



Test Plan

Deliverable D3.1

MPAT File ID: D3.1_test_plan.docx

Version: 1.0

Deliverable number: D3.1

Authors: Matthew Broadbent (ULANC)

Contributors: Doreen Ritter (RBB), Nico Patz (RBB)

Internal reviewers: Christian Fuhrhop (Fraunhofer FOKUS)

Work Package: WP3

Task: T3.2

Nature: R – Report

Dissemination: PU – Public

Status: Living

Delivery date: 31.03.2016

Version and controls:

Version	Date	Reason for change	Editor
0.1	10.02.2016	Initial draft version	Matthew Broadbent
1.0	31.03.2016	Final version	Matthew Broadbent

Acknowledgement: The research leading to these results has received funding from the European Union's Horizon 2020 Programme (H2020-ICT-2015, call ICT-19-2015) under grant agreement n° 687921.

Disclaimer: This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document may contain material, which is the copyright of certain MPAT consortium parties, and may not be reproduced or copied without permission. All MPAT consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the MPAT consortium as a whole, nor a certain party of the MPAT consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

Executive Summary

In this deliverable, we describe the initial usability testing procedure for the MPAT project. This includes the identification and description of two unique user groups: content creators and content consumers.

Noting their varied and diverse requirements, we explain how different usability characteristics are important to each group, as well as exploring common goals between the two.

Following this, we outline the details that must be considered for conducting usability testing. In particular, we focus on the use of focus groups, including the pre-requisites and dependencies that must be met before testing can take place.

We then provide a detailed procedure for conducting such testing, as well as identifying ethical considerations and the roles that each partner will take in the initial usability testing phase.

Table of Contents

1	Introduction.....	5
2	Test Plan	6
2.1	Areas of Measurable Usability.....	6
2.1.1	Creator Usability Goals.....	6
2.1.2	Consumer Usability Goals	7
2.2	WP Differences.....	8
2.3	Identification of Participants.....	8
2.3.1	Representing Creators	9
2.3.2	Representing Consumers.....	9
2.4	Details.....	9
2.5	Pre-requisites.....	12
2.6	Pilots	12
2.7	Procedure	13
2.8	Ethical Considerations.....	13
2.9	Partner Roles.....	14
3	Glossary	15
3.1	Partner Short Names.....	15
4	References.....	16

1 Introduction

The core aim of the MPAT project is to *design, build and evaluate* a tool for the creation of platform-agnostic media presentations. This is an iterative process which aims to create the final tool through a series of repeated development cycles. Part of this continuous process involves a study of usability from each of the prospective user groups: *content creators*; who produce and create applications using the MPAT platform, and *content consumers*; who will use these applications to interact and consume content via the applications produced previously.

This document contains a plan for testing the usability in the first phase of development, which will solicit initial feedback from the aforementioned user groups. For the content creator, this will be the design of the application authoring interface. In the case of the content consumer, this will be example applications that demonstrate the various functional features.

The remainder of this document identifies the relevant areas of usability, as well as how these can be measured and used in the project moving forward. Following this, we discuss the identification of participants for the usability study, and outline the considerations for the proposed testing.

2 Test Plan

2.1 Areas of Measurable Usability

The goal of this initial usability study is to discover potential deficiencies and highlight necessary improvements in both the *consumer*-focused and *creator*-focused User Interface (UI) designs. This evaluation forms part of the iterative process of development employed by the MPAT project. The findings collected through this process will be presented in D3.2. As such, they will provide an input for the next version of the UI design, presented in D3.3. The usability testing is repeated later in the project to complete the iterative cycle (as D3.4, D3.5 and D3.6, respectively).

Usability can be measured from a number of perspectives. These include learnability, efficiency, errors, memorability and satisfaction, and are described in more detail below:

1. **Learnability:** The capacity to understand and comprehend a system, especially without prior experience of using it.
2. **Efficiency:** Once the system is learned, the time and effort required to achieve a desired task.
3. **Memorability:** As users may not use the system regularly, the capability of a user to remember and subsequently use the system after such a period away.
4. **Errors:** The rate by which users make erroneous actions whilst conducting a task. Also considers their ability to recover from such situations.
5. **Satisfaction:** Whether the user has found that the system meets their subjective needs.

