Traction
Opera co-creation for a social transformation

Initial evaluation methodology definition

DISSEMINATION LEVEL	PUBLIC (PU)
WORK PACKAGE	WP4 COMMUNITY-CENTRIC METHODOLOGY AND VALIDATION
DELIVERABLE NUMBER	D4.1
VERSION	V1.0
SUBMISSION DATE	31/07/2020
DUE DATE	31/08/2020

This Project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870610

Authors

Authors in alphabetical order		
Name	**Organisation**	**Email**
Aline Striner	CWI	alina.striner@cwi.nl
Anna Matamala	UAB	Anna.matamala@uab.cat
François Matarasso	FM	matarasso@me.com
Jie Li	CWI	Jie.Li@cwi.nl
Olga Soler Vilageliu	UAB	Olga.soler@uab.cat
Pilar Orero	UAB	Pilar.orero@uab.cat

Control sheet

Version history			
Version	**Date**	**Modified by**	**Summary of changes**
0.1	14.05.2020	Anna Matamala	Initial draft
0.2	15.06.2020	Anna Matamala	Bibliography updated, evaluation sheet added, roadmaps proposed.
0.3	11.07.2020	Anna Matamala	Version for review, which includes input from partners
1.0	21.07.2020	Anna Matamala	Final version

Peer review		
	Reviewer name	**Date**
Reviewer 1	Stefano Masneri	14/07/2020
Reviewer 2	Gabriel-Miro Muntean	15/07/2020
Reviewer 3	Esther Novo	31/07/2020

Legal disclaimer

Table of content

List of figures

List of tables

Abbreviations

Abbreviation	Definition
AR	Augmented Reality
BVP	Blood Volume Pulse
D	Deliverable
DoA	Description of the Action
ECG	Electrocardiography
EDA	Electrodermal Activity
EKG	Electrocardiogram
EMG	Electromyography
QoE	Quality of Experience
T	Task
UX	User Experience
VR	Virtual Reality
WP	Work Package

Executive summary

This deliverable describes the initial overall evaluation methodology of Traction and is addressed, first of all, to the consortium, with an emphasis on trial leaders and evaluation coordinators. It is also addressed to the interested public, as it allows them to get a better understanding of how Traction evaluation activities are designed.

Section 1 describes the purpose of the deliverable, the general Traction approach to methodology, and the intended audience of the deliverable.

Section 2 provides a summary of the global methodological approach and puts emphasis on the role of community and on participatory approaches to research.

Section 3 describes the different evaluation axes, which correspond to a) tools and technology, b) co-creation and training processes, c) opera full performance and capsules, and d) social transformation impact.

Section 4 indicates where the evaluation actions will take place, describing the three Traction trials.

Section 5 provides a first categorization of the participants, including audiences, which will be involved in Traction evaluation activities.

Section 6 discusses ethical aspects and responsibilities in the evaluation process and provides a broad overview of possible methods to be implemented. This is complemented with specific details on planning and reporting evaluation activities, under Section 7.

Section 8 provides a timeframe for the evaluation activities, which is then expanded in Section 9 (evaluation activities before the performance), Section 10 (during the performance) and Section 11 (after the performance).

Section 12 provides a roadmap of the evaluation activities that have been planned in the initial stages of the project.

The annexes contain the proposed protocols for different evaluation methods.

1 Introduction

1.1 Purpose of the deliverable

This deliverable describes the overall evaluation methodology of Traction, and is a direct outcome of Task 4.1, *Overall Evaluation Methodology*, which is responsible for the definition of the evaluation methodology of the three Traction trials led by: Irish National Opera (INO), Gran Teatre del Liceu (LICEU), and Sociedade Artística Musical dos Pousos (SAMP). Task 4.1 defines a global evaluation methodology shared by all Traction trials in respect of four evaluation axes (tools and technology, co-creation and training, opera performance and capsules, and social transformation impact). The task also aims to provide a set of best-practice examples and recommendations to replicate the trials, or to develop new ones.

1.2 Traction evaluation approach

Overall, the Traction evaluation approach builds on the 4 core axes of WP4:

- Tools and technology
- Co-creation and training
- Opera performance and capsules
- Social transformation impact

Each of the core axes will work independently under this general framework.

Our evaluation approach is iterative, adaptable and flexible. The evaluation is **iterative** because different evaluation rounds will take place and the results may inform subsequent rounds. The evaluation is **open** and **adaptable** because, although a general framework is proposed in this document, the specificities of each trial will be taken into account during the course of the project in close coordination with WP3. This means that not all evaluation components and actions may need to be replicated in all Traction trials. Further, the evaluation is **flexible** because this initial proposal is open to being revised as the project evolves.

1.3 Intended audience

This deliverable is addressed, first of all, to the consortium, with an emphasis on trial leaders and evaluation coordinators. It is also addressed to the interested public, such as opera organisations that might be interested in approaching community operas, as it allows them to get a better understanding of how Traction evaluation activities are designed.

2 Traction methodological approach

This section reports on the main concepts that sustain Traction's methodological approach to evaluation activities.

2.1 From the human to the community

Human-centred design methodologies will be at the heart of the development of the Traction toolset. As defined by ISO 9241-210:2010(E), "Human-centred design is an approach to interactive systems development that aims to makes systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, usability knowledge, and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of the use on human health, safety and performance". Through **an iterative process**, the evaluation activities concerning technology and its use in community opera co-creation will be repeated during the project with the purpose of continuous improvement. Any evaluation action is expected to be preceded by a review of relevant literature to identify pre-existing work and research gaps in a list of areas in which Traction is interested in.

Evaluation of the social transformation impact of Traction will be informed by the capabilities approach to human development proposed by Martha Nussbaum in *Creating Capabilities* (2011). In that context, we will be concerned to understand whether and how participation in community opera creation can extend specific capabilities of the people who become involved, especially in the second axis, on co-creation and training. The resulting data will be used to understand whether and how individual change translates into social change in the communities to which the participants belong. This work will be done in close cooperation with the community members involved in the three community opera trials.

Participatory approaches to evaluation

The approach taken in the evaluation of Traction processes and outputs is participatory, meaning that it will not be an external task carried out by partners in WP4 on WP2/WP3 outputs and activities, but will instead be embedded in the trial activities and based on different "participants" that will be involved. To guarantee its feasibility, and a tight coordination between trial leaders and evaluation partners, evaluation coordinators have been selected for each trial. These coordinators –Patricia Malpas for INO, Eva Garcia for LICEU, Sandra Sousa for SAMP– will be in charge of keeping the records and coordinating with WP4 partners in evaluation activities. They will be provided adequate training and guidance where relevant so self-evaluation can take place successfully. The evaluation will also be participatory in the sense of seeking to work *with* the people involved in the trials at every stage. This is both out of respect for human dignity –no one should be considered or treated merely as a subject of the research process– and because the quality of the data gathered will be enhanced if people understand and value the reasons for which they are

being asked to provide. We will retain the possibility of involving participants and community members as co-researchers e.g. asking young people to interview each other, or people to record the views of family members. The value and ethics of such research methods will be assessed if and when they are felt to be necessary.

2.2 Starting point: aims and objectives

The aims of the three trials as reported in D.1.4 (final draft 25 March 2020) are phrased as follows:

Table 1: Aims of the trials

TRIAL	AIM
INO	To make contemporary opera a route for social cohesion in Ireland.
LICEU	To increase participation by persons with disabilities and migrants throughout the LICEU's work
SAMP	To make opera a viable route to rehabilitation of offenders

These aims, which may need to be refined as the project evolves, are broken down into more specific objectives, linked to the different perspectives taken in Traction (Table 2).

