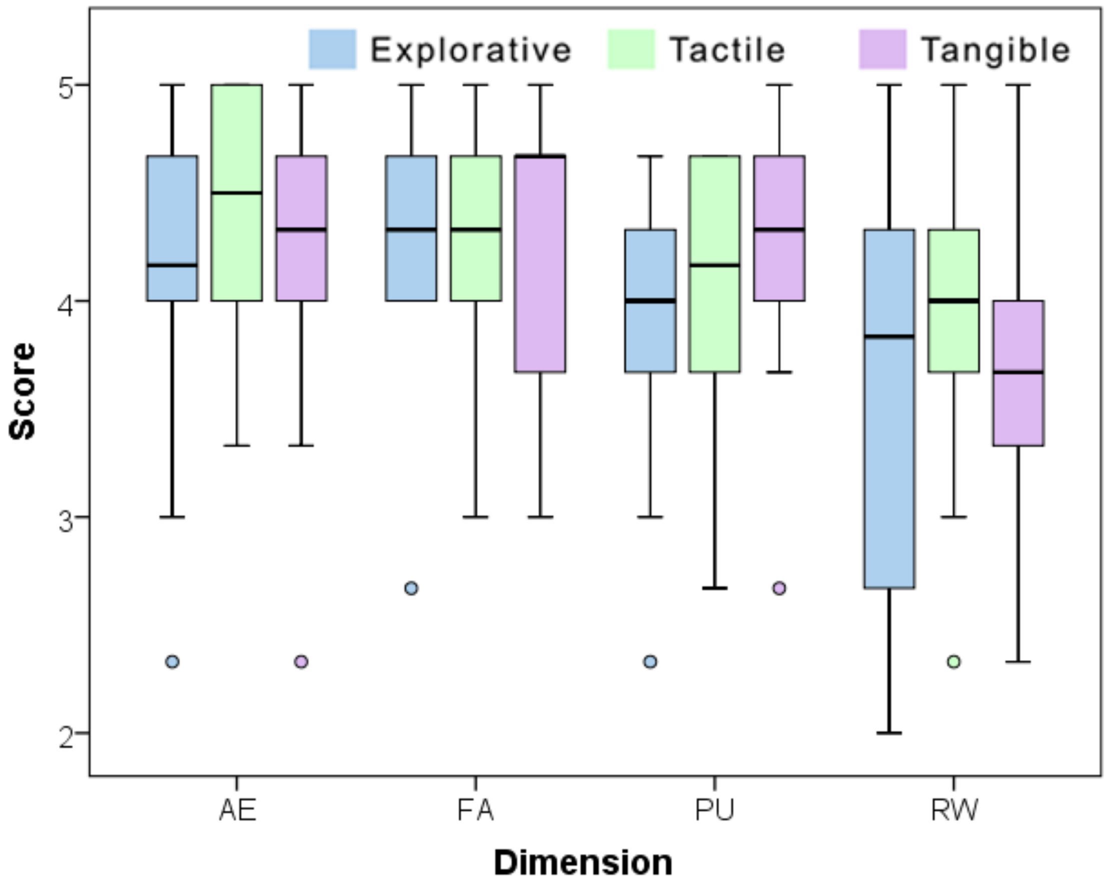


Figure 9. The boxplot chart depicting UES score dimensions of the three systems. Higher score is better.

Workload

The workload data gathered through the NASA-TLX are depicted in Figure 10 (*Explorative* $\bar{x}$ = 38.24, SD =12.21; *Tangible* $\bar{x}$ = 33.04, SD =12.43; *Tactile* $\bar{x}$ =29.87, SD =10.30). The ANOVA test revealed a significant differences between the three systems ($F(2, 34) = 4.187$, $p = .024$, partial $\eta 2 = .198$), and a posthoc analysis helped us identify a significant difference of 8.37 points (*SE=2.992,* $p < 0.05$) between the *Tactile* and the *Explorative* systems.

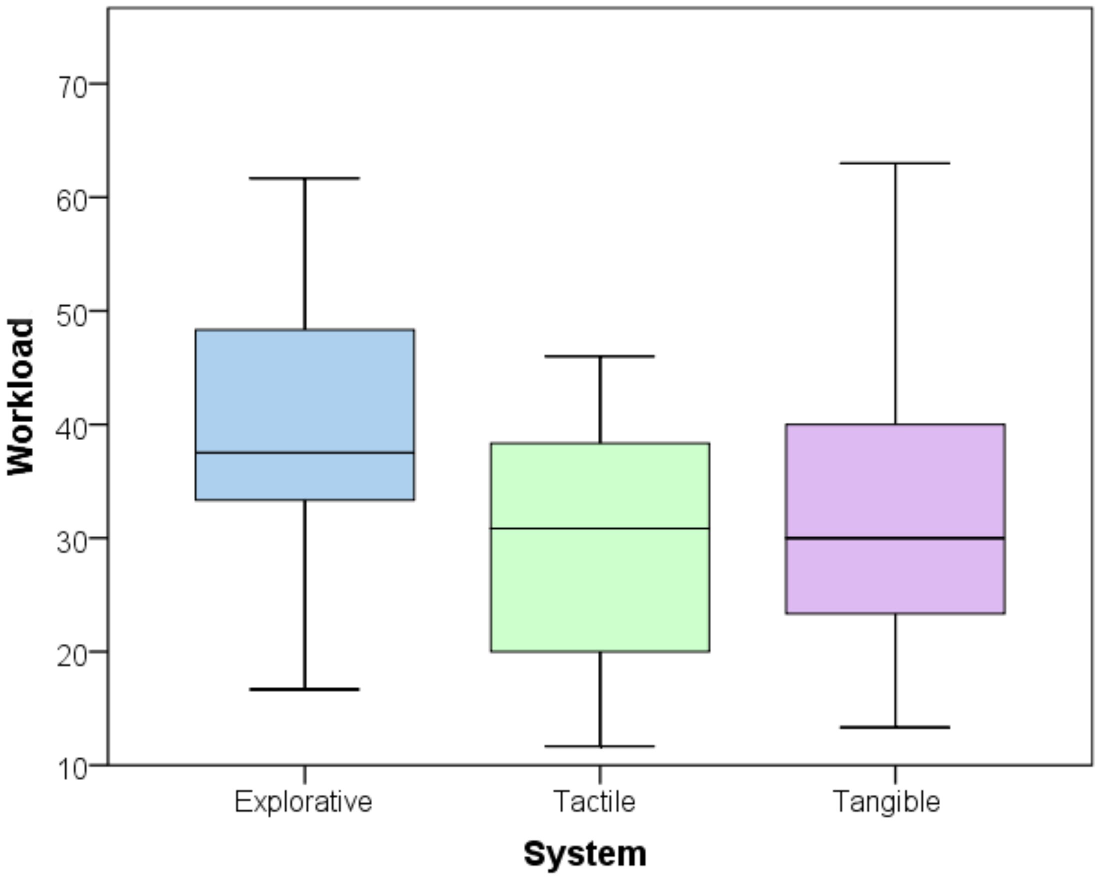


Figure 10. NASA-TLX workload of the three systems. Lower score is better.

The mean and the standard deviation of the NASA-TLX dimensions for each system were reported in Table 4 and depicted in Figure 11.

	Explorative		**Tactile**		**Tangible**	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Effort	43.33	24.73	36.11	20.62	37.22	22.18
Frustration	32.22	26.69	21.67	10.43	27.78	20.45
Mental Demand	57.78	25.33	38.33	21.49	48.33	25.72
Performance	35.56	16.53	33.33	12.83	33.33	14.55
Physical Demand	18.89	8.32	19.44	9.98	17.78	14.37
Temporal Demand	41.67	25.50	30.56	19.24	33.89	22.27

Table 4. Mean and the standard deviation of the NASA-TLX dimensions for the three systems.