As MPAT considers two distinct groups, the areas of measurable usability will be different between them. In the following sections, we consider each group in the context of the above mentioned perspectives.

2.1.1 Creator Usability Goals

One group considered within the scope of MPAT is the content creators; that is, those that will be authoring the applications used to display and present content to the end user. This may include traditional television broadcasters, new-media 'creatives' and those with little or no experience in the area of content presentation.

As such, there is two distinct sub-groups within this group. The first subgroup is the *content editors*, who need an intuitive and simple user interface to be able to author applications with minimal understanding and in a short period of time. The second group is the *technical staff*, who is responsible for roles and right management, as well as the implementation of additional themes and integration of new plug-ins. They will mainly support the *content editors*, but may include building the applications themselves.

For this group of users and the *content editors* in particular, *learnability* refers to the ability to understand and fathom the application authoring interface, particularly when first using it. As this interface is designed to be intuitive, this metric is a core concept for assessing the usability of MPAT.

WordPress^[1] is a candidate technology for use within MPAT. It also provides a basis from which to begin analysing *learnability*. As WordPress is intentionally designed to enable users to create websites easily, especially without previous experience of doing so, MPAT should aim to emulate this behaviour, or that of other such platforms.

Efficiency is also a key facet for the content creators, as the speed at which they can complete tasks is important in a production scenario. MPAT should enable the rapid creation of applications in response to spontaneous events. As well as considering the needs of novice users, MPAT will, to a lesser extent, facilitate those who have previous experience in creating interfaces for content presentation.

Memorability is another perspective that must be considered within MPAT, as it caters specifically for casual users. These are users who may occasionally use the application authoring interface, but have otherwise used the interface previously.

The rate of *error* is intrinsically linked to the other measures in that it can impact all of them in some way. The amount of these occurring should be minimised in all cases, and for all levels of user experience. Particularly for novice users, it must be clear how an error can be rectified, important when a content creator is creating a presentation intended for widespread public consumption.

The final usability measure that must be considered is the *satisfaction* achieved through the use of the application authoring interface. This is a subjective measure, and is directly related to the user's own personal opinion. In many respects, this is where the most interesting and valuable information comes from, especially when the feedback from this testing is to be used in future iterations of the UI design within MPAT. Without the acceptance of the content creators, MPAT will likely not see extensive usage.

2.1.2 Consumer Usability Goals

The same set of metrics is applicable to the consumer of the applications, albeit in different ways and with different priorities.

For example, *learnability* in this case refers to the ability for a user to become familiar with the application. Given that applications on televisions are often short-lived (in that they are only used for a special event or a one-off interaction), this metric is crucial to the success of an application built using the MPAT toolkit. It is now in fact expected that applications are intuitive for consumers to navigate and use; failure to make them so will ultimately result in low user retention and a lack of engagement.

Efficiency is less important in the context of MPAT applications, particularly if they are short-lived (as discussed above). However, there are some scenarios where an application built using MPAT may be more permanent, such as a channel launcher bar. In these cases, the efficiency becomes more important, especially if it is a feature that users will interact with regularly.

A similar situation occurs for the *memorability* metric. If an application is short-term, then there is a minimal chance that the user will spend considerable time away from the application and then return, given the availability of such. However, as with *efficiency*, there are likely to be more permanent applications where *memorability* becomes more prominent, such as the abovementioned launcher bar.

As with the content creator *error* metric, this should almost always be minimised. However, in the case of an MPAT application intended for the end-user, this can have an additional impact: if the user becomes irritated or unable to perform their required actions, they will likely be dissuaded from using the platform. This may only be a temporary measure, or may result in a prolonged or even permanent failure to return to an application of the same sort. As interactivity is a tool designed to increase user engagement and retention, an error-prone application can actually have the opposite effect, even if it is unintentional.