Table 2: Objectives of the trials

TRIAL	OBJECTIVES
INO	1. Providing outreach activities that engage several communities in exploring opera's potential in expressing their own cultures and lives. 2. Facilitating an opera co-creation project with communities and professional artists. 3. Testing how digital technology can enrich opera creation and production processes. 4. Producing and presenting a VR community opera in different parts of Ireland. 5. Building understanding between different communities through the opera. 6. Sharing the learning from the Traction experience with other opera companies and arts organisations.
LICEU	1. Working with community groups in Raval to build trust and common purpose. 2. Using technology to facilitate their creative inclusion in access, design and marketing requirements of Opera Prima Raval. 3. Working with them to identify how technology can enhance the audience experience. 4. Co-creating digital capsules with community groups to extend the life and reach of the community opera. 5. Research the process to identify learning of wider value to the opera sector.
SAMP	1. Building the skills and enthusiasm of inmates and staff at Leiria Prison through multidimensional opera workshops. 2. Developing a professional music and theatre education course for use in prisons.

	3. Testing ways of using digital technology to involve families and communities in the creative process. 4. Co-creating and performing small opera elements in and outside the prison. 5. Producing a new large-scale opera involving the prisoners and the community. 6. Promoting understanding of and support for the work in wider society.

The previous objectives are linked to the evaluation tasks in WP4 (see Section 3):

- Evaluation of Traction toolset (T4.2): INO3, LICEU2, LICEU3, SAMP3.
- Evaluation of co-creation and user training (T4.3): INO1, INO2, LICEU4, SAMP1, SAMP2, SAMP4.
- Evaluation of opera full performances and capsules (T4.4): INO4, LICEU4, SAMP5.
- Evaluation of social transformation impact (T4.5): INO5, INO6, LICEU1, LICEU 5, SAMP6.

The global aim of Traction and related objectives are summarised in Table 3. Evaluation has a central role, and is linked to technology (ALL3), process (ALL1), specific output (ALL4) and social transformation impact (ALL-AIM).

Table 3: Traction's global aim and objectives

AIM
To strengthen the potential of opera as a path for social and cultural inclusion
OBJECTIVES
1. Researching community opera practice in Europe, and the uses of new technology in co-creation, performance and audience experience. 2. Developing a set of new or adapted technological tools that can support inclusive community opera creation and production. 3. Testing the technology in three best-practice community opera projects in Spain, Portugal and Ireland. 4. Evaluating the outputs and outcomes of the trials to identify transferable lessons and improvements in the technology. 5. Publishing the findings and toolset in open access formats to support the future work of opera producers working towards social inclusion.

2.3 Research questions

Traction is a *research* project. As such it needs to state clearly and consistently what it intends to discover. That is most easily done by defining a research question its evaluation methodology is intended to answer, at least partially. The title of the project, in the grant agreement is 'Opera co-creation for a social transformation' and it seems sensible to root the research question there.

However, it is not yet a question, and it contains terms that are vague ('co-creation') and too large ('social transformation'). The first of these can be clarified and defined through the research process itself, but the second needs to be narrowed down. The grant

agreement speaks of 'rising inequality and social exclusion', and so inclusion is the aspect of social transformation we will focus on. It goes on to state that:

> *We do not want to make opera palatable to those who don't attend. We want to* ***define new forms of artistic creation through which the most marginalised groups*** *(migrants, the rural poor, young offenders and others) can work with artists to tell the stories that matter now.* ***By combining best practice in participatory art with digital technology's innovations of language, form and process, we will define new approaches to co-creation and innovate in three fields****: a) Opera creation and production; b) Immersive and interactive digital media; and c) Social integration and community development.*

From this we can draw the following research question for the Traction project:

- **How can participation in opera co-creation support social inclusion?**

The grant agreement also makes specific commitment to innovate in three fields: creation and production; digital media; and social integration. These have been reviewed and organised into four areas of research and evaluation, each with a specific question.

Technology

Immersive and interactive digital technology is at the heart of the Traction project, opening creative and experiential possibilities that are rarely, if ever, available in community opera co-creation. That may involve innovative technologies or innovative uses of existing technologies, according the particular needs of each trial.

In terms of technology Traction aims to evaluate:

- **How can digital technology contribute to opera co-creation?**

Co-creation and training

Traction is concerned with artistic co-creation (also termed participatory art), defined as 'the creation of a work of art by professional and non-professional artists' [1]. This element of the evaluation is focused principally on positive and negative outcomes pf participation by individuals in the community opera process, and what that indicates with regard to best practice in artistic co-creation.

In terms of co-creation and training Traction aims to evaluate:

- **How have participants benefited from the opera co-creation process?**

Opera performance and capsules

A work of art is completed only when it is presented to an audience, so the three community trials are all working towards one or other kind of public performance. Again, there is a role for digital technology either in enhancing the performance itself or in reaching audiences that are excluded for reasons of possibilities that are rarely, if ever, available in community opera co-creation. This will involve research into audience profile and responses to different aspects of the offer during and after the performance.

In terms of opera performance Traction aims to evaluate:

- **How can digital technology enhance people's access to and experience of opera?**

Social transformation impact

The aim of the project is to contribute to a social transformation through the increased social inclusion of the people who participate. The project will consider whether inclusion can be tracked by monitoring changes in a participant's fundamental 'capabilities' using the theories of Martha Nussbaum. It will also use community development theory to test whether individual change translate into broader social change.

In terms of social transformation Traction aims to evaluate:

- **How does participation in community opera co-creation build individual capabilities and lead to social outcomes?**

These research questions concern the Traction project and are used as a framework for the evaluation, where more specific aims and objectives are established for each of the axis and phases.

2.4 Planning: what, when, where, by whom, and how?

When planning the evaluation methodology, we employ a simple approach, focusing on 5 key questions:

- WHAT needs to be evaluated? (See Section 3)
- WHERE will the evaluation actions take place? (See Section 4)
- BY WHOM will the Traction development be evaluated? (See Section 5)
- HOW will the evaluation actions be performed, meaning what methods will be used and WHO will be responsible for (a) designing the evaluation protocols, (b) conducting the evaluation actions, (c) reporting on them, and (d) analysing them. (See Sections 7 and 8).
- WHEN will the evaluation actions take place? (See Section 8)

3 What? Evaluation axes

As mentioned, the evaluation of Traction activities is built upon four axes which will be followed independently under the umbrella of this overall methodological framework: tools and technology (3.1), co-creation and training (3.2), opera and capsules (3.3), and social transformation impact (3.4).

3.1 Tools and technology

As defined in the DoA, "the toolset will be evaluated to assess whether the technology being deployed is effective and meets the initial objectives defined (from a creative perspective) but also from a usability and accessibility perspective." More specifically, Task 4.2., *Traction toolset technical and user experience (UX) evaluation*, will "carry out experiments, mostly subjective ones, for the UX evaluation of the prototypes developed in WP2. The user experience testing will focus on two main aspects, considering on the one hand the participants that aim at designing/co-creating community opera experiences using the toolset, and on the other hand, the audience that will enjoy the interactive opera representations. The UX evaluation will provide a valuable feedback to WP2 for future iterations, as well as to WP5 in order to influence the exploitation strategies".

Although technology is expected to be present in all trials, the role of technology and the specific tools may differ: in SAMP, technology will likely be a tool to connect, in LICEU the role of technology will be to enhance the aesthetic experience through immersion, and in INO technology may allow to generate new opera genres and formats.

In order to get a better understanding of this issue, a document from technological partners VICOM, CWI and DCU was shared at the beginning of the project (February 2020). The result is summarised in Table 4.