A significant difference emerged for the Mental Demand dimension ($F(1.791, 30.441) = 3.558$, $p = .045$) and a posthoc analysis revealed that the *Tactile* system was scored better than the *Explorative* one, with a difference of 19.44 points (*SE=8.063,* $p = 0.082$). No significant differences emerged for Effort ($F(1.515, 25.759) = 1.377$, $p = .265$, partial $\eta 2 = .075$), Frustration ($F(1.944, 33.048) = 1.634$, $p = .211$, partial $\eta 2 = .088$), Performance ($F(1.885, 32.052) =.328$, $p = .710$, partial $\eta 2 = .019$), Physical Demand ($F(1.274, 21.665) =.126$, $p = .787$, partial $\eta 2 = .007$) and Temporal Demand ($F(1.509, 25.654) = 2.457$, $p = .117$, partial $\eta 2 = .126$).

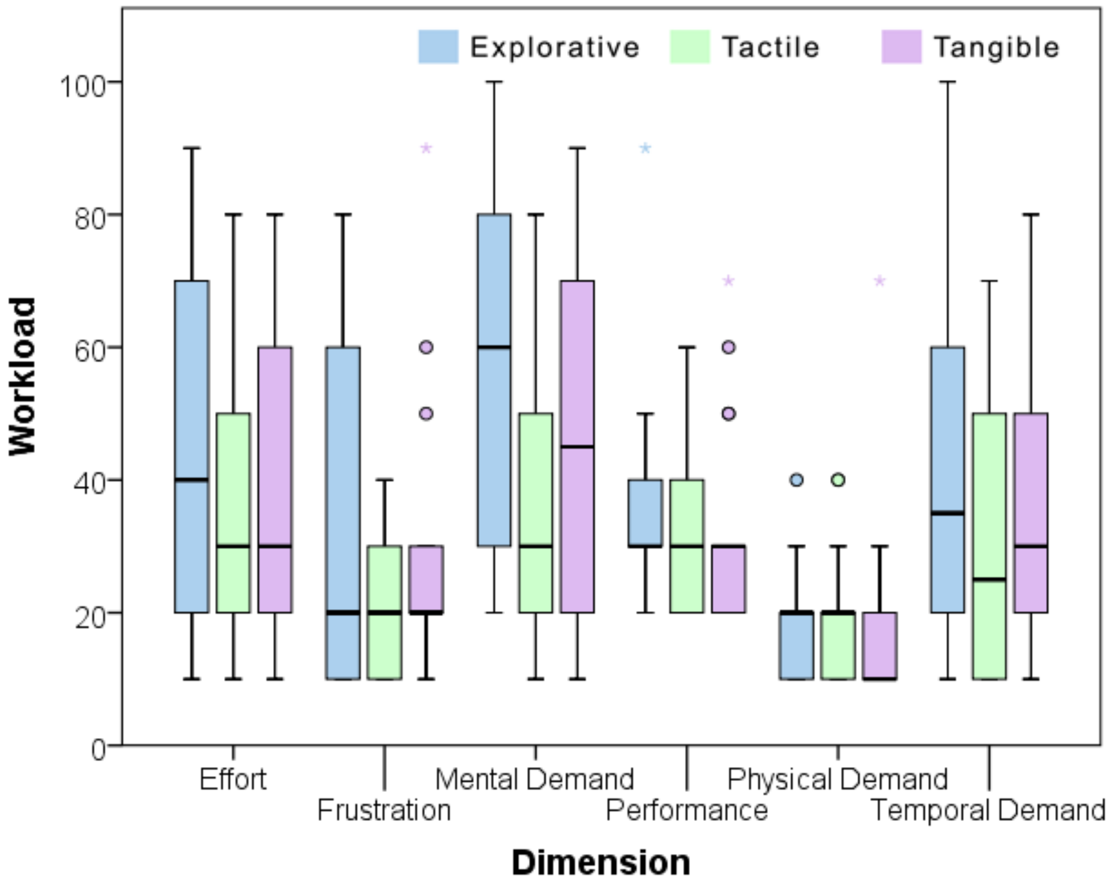


Figure 11. NASA-TLX dimensions workload of the three systems. Lower score is better.

Satisfaction

The third questionnaire revealed differences on how the participants considered the systems in relation to *Completeness*, *Utility* and *Ease of Use* and their overall preference on one of the three systems.

Regarding the Completeness, the *Explorative* and *Tactile* were considered the best systems (*Explorative* $\bar{x}$ = 1.72, *Tactile* $\bar{x}$ = 1.78, *Tangible* $\bar{x}$ = 2.50), with some significant difference ($\chi^2(2) = 7.778$, $p = 0.034$): *Tactile* was significant better than *Tangible* ($Z = -1.960$, $p = 0.050$) and *Explorative* was significant better than *Tangible* ($Z = 2.854$, $p = 0.004$). Regarding the Utility, the *Explorative* and *Tactile* were considered the best systems (*Explorative* $\bar{x}$ = 1.83, *Tactile* $\bar{x}$ = 1.78, *Tangible* $\bar{x}$ = 2.39), with significant differences ($\chi^2(2) = 4.111$, $p = 0.128$), even if the p-value was slightly greater than the 0.1 threshold: the *Tactile* system was perceived better than *Tangible* ($Z = -.824$, $p = 0.068$). Regarding the Ease of Use, the system rankings were quite similar (*Explorative* $\bar{x}$ = 1.89, *Tactile* $\bar{x}$ = 2.06, *Tangible* $\bar{x}$ = 2.06) without significant differences ($\chi^2(2) = .333$, $p = 0.846$).

These results are coherent with the way participants voted for the best system: the *Tactile* got 9 votes, the *Explorative* 6 votes and the *Tangible* 3 votes.

Threats to Validity

In this section, we analyse some issues that may threaten the validity of the experimental study, also to highlight under which conditions the study design offers benefits that can be exploited in other contexts, and under which circumstances it might fail.

Internal validity

Internal validity can be threatened by some hidden factors compromising the achieved conclusions:

- *Learning effect*. In our experiment, this factor was minimized by counterbalancing the order of the systems according to a Latin Square design.
- *Subject experience*. It was alleviated by the fact that none of the subjects had any experience with the experimented systems, as well as with similar systems in general.
- *Subject-expectancy effects.* Students are not the best participants for an experimental study due to the subject-expectancy effect they can produce, i.e., a form of reactivity that occurs when a research subject expects a given result and therefore unconsciously affects the outcome. We mitigated this effect by masking details that could produce bias. In particular, we presented the experiment to the participants in a way that suggests that we had no stake in the outcome. For example, we introduced all the experimental systems as already available tools that we wanted to analyse during the creation of SIEs; furthermore, in order to foster the credibility of this aspect, we developed our systems with a professional look-and-feel.
- *Method authorship*. We eliminated the biases that different facilitators running the experiment could introduce, as we had the same instructor for every session of the study. In this way, we avoided any variability in the initial training as well as in the way users had been observed.
- *Information exchange*. Since the study took place over 3 days, it is difficult to be certain whether the involved subjects did not exchange any information. However, the participants were recruited during the exams period thus, for many of them, it was difficult to communicate. The participants were asked to return all the material (e.g., the booklet) at the end of each session. We asked the participants that typically study and travel together to perform the test in the same session.
- *Study venue.* In order to create an environment familiar to the participants, the rooms where the experimental study was carried out were enriched with material typical of archaeologist's offices and other possible objects and tools that could be included in the SIE to be designed, included pictures of excavation campaigns tagged with QR codes, necessary for the Explorative system. However, we do acknowledge that performing the study in a natural environment, such as an archaeological museum, would probably have provided more leverage to the participants' creativity. Such a stimulus would have been particularly beneficial for the Explorative system.
- *Understandability of the material*. A pilot study involving three groups was performed to evaluate the system reliability and the research methodology (e.g., time constraints, coding techniques, video-recording activities), as well as the understandability of experimental procedures and materials.