In the case of the content consumer, the *satisfaction* can be even harder to measure, as each application will be developed for a specific audience and target market, and built by a different content creator. As such, the methods by which to measure satisfaction can vary greatly between each specific group. However, there will be a generic set of features and design elements which are common between the groups, regardless of the demographic. As such, this set of core features are those that must be used to determine *satisfaction*.

Since MPAT can be applied to an infinite amount of potential scenarios, and the application created through such can take on any of a number of forms, it is difficult to cover all potential use cases. To address this, we will identify core or representative user groups. Using these will ensure that what is created is acceptable to at least a number of important demographics.

MPAT will also provide, through documentation and templates, a set of 'golden-rules' and style guides. Through the use of these in usability testing, it will be possible to assess the suitability of these as examples from which others can base their designs. If a content creator follows these established guidelines, then the application can be said to hold at least some of the same usability attributes.

2.2 WP Differences

In WP3, we focus specifically on the usability of MPAT and the applications it will be used to create. However, evaluations of other sorts occur in a number of other packages. For example, in WP4, T4.1 will assess the functional and non-functional requirements of the MPAT platforms. Clearly, some of these requirements will encompass aspects of usability. Yet the focus will largely be on a broader analysis, including the extensibility of the platform and business-orientated analysis.

WP6, and particularly tasks T6.2 and T6.3, also perform evaluation. As with WP4, this evaluation will include a number of aspects which fall under usability category. However, the main difference in this WP is that it will focus on large-scale pilots in real-world production environments, whereas this WP (WP3), focuses on smaller-scale evaluation (through the use of focus-groups and one-to-one evaluations).

2.3 Identification of Participants

The identification of representative participants is key to the success of usability testing. If the testing fails to address the required participants adequately, results can be less useful. In some cases, the incorrect identification of participants can misinform the overall design and finished product, and be an otherwise counterproductive effort.

To claim statistical significance within the usability testing, it is necessary that group sizes are considered. This is true for both of the two distinct groups we have identified. However, there are many factors which can influence the amount used in testing, including the availability and willingness of participants. As such, it is left to each individual partner to determine an applicable amount.

In this phase of testing, it is likely that small groups of 4-5 participants will provide the most valuable feedback. In fact, larger groups may cause unnecessary overhead, especially given the relatively short time period the evaluation has to occur in. Larger groups may also produce results that are difficult to decipher and understand, particularly at this early stage where focus is required.

The usability testing described in this plan will take place at a number of different locations. This allows for a greater amount of coverage and diversity, particularly in capturing different user groups, experience levels and demographics. This includes the consideration of users who are otherwise underrepresented or unfamiliar with such platforms.

As mentioned previously, there are two distinct groups within the scope of the MPAT project. In the following sections, we describe how each of these can be best represented.

2.3.1 Representing Creators

As the MPAT project consists of multiple participating institutions, of which some specialise in broadcast television, these will likely be used to provide the main focus for representing content creators. In addition to these typical users, the MPAT project also seeks to engage with others who would not traditionally consider building such applications for platforms such as television.

Considering this, this group must incorporate users with varying degrees of experience. This includes familiarity with content creation and presentation, web presentation and or existing tools. Paramount amongst these are exposure to WordPress and HbbTV^[2]. As MPAT is hoping to emulate the ease of use offered by WordPress, having used the platform previously will aid greatly in the discussion and feedback in this early phase.

Similarly, some of the creators approached for this usability testing may also have developed for one or more of the platforms targeted in MPAT. This includes exposure to HbbTV. In these cases, this must be documented and noted appropriately, as it will undoubtedly influence the outcome of the usability testing. Even within an organisation, there may also be users who are competent in some areas, but are considered relative novices in others. Determining the experience of each participant is an important process that is usually performed shortly after the usability testing (as not to influence or guide the result by asking participants beforehand).

2.3.2 Representing Consumers

At a very general level, potential consumers can be considered to be anyone who may use an application created through the MPAT toolkit. As such, this includes anyone with access to a platform supported under MPAT, such as a HbbTV-compatible television. Evidently, such a large group is too broad and diverse to draw any meaningful conclusions, and as such, more targeted and specific groups should be considered.