Table 4: Technologies in Traction

Technology	Project expected developments	Who will test the technology?
FlexControl	The basic technology will be extended to be used during theatre performances for connecting different environments (theatre and families) and for enriching the performance (e.g. overlays). This will be used primarily in the SAMP trial.	1. Director or technical person in charge of the performance. 2. Viewers in the theatre enjoying the performance. 3. Remote viewers.
Cloud-based solution for media management	The basic technology will be extended to be used as a video vault: a database of video assets. They will be used as an asynchronous communication medium between users with	1. People uploading videos. 2. People (different profiles) viewing videos. 3. People communicating through the videos.

	different backgrounds and profiles (so different extra functionalities to be identified are needed). This will be used primarily in LICEU.	
Immersive Media experience (probably an extended version of the ImAc Player to include further immersive formats)	The basic technology will be extended to include an immersive media player for people enjoying a VR version of the opera or of the process of making the opera. It needs to be movable. This will be used primarily in INO and LICEU.	1. Professionals capturing an immersive media version of the media (360 or volumetric). 2. Professionals creating (editing) the user experience. 3. Users enjoying the immersive experience.

This first overview was expanded and pushed forward with a slide presentation by Pablo César –WP2 leader– (March 24th, 2020) on WP2 approach (Table 5).

Table 5: WP2 approach to technology

Technology	Technology	Timing	Synchronicity	Trial	Aim
Media vault	AWS-based system	Before performance	Asynchronous	INO LICEU	Gathering ideas, recording materials, conversation based on the material
Performance engine	FlexControl	During rehearsals, during performance	Synchronous	SAMP LICEU	Stage control, delivery of performance to remote locations, conversation in sync with performance
Installation or capsule	Unity-based system	After performance	Asynchronous	INO LICEU	Authoring of an experience based on media assets, allows for immersion and interactivity, moveable.

Further work has been carried out on this first approach to technology under WP2 and is documented in D.2.1. This was a first snapshot to start the dialogue between partners.

3.2 Co-creation and training

Task 4.3, *Co-creation and users training evaluation*, concerns the evaluation of:

- The artistic and developmental process involved when professionals and non-professionals work together on the creation of a new opera; and
- The acquisition of new capabilities for people involved in the trials, with a special emphasis on the non-professionals.

This will be done with reference to the aim and objectives elaborated for each trial and will focus on the experience of individuals who participate in the trials in order to establish an understanding of the results at the collective or community level under T.3.4. A necessary first step will be to define what is understood by co-creation and its associated so-called "training", which we envisage both as structured (for instance, workshops at LICEU) and unstructured (for instance, the new capabilities participants may acquire in all the trials). This task will be further developed in future iterations of this document, as the co-creation and training process are better defined through working with the participants in each trial. It should be noted that co-creation experiences will be different across trials and will evolve as the project progresses.

3.3 Opera full performances and capsules

Task 4.4, *Opera full performances and capsules evaluation*, aims to assess the methodology of creation of the capsules (for instance, the degree of participation of different professionals), the personal feelings of the people involved as artists, and the artistic output of the trials, in this case with a special role for expert evaluation. As the opera full performance and capsules are better defined, this task will be further developed in future iterations of this document. It is expected that audience responses will have a central role in this task.

3.4 Social transformation impact

Task 4.5, *Social transformation impact*, will "define the metrics to measure the social transformation impact in each of the trials, evaluating the effectiveness of the actions and how it affected the diverse communities". The language of measurement used in the project proposal presents conceptual and methodological difficulties where the social outcomes of artistic participation is concerned. Past experience suggests that the nature and extent of the social outcomes from each trial will be only partly quantifiable, if at all. Nor are replicability criteria that may be useful in evaluating technology readily transferable to social realities. The evaluation plan for tracking the social transformation impact will therefore need to propose more appropriate and accurate indicators than the metrics originally anticipated. This analysis is closely linked to the aims and objectives that have been identified in D.1.4 in relation to the trials and the project as a whole and uses the input gathered from individuals through the different tasks in this WP.

4 Where? Evaluation scenarios

The trials will take place in different countries and environments, as described next.

4.1 Trial 1: community opera with people from Raval (LICEU)

Gran Teatre del Liceu in Barcelona is one of the oldest opera houses in the world. Founded in 1847, the LICEU committed to a long-term project, *Opera Prima*, which aims to open the LICEU to the neighbourhoods through different 3-year specific activities. The first one is Opera Prima Raval (2018-2021), in reference to the closest LICEU neighbourhood of Raval, and involves collaboration between LICEU staff, an external artistic team, and the Raval neighborhood. It is within this broader context that the LICEU Traction trial will take place. The specific LICEU Traction trial will focus on how technology can help Opera Prima to be more efficient in making opera a positive force for social inclusion, especially involving local residents and persons with disabilities. Traction-related co-creation processes will focus on accessibility, design, and branding. The trial will take place in Barcelona, and most specifically in the Raval neighborhood.

4.2 Trial 2: opera co-creation with inmates in prison (SAMP)

SAMP is a community music school in Leiria, Portugal, that runs community outreach programmes in local schools, hospitals, and prisons. At Leiria Youth Prison SAMP has produced three Mozart operas performed by inmates and professional musicians. The SAMP Traction trial aims to create a new opera as a way to support rehabilitation, with a strong focus on how technology can facilitate the co-creation process and support the involvement of the families of young offenders. Artists, inmates, prison staff and social and artistic facilitators will define a co-created opera format, in which families and friends will also be vital contributors. The performance will take place in capsules and also in final full opera representations inside the prison and in a theatre in Lisbon.

4.3 Trial 3: digital opera with Irish communities (INO)

The trial aims to create a virtual reality opera through a co-creation work with three different communities: close to Dublin, in the Irish midlands, and off the Atlantic coast. The trial will test the potential of digital technologies and will involve communities that are usually outside the scope of INO. The different communities will share their stories and experiences, and they will develop new skills through the trial. This trial will allow them to increase social cohesion across Ireland and explore opera's potential in expressing community members' own cultures and lives.

5 With whom?

Defining who the persons that will be involved in the evaluation are and how to refer to them was a first step (5.1). A second step was to agree on a demographic questionnaire to gather consistent information across evaluation activities.

5.1 Defining the persons involved

In order to define the different persons involved in the Traction evaluation activities, a document was shared among trial leaders in April 2020, in which they were requested to identify the different "users" –that was the term used–, defining users as "any person involved in the different Traction trials in any stage (local residents, prison inmates, opera directors, subtitlers, etc.)".

As a starting point, two broad categories were identified:

- Professionals
- Non-professionals

The distinction between these categories is more subtle than the usual one of whether someone is paid. It includes payment, but also social position and identity: for example, very few poets earn their living by writing poetry (T.S. Eliot was a bank manager and publisher). The boundaries may not always be rigid and we will prioritise self-definition.

The result of the input gathered from the different trials highlighted the need to:

a) agree on a terminology to refer to these categories (see 5.1.1),
b) generate a broad categorization, as this can be used as a basis when deciding who to address for each evaluation action (see 5.1.2).

Different iterations of the document were produced until a categorization was agreed.

5.1.1.1 Terminology

The Traction glossary in its first version had the following definitions (Table 6).

Table 6. Initial set of definitions

Term	Glossary definition
User	A professional who uses the technology as part of an artistic process or creation to enhance the experience of the audience (e.g. a sound engineer, a director, etc.)
Participant	A person who takes part in the co-creation process. Participants cannot be paid, so giving their time must bring other benefits – e.g. psychological (confidence), social (friendship), political (recognition), educational (skills), etc.
Professional artist	Someone who is recognized by others as an artist. They have probably had training and spend most of their time making art, which provides an important part of their earnings.