External validity

External validity refers to the possible approximation of truth of conclusions in the attempt to generalize the results of the study in different contexts. With this respect, the main threats of our study are:

—*Users age and domain experience.* Since the study participants were young students not experienced with IoT and systems for SIE design, we have to take into account two potential limitations of the study results. The first one is the participants' age that limits the prediction of the benefits of the systems to older people. Thus, we can safely accept the experiment results for *digital natives* [53] but further studies have to be carried out including older people. The second potential limitation is related to the participants' domain experience. Rather than students of the Bachelor's or Master's degree in archaeology, the perfect participants would have been professional guides with experience in conducting guided tours, familiar with and able to master cutting-edge technologies. In our initial studies conducted to preliminarily assess the validity of the domain-oriented semantics [10], we already involved professional guides. Unfortunately, it is very hard to recruit people with this background with a statistically significant numerousness. Thus, students with a background in archaeology represented the best choice that mediates the need of adequate skills and a large sample. To mitigate this problem, we chose a theme of the exhibition consistent with the skills of archaeology students, i.e., the exhibition of tools they use regularly during their activities. In addition, we tried to include in as many pairs as possible a participant who had experience as a guide.

—*SIE Complexity*. The scenario used for the study asked participants to act as curators of a museum. It was designed with the help of professional guides that took into account the participants skills and age. Thus, the obtained results and the proposed design indications are valid for a particular class of scenarios, i.e., SIEs for the arrangements of traditional exhibitions. More complex and significant scenarios need to be evaluated.

—*SIE Design Process*. The SIE design is a three-phase process consisting of 1) conception of a high-level idea of the SIE, 2) creation of custom attributed for the SIE and 3) definition of ECA rules. Since the brainstorming phase at the beginning of the process is not affected by the use of a specific design paradigm, and we already assessed the validity of the ECA rule creation paradigm [18], we isolated the only variable not yet validated that could impact on the process, i.e., the custom attributes creation. Since we had a within-subject design, in order to lighten the participants workload, we focused on the custom attribute creation phase, by comparing the three systems proposed in this paper. To cover the remaining process, we included the ideation phase only once before the use of the three systems, and we asked participants to write down the ECA rules on a paper sheet, according to the schemas of rules supported by our system for ECA rules [18].

— *Resulting SIE*. This paper focuses on the design of SIEs; the quality of the resulting SIE was intentionally not evaluated. Anyway, as a future step, it is important also to assess the SIE quality, to deeply understand how the adoption of one of the three systems impacts on the final result. In this direction, we already planned other sessions to design SIEs implementing different scenarios, with the aim to evaluate the overall design process and also the final SIE from the perspective of SIE final users, i.e., museums visitors in our scenario.

Conclusion validity

Conclusion validity refers to the validity of the statistical tests applied for the analysis of the collected data. In our study, this validity was ensured by applying the most common tests that are traditionally employed in Empirical Software Engineering [54]. It is worth remarking that the significance level we used in this paper ($p < 0.1$) is slightly less strict than the conventional ones ($p < .05$ or $p < .01$). This because the overall goal of this research is to identify trends and trade-offs between the analysed dimensions of the three systems [55]; the higher threshold thus allowed us to consider also those results with p-values in the range 0.1 – 0.05, which still highlight trends.

5 Emerging trade-offs

Given the results illustrated above, we investigated whether any correlation occurred, to understand in which extent each analysis dimension was related to the others. To this aim, we compared all the dimensions deriving from the questionnaires: 1) *Creativity* from CSI, 2) *Workload* from NASA-TLX, 3) *UX* from AttrakDiff, 4) *Engagement* from UES, and 5) Satisfaction (decomposed in *Utility*, *Completeness*, *Ease of Use*). We also considered an additional dimension we call 6) *Ecology* of the system, which emerged during the focus group discussions and relates to the ease of deployment of the systems in real design settings, also considering their cost-effectiveness [19]. The analysis of the *Ecology* dimension reported in the following is an estimation that took into account both the participants' comments and an analysis of factors like costs and physical space required by the system installation. The *Ecology* ranges from 0 to 100, where high values indicate systems that are very cheap and easy to be installed, while lower values indicate systems that are expensive and require technological skills to be installed.

Since questionnaires adopt different scales, to facilitate the comparison of the resulting data we 1) normalized their values in the same interval (0-100) and 2) adjusted all the scales polarity so that higher scores indicate positive values. For AttrakDiff, we normalized the native scale, which ranges from 0 to 7, in the 0-100 interval. Similarly, we normalized the UES scale, which natively ranges from 0 to 5, and the Utility, Completeness, Ease of Use scales that range from 0 to 3. No adjustments were required for CSI. In the end, the polarity of the NASA-TLX values was inverted because, natively, a higher score means a higher workload. Thus, in the remaining of this section, higher NASA-TLX values indicate light workload and vice versa.

These dimensions were analysed to find significant correlations. In particular, the Pearson correlation coefficient was used to find, for each system, significant relationships between *Creativity*, *Workload* and *UX*, while the Spearman correlation coefficient was used to identify relationships between the previous dimensions and *Utility*, *Completeness* and *Ease of Use*[1]. Both tests were one-tailored. Table 5, Table 6, and Table 7 report the resulting correlations for each system. Cells in white report the Pearson results while those ones in light-grey show Spearman results. Values in bold indicate significant correlations at the confidence levels $p < 0.1$ while the number of asterisks indicates the strength of the correlations (1 for low strength ranging from 300 to 399, 2 for medium strength ranging from 400 to 499, 3 for high strength ranging from 500 to 599, 4 for very high strength ranging from 600 to 699). Negative significant correlations have been considered to identify trade-offs.

The next sub-sections discuss for each system the resulting trade-offs analysing the negative relationship existing between the quality dimensions, whose values are also depicted in radar-charts.

Explorative system

The most evident trade-offs in the correlation analysis are between *Ease of Use* vs *Workload* ($r_s(18) = -.498$, $p = .018$), as well as between *Creativity* vs *Engagement* ($r = -.466$, $n = 18$, $p = .026$), reported in bold in Table 5.

---

[1] The Pearson correlation coefficient measures the strength and direction of associations between two variables measured on at least an *interval scale*, like it happens for NASA-TLX, AttrakDiff and CSI.

The Spearman correlation coefficient measures the strength and direction of associations between two variables measured on at least an *ordinal scale*, which is the case of the other dimensions.

Table 5. Correlation strength among all the evaluation dimensions of the Explorative system

Explorative	*Crea.*	*Work.*	*UX.*	*Enga.*	*Util.*	*Comp.*	*Ease*
Creativity	1,000	,116	-,076	**-,466****	-,087	-,089	**,313***
Workload		1,000	,341	-,092	-,222	**,379***	**-,498*****
UX			1,000	,261	-,101	**,551*****	,017
Engagement				1,000	**,371***	,283	,070
Utility					1,000	,165	**,350***
Completeness						1,000	,105
Ease of Use							1,000

Regarding the first trade-off (*Ease of Use* vs *Workload*), the *Explorative* system was perceived as easy to be used as the other ones. However, as shown by the negative correlation identified between the two dimensions, a heavier workload affects it. A possible explanation of the participants' heavy workload can be found in the need of stopping the AR visualization on the mobile device to open the overview window showing the already defined custom attributes and their association with the smart objects. This problem is in line with the heavy *Mental Demand* of this system, and it was both reported by the observers and highlighted by the participants during the focus groups.

When we designed the *Explorative* system we considered the possibility of using a device with a larger display, i.e., a tablet, in order to visualize at the same time the status of the custom attribute definition and the AR mode to discover new custom attributes. We eventually came to the conclusion that sharing the screen between the workspace status and the AR visualization would make more difficult the exploration of the surrounding environment. Summing up, even though the AR mode fulfils the "Collect" design principle for creativity-support systems [39], the decision to simplify the visualizations to privilege the AR exploration worsens the workload (in particular, *Mental Demand*) due to the separate workspace overview. In other words, the workload was worsened due to the need to explore the environment in AR mode.