Potential groups that can be approached as part of this usability testing include those who have never used an application of the sort created through MPAT. Other potential groups may include viewers of a particular show or series, whom would be likely to use the application if deployed alongside the content. Similarly, there is also another potential group of users who have used similar applications in the past, such as 'Verknallt & Abgedreht'^[3] or 'Grenzenlos – 25h Mauerfall'^[4], these form an interesting baseline on which MPAT applications should seek to improve.

2.4 Details

In the following section, we outline a number of details that need to be considered before usability testing can take place. For testing to be successful, these must be addressed in advance of testing. Adhering to these will yield valuable information that can be fed back into the project iterative development cycle.

For the purposes of this initial usability testing, we will utilise a series of *focus groups*^[5]. These will attempt to match the group sizes determined above, and have a number of advantages, especially at this early stage of the project. These too are evidenced in the following section.

Where?

Given that a focus group involved the participation of numerous people, then the location chosen to conduct the test must be capable of hosting such an amount. Seating should be provided for all of these participants, given the length of time that the testing is likely to take. This ensures their comfort throughout the process.

Similarly, it should be ensured that the testing is not interrupted in any way once it has commenced. This might include preventing intrusion from others members of staff that share the build. These inconveniences may disturb or influence the process in some way, intentional or otherwise. Similarly, refreshments should be provided, consummate to the length and time of day in which the testing is taking place. Provision for comfort breaks is also required, and should be incorporated into the testing schedule if necessary.

The venue should also have in place the tools and equipment used to facilitate discussion. This may include whiteboards, pens, sticky notes and anything else deemed necessary. If these are not on hand, spontaneous discussions and demonstrations can be curtailed unnecessarily. Clearly this should be avoided if possible. Having these resources on-hand also caters for participants with different ways of visualising and presenting ideas; not all participants will do this in the same way.

How long?

Focus groups are typically designed to be less than two hours. The exact timing of the testing is however relatively fluid, with no prescribed minimum or maximum. It is important that discussion is facilitated and allowed; ending early or continuing beyond the natural conclusion can be counterproductive.

This is one of the major advantages of using focus groups: they allow a large amount of valuable data and feedback to be collected in a relatively small time frame. They also allow for a progressive approach to be taken, with later (Phase 2) usability testing taking on a more intensive and exhaustive role. This later testing (to be outlined in D3.4) will likely take the form of one-on-one testing, which is more appropriate given that prototype implementation will be available by this point in the project.

Who?

In the previous section (*Identification of Participants*), we outlined the size and composition of testing groups. This process includes working within reasonable bounds; the participants have to want to be present at the testing as their unwillingness may bias the overall result. To address this, it is important that participants have some incentive to be present at the focus group. This may include job requirements, an interest in using the finished MPAT, or other financial/material benefits offered.

If there is not a ready group of participants at a partner's institution, then it may be necessary to advertise and publicise the focus group in order to garner attention. This can be done in many ways, each of which will be unique to the partner in question. Nonetheless, this can be a resource intensive process in itself, and must be considered way ahead of the actual testing.

What?

The data and information collected as part of this testing will provide informative feedback for the MPAT development process. However, it is important to note that a focus group in itself can only collect subjective data. For more objective data, different testing techniques are required; as mentioned previously, these will likely be used later in the project, as they suit a more interactive demonstration and/or prototype usage. It would be difficult to conduct this type of testing at this stage, as these implementations are yet to be developed; they should be available for Phase 2 of the project. Nonetheless, a small amount of quantitative data may be collected, even at this early stage.

During the usability test itself, it is vital that any and all discussions are audibly recorded and stored in some fashion. They can then be used later in the process: initially as input in the reporting that is part of this WP, and also referred back to later (and in other WPs) if necessary. To supplement this information, other recording techniques can be used, such as video. It may also be possible to photograph and collect the material used to aid discussion, such as drawings made on a whiteboard or sticky notes that have been used in the process. All of this should be collected and stored appropriately for use later.