Non-professional artist	Someone who is making art but is not a professional. They are probably not seen by friends and family as an artist (though that might change as a result of the project). They can bring new ideas, new stories, a need to say something and a different kind of authority

However, during the process of defining the participants, some relevant questions were put forward:

- A non-professional –who may have different roles in the project, ranging from co-creator to audience member– may also use the technology. How should we call them?
- How do we define professionals and non-professionals?
- Is there an umbrella term that we can use?

After some discussion within WP4, it was agreed that:

- The terms “professional” and “non-professional” do not relate to expertise in a specific field but *to the relationship of the person with opera artistic creation*: if they are newcomers to the field, they will be considered non-professionals within Traction.
- The term “user” will be reserved to the context of technology. Professional users employ technology in the creation and presentation of opera. Non-professional users apply technology to access and enjoy the artistic experience created by the professional users.
- The term “artist” refers to someone who makes art, and can be divided into professional and non-professional.
- An additional category of “audience” is needed – the people who see the work that results from the co-creation of community operas by professional and non-professional artists. Being an audience member is not a passive role, as reception theory has demonstrated: audiences are active creators of meaning in artistic work.

The discussions lead to a revised version of the terms in the glossary (Table 7).

Table 7. Revised definitions

Term	Definition
Professional	Someone who is a professional in artistic creation or production and who is working on the co-creation process of the trials as such.
Non-professional	Someone who is not a professional in artistic creation or production. They may be professionals in other fields but newcomers to the field of artistic creation. For instance, a professional who is n expert in sound but who is a newcomer to opera creation would be considered a non-professional.
Professional user	Someone who uses the technology as part of an artistic process or creation to enhance the experience of the audience (e.g. a sound engineer, a director, etc.).
Non-professional user	Someone who uses the technology to access and enjoy an artistic output.
Professional artist	Someone who is recognised by others as an artist. They have probably had training and spend most of their time making art, which provides an important part of their earnings. They are hired to work on the Traction trials.
Non-professional artist	Someone who is making art but is not a professional. They are probably not seen by friends and family as an artist (though that might change as a result of the project). They can bring new ideas, new stories, a need to say something and a different kind of authority. Non-professional artists contribute solely out of choice and have no obligations except those they freely accept during the process.
Audience	Someone who accesses and enjoys the artistic output.
Participant	Someone who is involved in Traction trial activities.

5.1.1.2 A suggested categorisation of Traction participants

Based on the input provided by partners, a first iteration was used as a basis for a discussion in WP4, with many questions open to discussion. During the discussion it was agreed that the ultimate aim was to have a simple and broad categorisation to identify different participants during the life of the project. Participant numbers, roles, and involvement would vary across trials. This categorisation was then open to input from other Traction participants, and the following version was produced.

PARTICIPANTS IN TRACTION

a) PROFESSIONALS

- Artistic and creative team, including
  - Primarily musicians: for instance, conductor, composers,
  - Primarily theatre-related: for instance, artistic director, playwright, librettist, stage director, set designer, costume designer, choreographer, audiovisual director, visual artists, performers, translators, virtual reality artists, etc.
  - Persons involved in the co-creation process: facilitators, community arts experts, partners' staff.
- Technical and production team: for instance, lighting engineer, sound engineer, video producer/editors, translators, access staff, signers etc.
- Traction-related professionals, with a professional relationship with the project, such as Consortium members, funders and Advisory Board members.

b) NON-PROFESSIONALS

- Artistic and creative team: community members who actively participate in the opera project from an artistic and creative perspective.
- Technical and production team: community members who actively participate in the opera project from a technical and production perspective.

c) AUDIENCES

- Persons who enjoy the artistic output.

d) OTHERS, referring to any participant who enables the participation of non-professionals, such as psychologists, prison staff, foundations and community associations, to name a few. Their role is to accompany the process, but they are not actively involved in the co-creation.

5.2 Audiences

During the definition of evaluation activities and Traction project tasks, it was observed that a more refined categorization of audiences was needed. To this end, François Matarasso created a proposal of **community opera audiences**, which includes: live audiences, remote audiences, and Traction project audiences.

a) Live audiences

People who attend the live performances of the community opera

Live Opera audience	People who already attend opera/classical music
Live Participant audience	People who have participated but are not performing in the community opera production (e.g. in design, writing etc.)
Live Connected audience	People who are friends or relatives of those who have participated in the community opera production
Live Professional audience	People with a professional interest in opera or a connected field (e.g. technology, research, cultural management, sponsorship, government etc.)

b) Remote audiences

People who experience the community opera through digital technology (recordings, capsules, VR headsets etc.) independently of any live performance. There may be some remote audiences of the live events e.g. via Internet.

Remote Opera audience	People who already attend opera/classical music
Remote Participant audience	People who have participated in the community opera production (in any capacity, including performing)
Remote Connected audience	People who are friends or relatives of those who have participated in the community opera production
Remote Professional audience	People with a professional interest in opera or a connected field (e.g. technology, research, cultural management, sponsorship, government etc.)

c) Traction project audience

People with an interest in the knowledge produced by Traction.

Arts audience	People who work in opera and other cultural professions including policy and research
Technology audience	People who work in digital technology including policy and research
Policy audience	National and local government, foundations, European Commission etc.
General audience	Possibly people interested in the future of opera and culture.

5.3 Demographic questionnaire

To gather information about participants in a unified way, a demographic questionnaire has been designed for all project evaluation activities. After an initial discussion during the kick-off meeting, a draft version was created by UAB taking into account various factors, as described next, and subject to partner online discussion in February 2020. This was further improved during D.4.1 development in terms of phrasing.

How much information?

The approach agreed by partners was to keep the number of questions to a minimum, to secure a fast reply by participants. The initial proposal by UAB included 6 questions –on gender, location, age, ethnicity, disability, job– and one additional question to be filled in by the researcher about the relationship of the participant with the project, as initial internal face-to-face discussions identified these elements as relevant for the project development. During the online discussion, it was decided to add one question about opera, with a total of 8 questions.

What type of information?

Three principles underpin the collection of data about individuals in Traction. The first, in keeping with our ethical guidelines, is anonymity. The second is self-definition: if we ask people about themselves, they have the right to decide how they define their own identity. The third is that no one is obliged to provide any data: there is always the option to tick 'prefer not to say' or just leave a question blank.

First of all, it was agreed to ask about **gender**, and to give participants the choice to define themselves as "Female, Male, Other, I prefer not to say". It was decided to go beyond gender binarism to include other options and also the possibility not to answer this question. Under "Other" there is a wide range of options that could be included, but it was preferred not to list them to keep the possible answers short.

Concerning **location**, it was considered that asking for the postcode would be a fast way to get information on this item. For **age**, it was preferred to provide an open field.

As far as **ethnicity** is concerned, a broad and open question was preferred so that participants are free to phrase it in their own terms. Still, this question had to be deleted in the Portuguese version of the questionnaire due to specific regulations. Regarding **disability**, it was decided that the principal aim was to know how many persons with disabilities have been involved in Traction evaluation activities, rather than having details on their disability, hence the proposed question only asks if they have a disability. In the phrasing of this question, an approach consistent with the United Nations Convention on the Rights of Persons with Disabilities was prioritized, and the choice of not replying to this question was also considered.

It was considered relevant to ask about the **job or profession** of participants, principally as a proxy for social background. The question was phrased in broad terms ("What is your job/profession"?), so as to take into account all possible situations (active, unemployed, etc.) without gathering further data on this aspect.