The second trade-off (*Creativity* vs *Engagement*) refers to the correlation identified between the higher CSI score and the lower UES score obtained by this system. The excellent support to creative SIE design revealed by CSI was also confirmed by the highest number of custom attributes defined by the participants; for this number, the difference with the other systems was significant. During the focus groups, the participants also commented that the source objects installed in the environment provided inspiration for attribute creation. As also underlined by Shneiderman, the possibility to start from existing elements is one of the most important aspects for Creative Support Tools [39]. However, as revealed by the UES score, this system tends to be the less engaging of the three ones evaluated in the study. A possible explanation can be found in the *Reward* UES dimensions, which obtained the lowest scores, even if there are no statistically significant differences for these values between the systems. The participants felt not adequately rewarded by the usage of this system, likely due to the separate overview over custom attribute definition, which emerged as the most important limitation of this system.

Regarding the positive support to the creative process, this system presents another limitation that can yield to a third trade-off. It requires environments instrumented with tags, like QR-codes and NFC, which SIE designers can scan to copy custom attributes. Even if many different environments, e.g., museums, today exploit these technologies for an easy access to information on relevant elements, we cannot assume that SIE designers have available such an equipment. For this reason, the ecology of this system is quite low, indicating that its adoption in real contexts could be limited. Thus, the last *trade-off* we identified is *Creativity* vs *Ecology*. This means that our solution to let designers start from existing materials [39], which strongly improve creativity, negatively impacts on the system ecology because designers need to exploit tags installed in the environments. This trade-off could be generalized to the situations, even independent of the presence of specific technology in the environment where the design process takes place, in which the documentation of the available artefacts is not easily accessible.

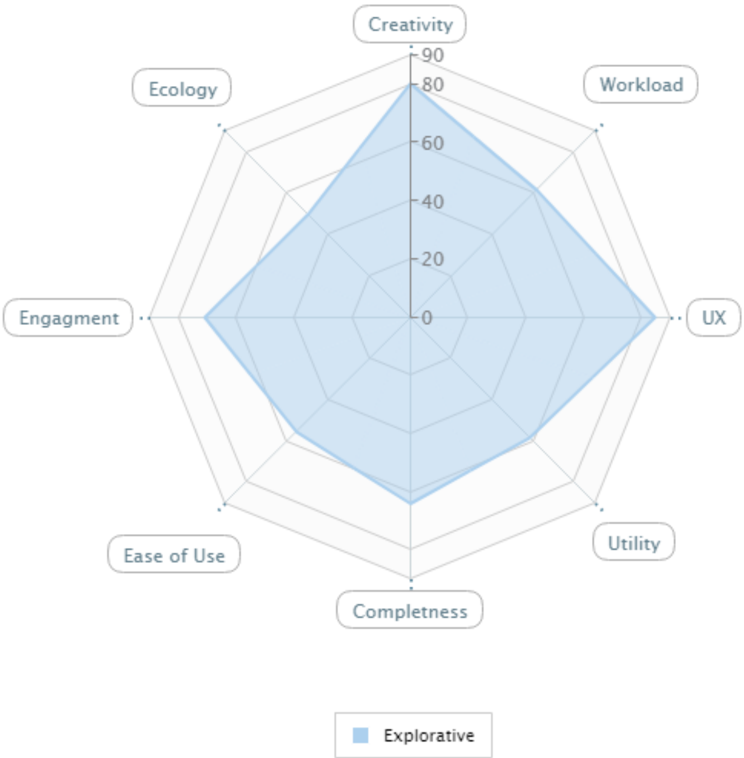


Figure 12. Radar chart that summarizes all the evaluation dimensions for the Explorative system

Tactile system

The most important trade-offs for the *Tactile* system emerged from the correlation between *UX* and *Ease of Use* ($r_s(18) = -.355$, $p = .074$), as well as between *Workload* and *Ease of Use* ($r_s(18) = -.381$, $p = .059$), reported in bold in Table 6.

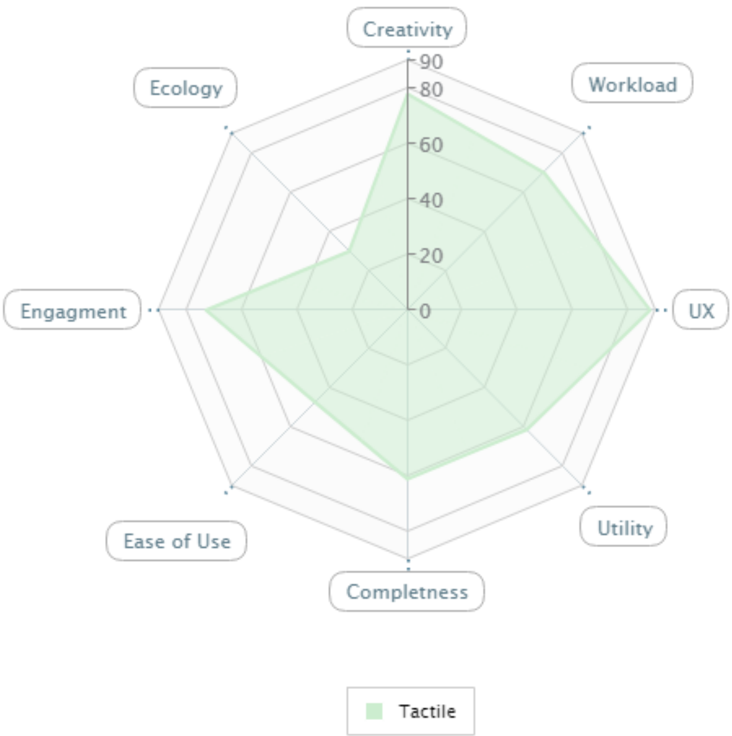
Table 6. Correlation strength among all the evaluation dimensions of the Tactile system

Tactile	*Crea.*	*Work.*	*UX.*	*Enga.*	*Util.*	*Comp.*	*Ease*
Creativity	1,000	,158	,065	,063	-,127	-,281	-,018
Workload		1,000	,121	,113	-,067	-,244	**-,381***
UX			1,000	,251	,017	-,023	**-,355***
Engagement				1,000	-,123	,042	-,307
Utility					1,000	**,344***	**,641******
Completeness						1,000	**,340***
Ease of Use							1,000

Regarding *UX* vs *Ease of Use*, this system was considered the most preferred by the study participants in terms of UX. This result emerged by the AttrakDiff data, the system ranking and the comments collected during the focus groups. However, the negative relationship with *Ease of Use* highlights that the good perception on the UX does not necessarily imply the system ease of use. A possible explanation can be found in the comments of the participants during the focus groups. They made it clear that the combination of tactile interaction and TUI enhances the UX and that, in line with some findings on tangible thinking [40], the creativity was stimulated by the use of tangible objects. The participants were indeed inspired by the objects to define new attributes. During the focus groups, some of them, for example, declared that the dice prompted the definition of an attribute representing game points. However, the combination of tactile interaction with TUI is more demanding, thus it reduces the Ease of Use of this system.

Regarding the second trade-off, i.e., *Workload* vs *Ease of Use,* the NASA-TLX results show that this system is the best one in terms of workload. This positive result can be explained considering both the observers' notes and the participants' comments: participants were facilitated in defining custom attributes without losing the overview on the overall status of the definition. This happened because the attribute definition tools and the visualization of the status were simultaneously available on the display. This had a positive benefit on participants' workload. However, the paradigm also had a negative impact on the *Easy of Use*, because the system asks users to manipulate tangible objects (i.e., both smart objects and tangible attributes) and simultaneously to use tactile interaction to manipulate the widgets visualized on the display.

Similarly to what we have seen in the *Explorative* System, the *Tactile* system presents another limitation that leads to a third trade-off, *Workload/UX* vs *Ecology*. Indeed, despite the good results for UX and Workload, the low *Ecology* of this solution could represent an important limitation to its adoption in real contexts. As also highlighted by the participants during the focus groups, a significant economic investment to purchase it and a space for its permanent installation are required. Both these aspects contribute to reduce the system ecology and thus the attitude towards using it.