Supplementary to the focus group itself, it may also be possible to conduct either a short survey or an interview with each participant individually. This would usually occur after the usability testing is complete, so as not to influence the result of this earlier process. This can be used to ascertain information such as previous experience (if this has not become evident through the focus group itself), or to provide tentative quantitative information using a survey. This might incorporate the use of a 'Likert' scale^[6], or similar, which provides the user with a scale on which to rank a particular feature (usually from *agree strongly* to *disagree strongly*).

Once all of this information has been collected, the data can be collated into a report. This will form the basis for the feedback once it has been interpreted and the findings have been ascertained. Statistical analysis techniques can also be employed at this time, although these will be limited in that the amount of quantitative data collected will be minimal and preliminary.

These individual reports (produced by each participating partner), will then be combined into a comprehensive *Usability Findings Report* (D3.2). In addition to containing the outcomes of each usability test, this report will also outline any overall findings, especially commonalities discovered between testing at different locations.

For this comparison to be successful, it is important that at least some harmonisation occurs between testing procedures and the resultant outputs; any differences in procedure should be clearly documented and noted. This ensures that results are comparable and that findings can be somewhat generalised. It also guarantees that the findings and considerations going forward are not only appropriate, but also representative of the target audiences that MPAT is hoping to address.

Why?

The purpose of the usability testing is to ascertain feedback relating to the early design of the MPAT interface and the applications it will be used to create. This feedback will be used to create a set of refinement and changes that must be considered in both the later designs, and the implementations and prototypes that result from these.

At this stage in the process, it would not be beneficial to consider definitive success criteria. However, early indications of general acceptance can be derived from the feedback recorded during the focus groups. To supplement this preliminary data may also be collected from the short surveys (if used). Although these results will not be conclusive, they will provide useful feedback early on in the project.

It is likely that a number of items will come to light through this process; to prioritise each of these, a score will be given to them. These will be agreed on between all partners to ensure consensus. These scores will indicate the severity of the defect, and highlight those that are of particular concern (and may be posing a risk to general usability of the system, rather than providing a minor refinement).

2.5 Pre-requisites

There are a number of prerequisites that must be in place before testing can occur. In addition to this test plan, it is necessary for the following actions to be undertaken:

- Delivery of the initial set of UI designs from D3.3: these will form an important part of the usability testing, as this will act as a focal point in the usability testing. The internal version (to be used in the focus groups) is to be delivered in M6.
- Derivation of a set of common questions to be used in these focus groups: these are to be agreed upon between all partners, but may be modified accordingly. These will be confirmed shortly and included in full within the *Usability Findings Report* (D3.2). Questions include:
 - 1-2 introductory questions
 - 5-6 additional questions
- Prioritisation and description of key scenarios from WP2: these will be used to refine questions and highlight usability aspects that must be discussed in the focus groups.
- Development of a script: to be used in the introduction and welcome of participants. It will primarily describe the testing procedure, outline the ethical considerations and introduce the project and scope of the work. Following the test, a second script may be used to thank the participants and conclude the session.
- Selection of a facilitator for the testing: this person acts as a moderator to facilitate discussion, and to stop it if necessary. This includes keeping the discussion within reasonable bounds, and preventing participants from hijacking or monopolizing the discussion if necessary. This person should be completely neutral if possible, but it is more likely that a participating member of staff acts as facilitator. In this case, the preparation is even more important, as they must consciously consider how their actions may impact the neutrality of the test.

2.6 Pilots

Piloting the usability tests is a necessary part of the testing procedure. In essence, this process involves running a mock usability test with a small set of participants (maybe 1-2). This will highlight any

immediate issues or significant deficiencies. These include the suitability of the room to be used, the availability of resources, and/or preparation on the organisers' behalf. These participants do not need to be representative of actual users (although this definitely aids the process). It is important to note that participants used in the pilot cannot participate in the main testing, at risk of influencing the outcomes.