A question that was added during the discussion process was "What is your **experience with opera**?", in which participants were given three possibilities to report how they view themselves: expert (e.g. expert audience, or director, performer or composer of operas), knowledgeable, and novice. However, it was decided not to include it explicitly in the questionnaire but to ask the researcher to record it.

Similarly, it was considered relevant to gather information on the **relationship of the participant to the project**. This was added to the questionnaire, but it was explicitly indicated that participants would not need to reply to this question, but the researcher would fill in this information.

When and how?

This questionnaire will be used in all Traction evaluation actions, in order to gather information on the participants. Ethical requirements will be followed, especially concerning the following aspects:

- Participants must be given a version that they can understand. For example, this means providing the form in the language of the participant.
- Participants must be given a version that they can access. For example, this means providing an oral version or an electronic version which can be read by a screen reader to a person with sight loss, or providing a Sign Language interpreted version to a person who is a Sign Language speaker.

The demographic questionnaire can only be administered once participants have understood the aim of the action and have signed the information sheet and consent form (see deliverable on ethics, D.1.6).

Depending on the action, the demographic questionnaire may be administered in different formats (paper/online) and languages. However, to keep good track of all the replies, a unique online form in English has been created. The evaluation action responsible will be in charge of filling it with the replies, together with an activity and participant codes.

Final demographic questionnaire

The draft created by UAB based on the previous factors was subjected to online partner discussion, and successive revisions, until the final demographic questionnaire was agreed, as follows:

1. Gender?
    - Female
    - Male
    - Other
    - I prefer not to say
2. Postcode?
3. Age?
4. In terms of ethnicity, how would you define yourself? (not applicable in the trial in Portugal)

5. Are you a person with a disability? Yes/No/I prefer not to say.
6. What is your job/profession?

Additional questions to be filled in by the researcher. Not to be asked directly:
* What is the participant's relationship to the project?
* What is the participant's experience with opera?

- Expert (e.g., expert audience, or director, performer or composer of operas)
- Knowledgeable
- Novice

6 How? Ethics, responsibilities and methods

Evaluation activities in Traction will be developed in an ethical manner by using the most appropriate methods for each specific goal. A participatory approach will be favored, meaning expertise will be shared among partners and trial members will be empowered to carry out the evaluation activities.

6.1 Ethics in evaluation activities

A dedicated webinar on ethical requirements was offered to the consortium on 23 March 2020 by Pilar Orero. Ethical requirements explained in the seminar, and further described in D.1.6 Ethical requirements, will be followed in all evaluation activities. Four important aspects are the following:

- INFORMATION. Participants need to be informed about the project and about the evaluation action. They also need to be given the means to access more information once the evaluation action is finished.
- CONSENT. Evaluation activities can only be carried out once informed consent has been obtained. Informed consent sheets need to be provided in a format and language participants understand.
- VISUAL RIGHTS. Video and audio recordings, images or notes can only be taken if participants have agreed to it.
- DATA PROTECTION measures need to be implemented.

Specific advice for the development of any evaluation action is the following:

- Treat participants with care and sensitivity.
- Be objective and transparent.
- Avoid ethnocentricity: show respect for their ethnicity, language, religion, gender or sexual orientation.
- Rigorously safeguard the dignity, wellbeing, autonomy, safety and security of their family and friends.
- Respect their values and right to make their own decisions.

Traction includes an ethics self-assessment report in which partners are asked to provide details on recruitment, inclusion and exclusion criteria, interactions with vulnerable groups and feedback procedures.

6.2 Responsibilities

The responsibilities in the different evaluation steps are distributed as follows, targeting at a participatory approach to research:

Step	Who?
Determine goal of the evaluation action and best method	Evaluation leaders
Developing instructions and protocol for the evaluation action, providing training if necessary	Evaluation leaders with support from relevant experienced partners
Performing the evaluation action	Evaluation action responsible individual
First level analysis and reporting on the evaluation action	Evaluation action responsible individual
Second level analysis of evaluation action	Evaluation leaders

An **evaluation leader** is a partner leading evaluation tasks in WP4, namely CWI (Alina Striner, Jie Li), UAB (Anna Matamala, Olga Soler-Vilageliu) and FM (François Matarasso).

An **evaluation action responsible** is an individual responsible for an evaluation activity. Evaluation actions will generally be performed by WP3 trial members in what can be called a self-evaluation, but WP4 evaluation leaders will be providing clear protocols and guidance, with adequate training when needed. To ensure an adequate communication between WP4 evaluation leaders and WP3 trial leaders, **evaluation coordinators** have been appointed for each trial: Patricia Malpas for INO, Eva Garcia for LICEU, and Sandra Sousa for SAMP.

The choice of method for each evaluation action will depend on the goal of the action and may differ depending on the phase and target participants. The selection will be done collaboratively among evaluation leaders and evaluation coordinators so as to guarantee it is feasible to actually implement it.

The data analysis will take place at two levels: on the one hand, the evaluation action responsible will provide a report based on a pre-established template in which key aspects will be summarized. On the other hand, evaluation leaders will take this analysis further by providing a second-level analysis, in which the different evaluation actions are interrelated. All reports will be made in English in order to ensure that all partners can have access to the data. This information will be shared with the consortium and will inform further developments and next steps to be taken.

6.3 Evaluation methods

As a first step in the project, three actions were performed: first of all, an overview of possible methods was shared and discussed with partners; secondly, an overview of the consortium expertise on methodology was gathered; finally, a protocol for the methods selected for the first phase was developed.

6.3.1 Traction overview of methods

An overview of methods relevant to the project, especially when evaluating user experience of different Traction components, was developed at the beginning of the

project by CWI and commented by partners. The suggested list includes the following methods:

- In-depth interview
- Survey
- Focus group
- Usability study
- Expert/heuristic evaluation
- Self/peer evaluation
- Co-design study
- Physiological sensors

In the following subsections we 1) summarize the general method and/or procedure, 2) discuss constraints and points of caution, 3) overview types of analyses that can be performed for this method, and 4) provide a list of academic and professional references for the method. As the project evolves, other methods may be included.

6.3.1.1 In-depth interview

In-depth interviewing is a technique that engages with the research subject in a way that encourages a detailed exchange of information. The general purpose of in-depth interviewing is to discover the respondents' perceptions or to probe into a subject to explore nuances and details. In-depth interviews vary from informal conversations to more formal interviews, which may be unstructured, semi-structured or structured. In-depth interviews are also referred to as “ethnographic interviews”, “open-ended interviews” or “case study interviews”. Case study interviews focus on exploring a particular case, which might be, for example, a single organization, community or group that is being used as a detailed illustration of a particular phenomenon [2] [2]. Table 8 provides a summary of three different types of in-depth interviews, namely structured, semi-structured and unstructured. In-depth interviews can be conducted with one interviewee or a group of interviewees [4].

Table 8: A summary of different types of in-depth interviews

Types of in-depth interviews	Individual	Group
Structured	A set of broad questions asked more or less in order but flexible to adjust to the interviewee’s answers. Primarily, it is a one-way data collection technique (data from the interviewee).	Group interviews with question-by-question structured discussion guided by the interviewer.
Semi-structured	A set of broad topics/questions for discussion, the interviewee is encouraged to develop ideas and the interviewer ensures the interviewees to tell concrete stories and details.	Questions posed by the interviewer can be answered by any group interviewees. They may elaborate based on responses of other group members.

	The interviewer may contribute but it is mainly a one-way information flow from the interviewee.	
Unstructured	Open conversation between the interviewer and the interviewee on a set of broad topics, which are referred to as "open ended". In this case, the one-way information flow may develop into a dialogue where the interviewer contributes as much as the interviewee.	A guided conversation by the interviewer (facilitator) with a group of interviewees. The interviewer facilitates the information exchanges and prompts further discussion amongst the group.