3

Figure 13. Radar chart that summarizes all the evaluation dimensions for the Tactile system

Tangible system

In relation to the negative correlations reported in Table 7, three significant trade-offs can be identified, namely *Creativity* vs *Workload* (r = -.509, n = 18, p = .015), *Utility* vs *Workload* (r_s(18) = -.355, p = .074), and *UX* vs *Satisfaction*, in particular *UX* vs *Utility* (r_s(18) = -.598, p = .004), *UX* vs *Completeness* (r_s(18) = -.467, p = .025) and *UX* vs *Ease of Use* (r_s(18) = -.333, p = .088).

Table 7. Correlation strength among all the evaluation dimensions of the Tangible system

	Crea.	*Work.*	*UX.*	*Enga.*	*Util.*	*Comp.*	Ease
Creativity	1,000	**-,509*****	,157	,312	-,158	-,212	-,249
Workload		1,000	-,050	,053	**-,355***	-,060	**,373***
UX			1,000	-,158	**-,598*****	**-,467****	**-,333***
Engagement				1,000	,151	-,061	,033
Utility					1,000	,263	**,665******
Completeness						1,000	-,057
Ease of Use							1,000

Considering the first trade-off, i.e., *Creativity* vs *Workload,* from the CSI and the focus group comments it results that the *Tangible* system adequately supports a creative design process. This can be ascribed to the use of the proposed TUI that stimulates tangible thinking [40]. However, as also observed for the *Tactile* system, the TUI worsen the workload because different additional operations, like typing attributes name and values on post-it notes, are required.

The second trade-off, *Workload* vs *Utility,* confirms the light workload of this system but also makes evident a low *Utility* score. From the participants' comments it is possible to assume that the low *Utility* score is due to the almost total absence of digital facilities available during the SIE design. The participants said that such missing digital features could be instead useful to speed up and improve the SIE design process (e.g., digital typing of attribute name and values used for the Tactile system, the AR exploration mode of the *Explorative* system).

The third trade-off is between *UX* and the *Satisfaction* sub-dimensions. The *UX* measured through the AttrakDiff questionnaire was very promising because, as also emerged during the focus groups, the participants were positively impressed by the possibility to take a picture of the physical workspace that is automatically converted in a digital workspace. However, for the same reasons already explained above, constraining the participants in using this kind of TUI lowered the *Satisfaction* sub-dimensions.

The last trade-off, *Ecology* vs *Usefulness,* derives from the participants' opinion that a positive aspect of this system is its cheapness and minimalism. Indeed, this system can be conceived like an in-the-box kit that designers can put on a shelf and use as needed just by placing objects on a desk. Thus, all the dimensions that obtained negative results for this system, like *Engagement* and *Usefulness,* can be considered negatively related to the system ecology.

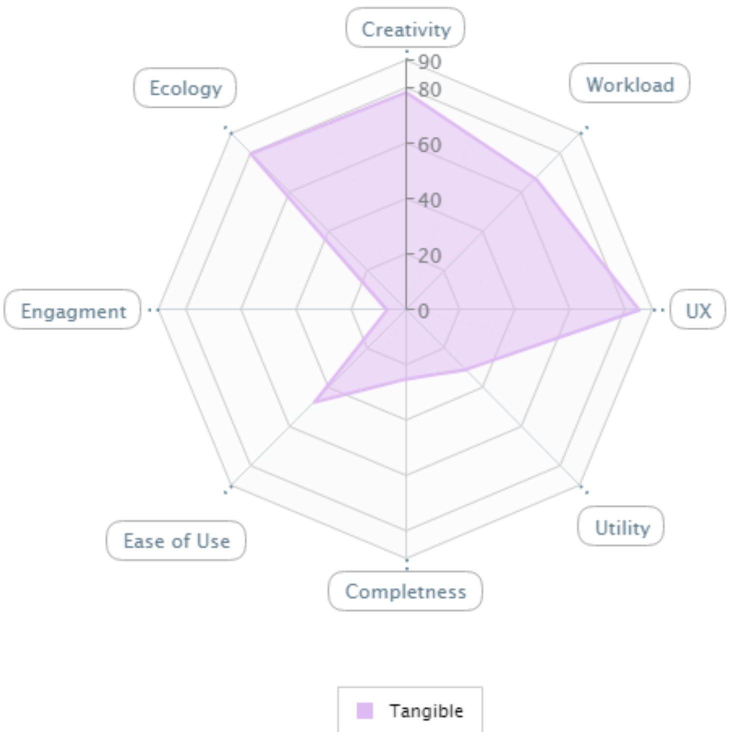


Figure 14. Radar chart that summarizes all the evaluation dimensions for the Tangible system

6 Alternatives for design paradigms and system architectural support

In this section, we discuss alternatives that can be taken into account when creating frameworks for SIE design. We also illustrate the architecture of a platform prototype, EFESTO-4SE (4 Smart Experiences), that we developed to support our methodological framework. Its modular architecture initially served the purpose of building the prototypes implementing the three design paradigms compared in the experimental study. The same platform has been then extended to support the combination of the three design paradigms.

SIE design space

One notable aspect emerged from the study is that, even if the participants were able to carry out the design process by using one system at a time, during the focus groups they debated a lot on the possibility to use different systems in a cross-paradigm and cross-device fashion. This would allow accommodating different needs and requirements. For example, the *Explorative* system resulted as the best solution for material collection, an important phase to get inspiration from existing ideas, as also underlined by Shneiderman [32]. However, this system had a negative impact on the *Workload* due to the need to switch from the exploration mode, supporting the discovery of materials, and the overview mode, to verify the current status of the custom attribute definition. TUIs, implemented both in the *Tactile* and *Tangible* systems, then resulted as an important aspect for stimulating digital thinking and creativity. Nevertheless, the *Tangible* system has a low *Ecology* and *Ease of Use*, while *Tactile* system, which has an excellent *Ecology*, has a bad *Satisfaction*. In other words, each of these systems has complementary peculiarities that, if combined, can facilitate SIE design in different ways. The radar chart reported in Figure 15 puts the emphasis on the comparisons of the three systems in relation to the different evaluation dimensions.

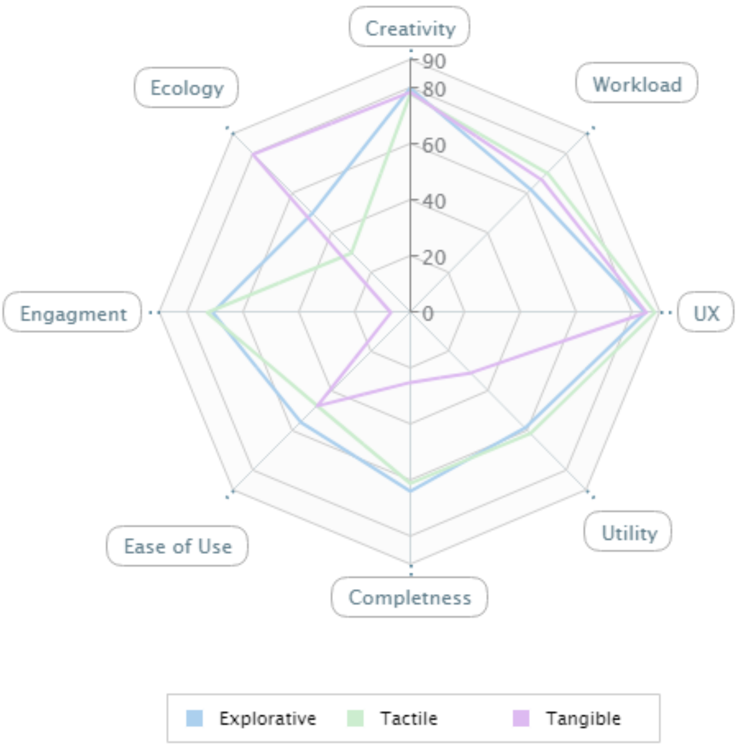


Figure 15. Radar chart comparing the three system along the evaluation dimensions