2.7 Procedure

The specific procedure to be executed on the day of the testing is as follows:

1. The facilitator should arrive before the participants and prepare for the arrival of participants by setting up the room, ensuring that everything is in place for the testing. This may include any necessary signage, refreshments, documentation, stationery etc.
2. The facilitator (or another member of staff) should welcome participants to the building / department / company as appropriate. Evidently, this may not be necessary in all cases, particularly if the participants are already familiar with the surroundings.
3. Once all the participants have arrived at the venue, introduce them to the usability test using the previously prepared script. This should explain the details of the focus group, venue specific details, schedule etc.
4. Ensure that recording provision is in place and started.
5. Have each participant introduce themselves; this may include providing their name, a brief description of themselves and their role in the organisation.
6. Begin the testing with the introductory questions.
7. Move on to the main questions, exploring each one in turn: facilitating any discussions by providing direction. This should be done with minimal input and/or influence. It may also be necessary to provide resources to aid this discussion as required.
8. After the test is complete and all questions are exhausted, provide a summary of the session to participants and thank them for their participation.
9. Conduct any post-testing procedures, such as distributing surveys or conducting short interviews.
10. Review the session and begin to record outcomes. This is best done shortly after to ensure that everything discussed within the session is recorded accurately; this is particularly important if something won't be evident from the recording and may be lost otherwise.
11. Once these findings are collated, analysis can begin: looking for trends and general findings moving forward. It may also highlight specific recurring issues that need to be addressed in the MPAT development process.
12. Finally, these findings will form the basis for D3.2. Each report will need to describe the specific experiment (which although similar, may be slightly different between partners), the experimental setup and any findings (small or otherwise).

2.8 Ethical Considerations

The testing procedure described above involves the collection of potentially personally identifiable data. In such cases, it is necessary to anonymise any such data to ensure privacy of users is maintained. Users must therefore be informed as to what will be collected and how it is intended to be used.

The minimum procedure that must be adhered to involves receiving consent from each participant, usually in the form of a signed letter or form, which allows the collected data to be used in the development of MPAT. It needs to be made clear in the provided information that the data collected will be anonymous, and as a result, anything that they say or contribute to the discussion cannot be personally linked back to themselves. It should also be made clear to users that they are free to leave the focus group at any time, and to do so without the need to provide an explanation.

These ethical considerations must be enforced and strictly adhered to if the data ascertained from these tests is to be shared between the partners and/or made publically available. It is important to note that the abovementioned considerations only form the *minimum requirements*; each jurisdiction and country will have its own specific requirements that must be met when handling user data. Complying with these requirements may take significant effort and time; it is important this process is started immediately.

2.9 Partner Roles

MPAT consists of a number of partners, each with their own unique circumstances and capabilities. Leveraging the diversity of these partners, the following roles and groups have been identified (although these are likely to change):

- **Lancaster (content creator, content consumer)**
 - Students (consumer)
 - Lancaster Arts patrons (consumer)
 - Lancaster Arts staff (creator)
- **RBB (content creator)**
 - Content editors
 - Technical staff
- **Mediaset (content creator)**
 - Content editors
 - Technical staff
- **Leadin (content consumer)**

3 Glossary

3.1 Partner Short Names

Short Name	Name
FRAUNHOFER	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (DE)
IRT	Institut für Rundfunktechnik GmbH (DE)
RBB	Rundfunk Berlin-Brandenburg (DE)
ULANC	Lancaster University (UK)
MEDIASET	Reti Televisive Italiane S.p.A. (IT)
LEADIN	Leadin Oy (FI)
FINCONS	Fincons SpA (IT)
IMT	Institut Mines-Telecom (FR)

4 References

- [1] *WordPress* – <https://wordpress.org/>
- [2] *Hybrid Broadcast Broadband TV* – <https://www.hbbtv.org/>
- [3] *On the air again: „verknallt & abgedreht“* – <https://www.fokus.fraunhofer.de/60e05a5e365eadb8>
- [4] *rbb: Social TV App* – https://www.fokus.fraunhofer.de/go/rbb_socialtvapp
- [5] *Likert Scaling* – <http://www.socialresearchmethods.net/kb/scallik.php>
- [6] *Conducting A Focus Group* – <http://www.cse.lehigh.edu/~glennb/mm/FocusGroups.htm>