During in-depth interviews, the interviewer should avoid asking questions that only trigger "yes" or "no" answers. The questions should guide the interviewees to tell detailed stories. The analysis of the interview data involves transcribing the audio recordings and open coding the transcripts with a group of researchers [5].

6.3.1.2 Survey

Surveys are a flexible research technique to efficiently gather data from a range of different participants. Surveys can be conducted online, by telephone, by post, in person, or as a hybrid of several methods [6]. Surveys methods allow researchers to make statistical inferences about populations, however this depends on the types of questions that are asked of participants, and the size of the participant pool.

To conduct a survey, investigators must first determine the goals of their survey, and what kind of data they want to collect [7]. Then, investigators should pinpoint their target demographic, and identify the number of respondents they need. Finally, they should determine when an appropriate time to conduct the survey is, and what format they should use.

Investigators should use caution when designing surveys, taking survey length, wording, and counterbalancing into consideration [8]. Survey length is important to consider because audiences are less likely to respond to long surveys, so shorter surveys are likely to be completed by more participants. Further, long answers may lose audience attention, so it is vital to determine which questions are the most important to elicit.

Likewise, wording can affect survey quality because it may bias participant responses [7]. Researchers should be careful about word choice, and should avoid asking "double barrelled" questions composed of two or more issues or topics with one response (e.g. Is this tool interesting and useful?). Finally, surveys are subject to question and scale ordering bias, so it is important to counterbalance questions and scale order to ensure internal survey validity. These factors affect the quality of data and types of inferences investigators can make, and thus are paramount to take into consideration during survey design.

6.3.1.3 Focus group

A focus group is a group discussion on a particular topic organized for research purposes. This discussion is guided, monitored and recorded by a researcher or a facilitator [8] [9]. A focus group usually consists of a group of people in a room who are asked to draw on their personal experience to discuss a specific topic selected by the researcher. The structure of the discussion is often loose, lacking specific questions and inviting any views on the general topic of the focus group. When the group consists of up to six people, they usually discuss the topic in the group as a whole. When the focus group is larger, the process often involves the group breaking into smaller clusters to explore the topic in detail followed by a plenary where the key points from the subgroups are discussed by the whole group [10]. Focus groups are used for generating information on collective views, and the meanings that lie behind those views. They are also useful in generating a rich understanding of participants' experiences and beliefs. Suggested criteria for using focus groups include:

- As a standalone method, for research relating to group norms, meanings and processes
- In a multi-method design, to explore a topic or collect group language or narratives to be used in later stages
- To clarify, extend, qualify or challenge data collected through other methods
- To feedback results to research participants.

When running a focus group as a facilitator, one should ensure every participant of the focus group equally contributes to the discussion. Sanders [11] developed a set of generative tools (i.e., probes, see [12]) to inspire the participants to reflect in depth about their experiences, options and ideas. The analysis of the focus group involves transcribing the audio recordings and open coding the transcripts with a group of researchers [5].

6.3.1.4 Usability study

Usability studies are an important research technique to evaluate the design of a system, product, or experience by testing it directly on users. The goal of usability testing is to measure how well a design meets its intended purpose, and how intuitive it is for users with no prior experience to learn to use the tool [13]. There are several ways to conduct usability testing including 1) hallway testing, 2) formal user studies, and 3) A/B Testing. Hallway testing is a quick and cheap method that asks people passing in the hall to quickly try a product or service. This can help designers identify major problems in the early stages of a new design. A second method is user testing, in which an investigator asks a participant to perform tasks using an interface or system. As participants complete each task, the investigator observes participant behaviour and listens for feedback. The third method, A/B testing, compares two versions (A and B) of an experience with a variation that may impact a user's behaviour.

Formal usability studies often follow rigorous research practice [14]. To conduct a usability study, investigators should first decide what they want to test (e.g. interaction or workflow), and choose a goal for the study. Then, they should choose study tasks that reflect study goals, such as finding out if one workflow is better than another. Next, they

should create a hypothesis for the study, predicting what results they think will happen. Having determined a hypothesis, they can choose or create an evaluation metric based on the study tasks, and decide on an appropriate sample size. This may include biofeedback and behaviour data (e.g. eye gaze), quantitative responses to survey questions (e.g. the System Usability Scale), qualitative open-ended responses, and user feedback through audio and/or video recording.

Investigators should create an analysis plan for their study based on their evaluation metrics, write a study plan and script, and delegate moderator roles [14]. Having created a study plan, they should pilot test the study to make sure it is not overly long and runs as they expect, recruit participants, and run the study. After running the study, investigators should conduct qualitative or quantitative analyses based on their study plan, report main takeaways, and plan for the next round of testing.

Several factors affect the quality of usability testing [13]. First, iterative testing early in design is vital before conducting a formal usability study. When conducting usability testing, investigators should be aware of user body language (e.g. squinting eyes, hunched shoulders, shaking heads, and deep, heart-felt sighs), and when users face a problem, should directly ask users why they are confused. Finally, investigators should not ask leading questions that can bias their participants' opinion.

6.3.1.5 Expert/heuristic evaluation

Expert or heuristic evaluation is a quick and cheap method for identifying usability problems in interface design [15]. During a heuristic evaluation, a set of domain experts examine an interface, and judge how well it complies with recognized usability principles, often referred to as the "heuristics." A typical evaluation lasts 1-2 hours, and literature suggests approximately 5 evaluators will be able to find 75% of issues [15].

To conduct a heuristic evaluation [15], investigators should choose a small set of domain expert evaluators that are representative of the user population. If the system is ready to use and/or evaluators are experts, evaluators should use the system without assistance. If the system is not fully completed or evaluators are not experts, an investigator can provide assistance, by giving evaluators a use case scenario and a list of workflow tasks.

During the evaluation, investigators ask evaluators to verbalize (or "think aloud") their comments to an observer as they go through the interface, inspecting elements of the interface and dialogue, and noting problems. During the evaluation, the expert users should compare interface elements with a list of heuristics. Heuristics can be recognized usability principles described by Nielson [16], or could be domain-specific. During the evaluation, an observer takes notes and interprets how user actions are related to the usability issues in the design of the interface.

Several considerations should be taken into account for this method. First, one expert will not find all of the usability problems in an interface, so investigators should not rely solely on one evaluator's findings. Next, when an evaluator has problems using the interface, they can be given help to proceed in order to not waste time. However, they should only be given help when they have significant problems, and have clearly identified the usability problem. At the end of the individual evaluations, experts should summarize their findings

to eliminate duplicates, and create a list of issues that should be addressed. The output of a heuristic evaluation should be a list of specific usability problems in the interface, referencing usability principles that were violated by the design.

6.3.1.6 Self/peer evaluation

Self and peer evaluation is an educational method where students evaluate or grade each other for learning purposes. Peer assessment can help students take responsibility for and manage their learning, motivate students to engage deeply with course materials, encourage them to give others constructive feedback, and increase knowledge diffusion and exchange of ideas. Several online survey tools can be used to implement a self or peer evaluation, such as Qualtrics, Google Forms, and SurveyMonkey [17].

To create a self or peer evaluation, it is important to set clear goals and expectations for the process before beginning evaluation, and explain rationale for peer feedback. First, investigators should create a detailed rubric or checklist to ensure evaluations are respectful and constructive, and provide examples of effective evaluation [17]. To encourage self-direction and responsibility, they should allow students to create personal rubrics or checklists that are approved.