The previous observations highlight that the tuning of the final framework for SIE design has to consider different factors. For example, if designers have the possibility to collect material within a stimulating environment (e.g., at archaeological sites or museums rich of tagged works), the *Explorative* system can be used at the beginning of the design process, when collecting is the most important activity while getting an overview of the workspace definition can still be delayed. Afterwards, the *Tactile* system can be used to refine the initial workspace definition. Another proposal coming from the study participants was to start the identification of the attributes through the *Tangible* system and then to use the *Explorative* system to walk around in the environment (e.g., museum rooms) to enrich the attribute definition. Alternatively, if a brainstorming cannot be performed, due to time constraints or because the target SIE is rather simple, the *Tangible* system can be used in a first design step because tangible attributes and post-it notes anyway support a reasoning flow that participants perceived as a good substitute for brainstorming. Starting from these observations, we propose an architecture where the strengths of the three paradigms can be combined flexibly, i.e., depending on the constraints and the requirements that characterize specific design situations.

Platform organization

We here illustrate the architecture of EFESTO-4SE, which extends the one already implemented for the EFESTO-5W platform devoted to the creation of ECA rules [18]. The new platform now also supports our design methodology by offering a cross-modality and cross-device paradigm for semantic enrichment based on custom attributes. It can be considered an architectural pattern that can guide the development of flexible software environments for the creation of SIEs.

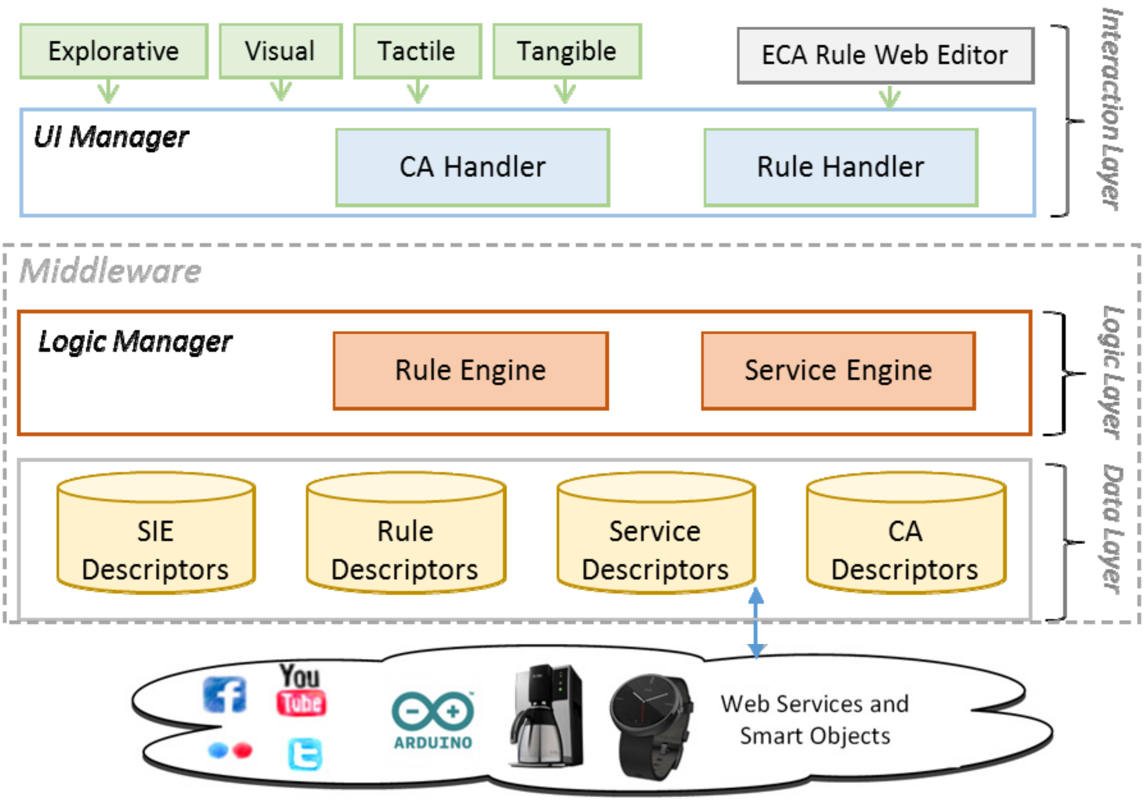


Figure 16. The overall organization of the plug-and-play platform architecture.

As illustrated in Figure 16, the architecture is based on the MVC (Model–View–Controller) pattern, which facilitates the development of a decoupled interaction layer implementing different UIs characterized by different interaction paradigms. Indeed, the peculiarity of this architecture is the possibility to adopt two kinds of "plug-and-play" UIs, i.e., the ones for creating ECA rules (characterized by multiple events/actions, as well as temporal and spatial constraints [18]) and the ones to create Custom Attributes (CAs), for example the UIs described in this article.

At the Interaction Layer, the UIs for the definition of CAs have to implement the *CA Handler* module, which is in charge of translating the users' actions into proper CA descriptions stored in the *CA Descriptors* repository. Similarly, the ECA UIs implements the *Rule Handler*, in order to convert the users' actions on the UI for rule definition into descriptors that summarize the resulting rule in terms of events, actions, conditions for spatial and temporal constraints.

The logic and data layers then constitute a middleware, installed on a Web server, which provides generalized functions and data that are shared by all the UIs. The *Logic Manager* is composed of two modules that expose RESTful APIs invoked by the UIs. In the platform, the resources that can be managed by a SIE, including the smart objects, correspond to services in charge of listening to events and activating actions. The *Service Engine* thus provides an API that the UIs can invoke to retrieve all the service events and actions. In particular, this API provides events and actions available in the *Service Descriptors*. It also interprets and translates the conditions expressed on the user-created CAs into low-level conditions on the actual events and actions exposed by services.

Using the elements defined through the UIs, the second module, the *Rule Engine*, instantiates an object representing the rule, based on a publish-subscribe event-action model [56; 57]. The *Rule Engine* then enables an immediate execution of the rule. Once the rule is instantiated, it checks every N minutes (1 minute of sample rate in our systems) if the published events are triggered; if they are triggered, the *Rule Engine* runs all the subscribed actions associated with the rule.

The *Data Layer* includes four repositories to store services, ECA rules, CAs, and SIE descriptors. Each repository exposes CRUD (Create, Read, Update, Delete) APIs to allow the remote clients, e.g., the ECA and CA UIs, to manage their data. Descriptors provide models that enable a seamless switching between different modalities at the interaction layer. This is indeed possible because all the modalities share the same representations of the SIE and of its resources. The *Service Descriptors* store all the information useful to query a service API and contributes to decoupling the registered services from the rest of the platform. The *Rule Descriptors* stores all the rules created by users. The *CA Descriptors* stores all the CA created by designers, also including the information about the service they are associated with. Finally, the SIE descriptor includes all the information about the SIEs created by each user, in terms of references to the involved services, rules and CAs used in each SIE. This descriptor orchestrates the execution of the different rules.

Combining the three design paradigms

Let us analyse how the three systems proposed in this article implement the EFESTO-4SE architecture, also reporting some technical details.

The *Tactile* solution is based on a tabletop built with a 55" LED TV. The interaction is enabled by the *UBI device*, a plug-and-play technology that converts any display or projected surface into a touchscreen [58]. A depth camera installed on top of the tabletop recognizes the smart objects and tangible attributes. The smart object recognition is based on the *SPRITS* framework [59], which uses the camera video to recognize the objects placed on the surface and labelled with fiducial markers. The tangible attributes are recognized by the *CA Handler* module, which invokes external APIs[4] to analyse the camera video and to detect the tangible attributes, also establishing their size/position/orientation on the tabletop surface. Every time users put on the tabletop a smart object or a tangible attribute, repository APIs are invoked to retrieve CAs and service information and to update the SIE repository.