It is important to break larger assignments or tasks into small pieces, and to include peer feedback at each stage of the task [18]. Investigators should create small feedback groups where written comments can be explained and discussed. Further, investigators should listen to group feedback discussions, and provide guidance and input when necessary. For effective evaluation, students should be given the opportunity to practice self or peer evaluation during a low stakes activity, such as class discussion. Self or peer evaluations should be evaluated based on the rubric or checklist created by investigators.

6.3.1.7 Co-design study

The practice of co-design allows users to become an active part of the creative development of a product by interacting directly with design and research teams. It is grounded in the belief that all people are creative and that users, as experts of their own experiences, bring different points of view that inform design and innovation direction. Co-design is a method that can be used in all stages of the design process, but especially in the ideation or concepting phases. Partnering with users ensures their inclusion in knowledge development, idea generation, and concept development on products whose ultimate goal is to best serve these same users [19].

Co-design sessions usually have two main stages [20]:

- Stage 1: Prepare participants beforehand, allowing them to think and reflect about the topic to be studied. Give them sufficient time (e.g., one week) to do the preparation. The research methods for this initial preparation prompt the participants to reflect on activities that otherwise might be taken for granted in everyday life, and reflect on experiences they normally perceive as routine. Diary studies, journaling, workbooks, and day-in-the life exercises are generally used in this stage.
- Stage 2: Run co-design workshops. Co-design workshops can be conducted with a mixed group of users, designers, and researchers, which can yield rich results. It is

important that the researcher guides the sessions at all times and equips participants with tangible and visual materials to help them express the ideas.

Figure 1 and 2 provide some examples of the tangible and visual materials [20]:

Figure 1: Example of collage toolkit developed by Liz Sanders [20]

Figure 2: Paper prototyping and sketching: Example of co-design sessions in software interface design [20]

When working with a large group of participants in a co-design session, it is best to break the group into subgroups of a maximum three members and monitor their work and internal discussions as best as possible. Each subgroup can work on the same theme or variations of it, presenting their work to the rest of the group at the end of the session. Most co-design workshops have a duration of 1.5-2 hours, although some can be longer if working with expert and highly passionate users. The conversation is generally opened by examining the results from the self-reflection research exercises completed in stage one, and is followed by one or more hands-on exercises [20]. Common analysis methods for co-design sessions are similar to those used for focus groups and in-depth interviews, including open coding of the recorded audio, open annotation of the recorded video and the collected artefacts [5].

6.3.1.8 Physiological sensors

Physiological sensors provide an objective measure of physiological signals. In comparison with other sources of emotional information (facial expressions, gestures, and speech), physiological signals offer substantial advantages; they are mostly involuntary and, as such, are quite insensitive to deception, they can be used to measure the affective states continuously, they can be observed as early as 200ms after emotional stimuli in the case of neurophysiological signals and, usually, they are non-disruptive to task performance [22] [23] [24] [25].

Most of these measures are based on recordings of electrical signals produced by brain, heart, muscles, and skin [26], including (1) EMG that measures muscle activity; (2) Electroencephalography that measures brain activity; (3) EDA or skin conductance that measures the hydration in the epidermis and dermis of the skin. It is typically recorded from the surface of the hand or wrists; (4) EKG or ECG that measures heart activity (heart rate, inter-beat-interval, heart rate variability); (5) Electrooculogram measuring eye pupil's size and movement (eye tracking); (6) BVP measures blood pressure; and (7) Respiration, where rate of respiration and depth of breath are the most common measures. Figure 3 illustrates the above-mentioned seven types of physiological sensors.

A potential pitfall of physiological metrics is that they are often obtrusive or even invasive, troubling the user's experience with the interface. They also necessitate specialized and frequently expensive equipment and technical expertise to use the equipment [27]. However, these drawbacks are alleviated with the development of new light physiological sensors, which are more and more user friendly [28].

Figure 3: An overview of physiological signals [25]

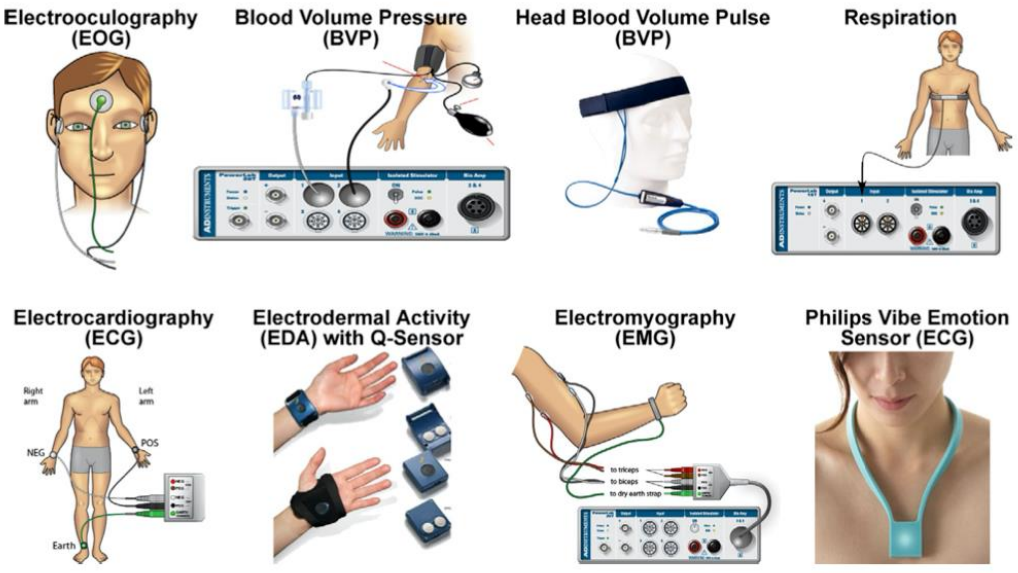


FIGURE 11.2

Bio-physiological signals (Respiration, Electrocardiography (ECG), Electromyography (EMG), Blood Volume Pressure (BVP), 2015; Electrooculography (EOG), 2015; Q Curve EDA sensor, 2015; Head Blood Volume Pulse Sensor, 2015; Vibe Emotion sensor, 2015).

6.3.1.9 Traction methodological expertise

To gather input on the different partners' methodological expertise an online survey was circulated at the beginning of the project (March 2020) asking the consortium the following questions:

- Identification data: name, partner, trial and WP in which they are involved.
- Research methods:
  - Experience with the above-mentioned methods (in-depth interviews, surveys, expert/heuristic evaluation, usability study, focus group, physiological metrics, self/peer evaluation, co-design study, participant observation) in which they could choose three options: no experience, novice, expert.
  - An open field asking survey respondents to describe their experience in these methods.
  - Is there another research method relevant for Traction that is not listed above?
- Research questions and interaction techniques. Respondents were asked
  - Are there any interaction techniques that you would like to advance? Please explain.
  - Do you have any research questions related to these trials that you are interested in answering (education, workflow, community building), or directions you would like to explore? Please explain.

Answers provided us with a wide overview of the consortium in terms of methodological expertise.

Table 9: Consortium methodological expertise

	Expert	Novice	No experience
In-depth interview	6	5	9
Survey	7	10	3
Expert/heuristic evaluation	0	7	13
Usability study	6	4	10
Focus group	5	7	8
Physiological metrics	1	14	5
Self/peer evaluation	4	5	11
Codesign sessions	3	3	14
Participant observation	6	7	7

When looking at the information by partners, we observed that putting together the expertise of all WP4 evaluation partners, experience in all methods is guaranteed:

- CWI is experienced in all methods.
- UAB is experienced in all methods except for co-design.
- FM is experienced in interviews, surveys, focus groups, self/peer evaluation, and participant observation.