The *Tangible* system has been implemented as an Android app. When users take a picture of the desk, its *CA Handler* module invokes an external API to recognize the smart objects starting from it QR-codes[4], as well as a visual recognition API to interpret the post-it text[2]. Similar to the previous system, the services, CAs and SIE data are saved in the repository.

The *Explorative* system is an Android app that implements the *CA Handler* to manage the user's actions while exploring the environment. In particular, every time a QR code related to a set of CAs is scanned, it is elaborated through an APIs (the same used for the *Tangible* system) and automatically converted in a set of attributes, which are visualized in a pop-up where the designer can select and edit the attributes, and associate them to the smart object. When users scan target smart object, the *CA Handler* stores CA, service and SIE information on the related repositories.

Since these systems share the same middleware, designers are free to use all of them, or new ones, in a cross-device fashion. For example, if they start by creating CAs with the Explorative system, then they can enrich the workspace with more CAs by using other CA UIs. Afterwards, they can move to the ECA rule Web Editor to create ECA rules on the CAs newly created. During the creation of ECA rules, designers are free to come back to the CAs creation to enrich smart objects with new and useful semantic. Indeed, the cross-device behaviour is allowed not only across the CA UIs but across all the UIs connected to the middleware and that implement the Interaction Layer modules.

7 Conclusion and Future Work

This article has discussed trade-offs that emerged by correlating some quality dimensions that we analysed in an experimental study comparing three paradigms for SIE design. A previous analysis of the data collected during the experimental study and presented in [19] allowed us to understand in which measure the three paradigms are able to support the creative process of SIE design. The results highlighted that there is not any paradigm performing better than the others; rather, each paradigm offers elements that potentiate specific aspect of the design process. For this reason, further analyses were carried out in order to identify whether any trade-offs between the consider dimensions existed and could guide the definition of a design framework able to potentiate different quality dimensions depending on the specific design context.

The study, in general, and in particular the trade-off analysis led us to identify some interesting implications for the organization of the methodology for SIE design as well as of the enabling software systems. However, the study has also paved the way for interesting directions for future work.

First of all, although the study was conducted in an environment that was purposely configured to be as realistic as possible, we recognize the need to perform field studies, conducted in real design settings. This would allow us to further validate the impact that the exploration of the environment, where the SIE is supposed to be rendered, might have on the creative design.

The combination of different design paradigms, which now is possible through the extended EFESTO-4SE platform, also needs to be evaluated and tuned. The current organization of the platform allows designers to freely adopt the techniques offered by the three paradigms in each different phase of the design process. However, this combination leads to a "new" paradigm, which needs to be validated and compared with

[2] In the current prototype, both the Tactile and the Tangible systems use Google Vision APIs (https://cloud.google.com/vision) for visual recognition of tangible attributes, smart objects, and post-it notes and bar-code.

the original ones, along the same quality dimensions that we already considered in the first study. For this, we already planned to conduct a new user study.

The quality of the final SIE and the experience of the SIE final users, namely the museum visitors in the scenario presented in this article, can also be evaluated, as a proxy for further assessing the quality of the design process.

From a technical perspective, we need to deeply understand and improve the way the three design paradigms can be combined through cross-device interaction. Finally, new interaction techniques could be investigated, such as the use of speech-based interaction to allow designers to express custom attributes and their association with smart objects. The exploration of properties offered by the surrounding environment could also exploit the automatic recognition of objects and the retrieval of related content, by similarity matches, from online repositories. Regarding this, mashup technologies could serve the purpose of collecting and integrating material from distributed data sources [60].

References

[1] Atzori, L., Iera, A., and Morabito, G. 2010. The Internet of Things: A survey. *The International Journal of Computer and Computer Networks*, 54, 15 (2010), 2787-2805.

[2] Guo, B., Zhang, D., Wang, Z., Yu, Z., and Zhou, X. 2013. Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, 36, 6 (2013), 1531-1539.

[3] Uskov, A. and Sekar, B. 2015. Smart Gamification and Smart Serious Games. In: *Fusion of Smart, Multimedia and Computer Gaming Technologies: Research, Systems and Perspectives*, D. SHARMA, M. FAVORSKAYA, L.C. JAIN and R.J. HOWLETT Eds. Springer, 7-36.

[4] Madeira, R.N., Correia, N., Dias, A.C., Guerra, M., Postolache, O., and Postolache, G. 2011. Designing personalized therapeutic serious games for a pervasive assistive environment. In *Proc. of SeGAH '11*, 1-10.

[5] Andreoli, R., Corolla, A., Faggiano, A., Malandrino, D., Pirozzi, D., Ranaldi, M., Santangelo, G., and Scarano, V. 2017. A Framework to Design, Develop, and Evaluate Immersive and Collaborative Serious Games in Cultural Heritage. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 11, 1 (2017), 1-22.

[6] Cuomo, S., Michele, P.D., Piccialli, F., Galletti, A., and Jung, J.E. 2017. IoT-based collaborative reputation system for associating visitors and artworks in a cultural scenario. *Expert Systems with Applications*, 79 (2017), 101-111.

[7] Marshall, M.T., Dulake, N., Ciolfi, L., Duranti, D., Kockelkorn, H., and Petrelli, D. 2016. Using Tangible Smart Replicas as Controls for an Interactive Museum Exhibition. In *Proc. of TEI '16*. ACM, New York, NY, USA, 159-167.

[8] Garzotto, F. and Gelsomini, M. 2018. Magic Room: A Smart Space for Children with Neurodevelopmental Disorder. *IEEE Pervasive Computing*, 17, 1 (2018), 38-48.

[9] Conti, M., Das, S.K., Bisdikian, C., Kumar, M., Ni, L.M., Passarella, A., Roussos, G., Tröster, G., Tsudik, G., and Zambonelli, F. 2012. Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber–physical convergence. *Pervasive and Mobile Computing*, 8, 1 (2012), 2-21.

[10] Ardito, C., Buono, P., Desolda, G., and Matera, M. 2017. From smart objects to smart experiences: An end-user development approach. *International Journal of Human-Computer Studies*, 114 (2017), 51-68.
[11] Costabile, M.F., Fogli, D., Mussio, P., and Piccinno, A. 2006. End-User Development: The Software Shaping Workshop Approach. In: *End User Development - Is-EUD 2006*, H. LIEBERMAN, F. PATERNÒ and V. WULF Eds. 9. Springer Netherlands, 183-205.
[12] Lieberman, H., Paternò, F., Klann, M., and Wulf, V. 2006. End-User Development: An Emerging Paradigm. In: *End User Development*, H. LIEBERMAN, F. PATERNÒ and V. WULF Eds. Human-Computer Interaction Series. Springer Netherlands, 1-8.
[13] Ardito, C., Costabile, M.F., De Angeli, A., and Lanzilotti, R. 2012. Enriching exploration of archaeological parks with mobile technology. *ACM Transaction on Computer-Human Interaction (TOCHI)*, 19, 4 (2012), 31 pages.
[14] Zancanaro, M., Not, E., Petrelli, D., Marshall, M., van Dijk, T., Risseeuw, M., van Dijk, D., Venturini, A., Cavada, D., and Kubitza, T. 2015. Recipes for tangible and embodied visit experiences. In *Proc. of MW '15*.
[15] Petrelli, D. and Lechner, M. 2014. The meSch project – Material EncounterS with digital Cultural Heritage: Reusing existing digital resources in the creation of novel forms of visitor's experiences. In *Proc. of CIDOC '14*.
[16] Ibrahim, N. and Ali, N.M. 2018. A Conceptual Framework for Designing Virtual Heritage Environment for Cultural Learning. *Journal on Computing and Cultural Heritage (JOCCH)*, 11, 2 (2018), 1-27.
[17] Rennick-egglestone, S., Brundell, P., Koleva, B., Benford, S., Roussou, M., and Chaffardon, C. 2016. Families and Mobile Devices in Museums: Designing for Integrated Experiences. *Journal on Computing and Cultural Heritage (JOCCH)*, 9, 2 (2016), 1-13.
[18] Desolda, G., Ardito, C., and Matera, M. 2017. Empowering end users to customize their smart environments: model, composition paradigms and domain-specific tools. *ACM Transactions on Computer-Human Interaction*, 24, 2 (2017), 52 pages.
[19] *Paper under review*
[20] Fischer, G. 2017. Exploring design trade-offs for quality of life in human-centered design. *interactions*, 25, 1 (2017), 26-33.
[21] Weiser, M. 1991. The computer for the 21st century. *Scientific american*, 265, 3 (1991), 94-104.
[22] Delprino, F., Piva, C., Tommasi, G., Gelsomini, M., Izzo, N., and Matera, M. 2018. ABBOT: a smart toy motivating children to become outdoor explorers. In *Proc. of AVI '18*. ACM, New York, NY, USA, 1-9.
[23] Cook, D.J., Augusto, J.C., and Jakkula, V.R. 2009. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing*, 5, 4 (2009), 277-298.
[24] Mighali, V., Fiore, G.D., Patrono, L., Mainetti, L., Alletto, S., Serra, G., and Cucchiara, R. 2015. Innovative IoT-aware Services for a Smart Museum. In *Proc. of WWW '15*. ACM, New York, NY, USA, 547-550.
[25] Piccialli, F. and Chianese, A. 2017. The Internet of Things Supporting Context-Aware Computing: A Cultural Heritage Case Study. *Mobile Networks and Applications*, 22, 2 (April 01 2017), 332-343.
[26] Coronado, M. and Iglesias, C.A. 2016. Task Automation Services: Automation for the Masses. *IEEE Internet Computing*, 20, 1 (2016), 52-58.

[27] Corno, F., Russis, L.D., and Roffarello, A.M. 2017. A Semantic Web Approach to Simplifying Trigger-Action Programming in the IoT. *Computer*, 50, 11 (2017), 18-24.
[28] Tutenel, T., Bidarra, R., Smelik, R.M., and Kraker, K.J.D. 2008. The role of semantics in games and simulations. *Computer Entertainment*, 6, 4 (2008), 1-35.
[29] Corno, F., Russis, L.D., and Roffarello, A.M. 2017. A High-Level Approach Towards End User Development in the IoT. In *Proc. of CHI '17*. ACM, New York, NY, USA, 1546-1552.
[30] Beghetto, R.A. and Kaufman, J.C. 2007. Toward a broader conception of creativity: A case for" mini-c" creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 1, 2 (2007), 73.
[31] Kaufman, J.C. and Beghetto, R.A. 2009. Beyond big and little: The four c model of creativity. *Review of general psychology*, 13, 1 (2009), 1.
[32] Shneiderman, B. 2007. Creativity support tools: accelerating discovery and innovation. *Communications of the ACM*, 50, 12 (2007), 20-32.
[33] Cherry, E. and Latulipe, C. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transaction on Computer-Human Interaction (TOCHI)*, 21, 4 (2014), 1-25.
[34] MacCormack, A., Kemerer, C.F., Cusumano, M., and Crandall, B. 2003. Trade-offs between productivity and quality in selecting software development practices. *IEEE Software*, 20, 5 (2003), 78-85.
[35] Fischer, G. and Giaccardi, E. 2006. Meta-design: A Framework for the Future of End-User Development. In: *End User Development*, H. LIEBERMAN, F. PATERNÒ and V. WULF Eds. Springer Netherlands, 427-457.
[36] Fischer, G. 2018. Identifying and exploring design trade-offs in human-centered design. In *Proc. of AVI '18*. ACM, New York, NY, USA, 1-9.
[37] Ardito, C., Desolda, G., Lanzilotti, R., Malizia, A., and Matera, M. 2017. Advanced Interaction Paradigms to Define Smart Visit Experiences in the Internet of Things Era In *Proc. of CHItaly '17*, CEUR Workshop Proceedings, Vol-1910, 1148-1152.
[38] AtomGraph. https://github.com/AtomGraph/Web-Client. Last access: February 12, 2018
[39] Shneiderman, B. 2009. Creativity Support Tools: A Grand Challenge for HCI Researchers. In: *Engineering the User Interface: From Research to Practice*, M. REDONDO, C. BRAVO and M. ORTEGA Eds. Springer London, 1-9.
[40] Ishii, H. and Ullmer, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proc. of CHI '97*. ACM, New York, NY, USA, 234-241.
[41] Kim, M.J. and Maher, M.L. 2008. The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies*, 29, 3 (2008), 222-253.
[42] Doering, T., Beckhaus, S., and Schmidt, A. 2009. Towards a sensible integration of paper-based tangible user interfaces into creative work processes. In *Proc. of CHI '09*. ACM, New York, NY, USA, 4627-4632.
[43] Hunsucker, A.J., Baumgartner, E., and McClinton, K. 2018. Evaluating an AR-based museum experience. *interactions*, 25, 4 (2018), 66-68.
[44] Oppezzo, M. and Schwartz, D.L. 2014. Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of Experimental Psychology: Learning, Memory, and Cognition (PeePs)*, 40, 4 (2014), 1142.

[45] Catala, A., Jaen, J., Dijk, B.v., and Jordà, S. 2012. Exploring tabletops as an effective tool to foster creativity traits. In *Proc. of TEI '12*. ACM, New York, NY, USA, 143-150.
[46] Burnam-Fink, M. 2015. Creating narrative scenarios: Science fiction prototyping at Emerge. *Futures*, 70 (2015), 48-55.
[47] Kohno, T. and Johnson, B.D. 2011. Science fiction prototyping and security education: cultivating contextual and societal thinking in computer security education and beyond. In *Proc. of SIGCSE '11*. ACM, New York, NY, USA, 9-14.
[48] Kudrowitz, B.M. and Wallace, D. 2013. Assessing the quality of ideas from prolific, early-stage product ideation. *Journal of Engineering Design*, 24, 2 (2013), 120-139.
[49] Hart, S.G. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 9 (2006), 904-908.
[50] Hart, S.G. and Staveland, L.E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52 (1988), 139-183.
[51] O'Brien, H. 2016. Theoretical Perspectives on User Engagement. In: *Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media*, H. O'BRIEN and P. CAIRNS Eds. Springer International Publishing, 1-26.
[52] O'Brien, H.L., Cairns, P., and Hall, M. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112 (2018), 28-39.
[53] Prensky, M. 2001. Digital natives, digital immigrants part 1. *On the horizon*, 9, 5 (2001), 1-6.
[54] Juristo, N. and Moreno, A.M. 2010. *Basics of Software Engineering Experimentation*. Springer Publishing Company, Incorporated.
[55] Fisher, R.A. 1992. Statistical methods for research workers. In: *Breakthroughs in statistics*. Springer, 66-70.
[56] Cappiello, C., Matera, M., Picozzi, M., Sprega, G., Barbagallo, D., and Francalanci, C. 2011. DashMash: A Mashup Environment for End User Development. In: *Web Engineering - ICWE 2011*, S. AUER, O. DÍAZ and G. PAPADOPOULOS Eds. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 152-166.
[57] Cappiello, C., Matera, M., and Picozzi, M. 2015. A UI-Centric Approach for the End-User Development of Multidevice Mashups. *ACM Transaction Web*, 9, 3 (2015), 1-40.
[58] UBI Interactive. http://www.ubi-interactive.com/. Last access: September 12th, 2018
[59] SPRITS. https://github.com/tommasoturchi/SPRITS. Last access: September 12th, 2018
[60] Daniel, F. and Matera, M. 2014. *Mashups: Concepts, Models and Architectures*. Springer.