Concerning trial leaders:

- LICEU reports being an expert in all methods, and a novice in expert/heuristic evaluation and usability studies.
- INO reports having some experience in surveys but no experience in the other methods.
- SAMP reports being novice in focus groups, surveys, participant observation, and self/peer evaluation.

Concerning technological partners:

- CWI is experienced in all methods.
- DCU is experienced in all methods except for co-design sessions.
- VICOM has some experience with surveys, usability studies, focus groups, self/peer evaluation, codesign sessions and participant observation.
- VRI has some experience with interviews, surveys, focus groups, and co-design sessions.

This broad picture allows us to confirm the partnership has the methodological expertise to implement any of the presented methods, as it has experts in all of them. However, it also highlights the need to provide adequate guidance to trial leaders taking into account the participatory approach taken in which trial leaders will be performing the evaluation activities themselves.

This questionnaire also allowed us to identify additional suggested methods and interaction techniques proposed by the respondents.

Suggested methods:

- Recorded transformations
- Creative evaluation methods
- Quality of experience assessment methods
- Perception questionnaires
- Discourse analysis

Interaction techniques:

- Validate an audience interactivity survey
- Interaction with virtual artifacts
- Automated video analysis

When asking the consortium about their research questions, we obtained a map of the main research interests. This map shows how the different evaluation axes identified before are closely interrelated: technology and tools, co-creation and training, opera performance and capsules, and social transformation.

- **Audience participation and engagement**, linked to the co-creation process, the actual performance and the capsules, and the technology.
    - Audience participation questions such as defining different types of audience goals, design interactivity for multiple levels of audience participation,

designing audience emotional and physical well-being, making passive audience experiences feel more interactive.
    - How to engage audiences with VR operas? How to convert the passionate energetic performance of live Opera to VR e.g. using rumble wear/other haptics along with the VR visual and audio content?
    - See how engaging VR opera experiences are.
- **Technology-based co-creation**, linked to the co-creation process, the actual performance and the technology, with an emphasis on **VR/AR** in one of the trials.
    - Unlocking new creative parameters in co-creation through technology
    - How does immersive technology improve the experience of the artists in a co-creation process? What benefits can be achieved in terms of the final artistic process?
    - Better understand user experience and/or expectations involving the addition of technology into the co-creation.
    - How does immersive technology improve the experience of users (community members/non-professional artists) in a co-creation setting?
- **Creation/perception of adaptive** content, linked to technology.
    - Adapting content to fit the operational environment, e.g. device type, network conditions to improve QoE.
- **Quality of experience** in technology.
    - Perceived quality of technology (video, audio, etc.), technology sustainability, QoE meeting expectations.
- **Measuring and strengthening relationships**, closely related to social transformation impact.
    - Consequences on self-image and link/belonging to the community
    - How to measure the relationship between the inmates, their families and the guards.
- **Dissemination/sustainability/impact**, related to technology, capsules and social transformation.
    - What are the benefits of immersive technology as a dissemination tool for opera? Does the VR experience add anything to the opera experience?
    - Success factors in creating an impact (societal, technological, etc.).
    - Audience experience at the dissemination stage.
- **Impact on education/rehabilitation,** closely related to skills acquisition in co-creation and social transformation.
    - Impact of opera experience on rehabilitation of the offenders.
    - Traction technology impacting long-term education (e.g. learning new skills) of the participants.
- **Accessibility**, which impacts on all the other dimensions.

7 How? Planning and reporting

Evaluation activities developed in Traction will follow the steps described in this section.

7.1 Planning

Any activity needs to be carefully planned to guarantee its success. For each evaluation method, some guidance will be provided and a specific protocol will be developed. This will allow us to gather consistent data across trials of all evaluation activities.

Each activity will have a unique ID code.

This protocol will include:

a) General information to be gathered for any evaluation action. For instance:

- Aim:
- Participants:
- Recruitment process:
- Timeline:
- Place:
- Language of the test:
- Materials:
- Preparatory tasks:
- Evaluation action responsible:

b) Specific protocol: step-to-step process indication, with guiding questions or activities where relevant.

7.2 Reporting

Any evaluation action will generate at least two outputs:

Evaluation report

A report will be generated for each evaluation action, which will include:

- General information such as:
    - Aim:
    - Date:
    - Place:
    - Number of participants:
    - Evaluation action responsible:
- A summary of the participants' profile
- Main findings
- Researcher's observations, where relevant.

8 When? Evaluation timeframe

The timeframe for the evaluation actions depends the timeframe for each trial (Table 10), on the technological development (Table 11) and on WP4 timeframe (Table 12). Evaluation actions will take place before the performance (see Section 9), during the performance (see Section 10) and after the performance (see Section 11), following the same steps (Table 13).

Table 10: Timeframe for trials

STAGES/TRIALS	LICEU	SAMP	INO
Community dialogue	M1-M6	M1-M12	M1-M12
Co-creation, training, production	M7-M22	M13-M29	M13-M24
Performance	M23	M30	M25-M36
Digital capsules	M24-M36	M31-M36	M30-M36

Table 11: Timeframe for technology

TOOLS/TECHNOLOGY	Iteration 1	Iteration 2	Final version
Media vault	M9-M12	M17-M22	M31-M36
Performance engine	M9-M12	M17-M22	M31-M36
Installation or capsule	M9-M12	M17-M22	M31-M36

Table 12: Timeframe for WP4 evaluation activities and deliverables

Task		Timeframe	Deliverables
4.1.	Overall evaluation methodology	M4-M36	M8 (D4.1), M23 (D4.6)
4.2.	Traction toolset technical and UX evaluation	M8-M36	M12 (D4.2), M23 (D4.7), M36 (D4.8)
4.3.	Co-creation and user training evaluation	M7-M30	M16 (D4.3)
4.4.	Opera full performances and capsules evaluation	M23-M36	M36 (D4.4)
4.5	Social transformation impact	M7-M36	M34 (D4.5)

Table 13: Steps to be followed in all evaluation actions

STEP	SUGGESTED TIMING
Defining a protocol based on the template for each evaluation action (or groups of actions that follow the same protocol)	6 weeks before the action
Preparation of the action following the protocol	4 weeks before the action
Performing the evaluation action	Variable length and timing
Reporting on the evaluation action using the template	10 days after the activity at the latest
Analysing the evaluation action	In the deliverable

9 Phase 1: before the performance

The first phase will include all actions before the performance takes place. During this phase, the technology will be developed and the process of co-creation, with its related training, will take place. Next we provide an overview of the specific objectives, methods and related protocols, keeping in mind that this is a general framework and that any necessary adjustments can be made as the project progresses.

9.1 Specific objectives of the phase

The first objectives would be to:

a) gather user requirements in an iterative process on the different Traction tools and technologies, and evaluate the technological solutions developed in different iterations (T4.2);
b) evaluate the co-creation and training process as it develops towards the performance, based on a previous understanding of the concept of co-creation (T4.3);
c) gather aims and objectives for each trial, as a necessary first step to assess the social transformation impact (T4.5).

9.2 Summary of methods

The main expected methods at this stage are:

- Focus groups, to gather requirements (T4.2.).
- In-depth interviews, to gather aims and objectives (T4.5), define what is understood by co-creation (T4.3) and, if relevant, to gather technical requirements (T4.2).
- Evaluation sheets or questionnaires to assess the co-creation and training process (T4.2).
- Participant observation (T4.5).

9.3 Protocols

To follow a unified approach in the development of focus groups and interviews, a protocol for each method has been developed, including a template for focus groups (see Annex 1) and a template for interviews (Annex 2). Additionally, an initial version of the co-creation and training activities evaluation log has been developed (Annex 3), as well as an agreed list of questions to be asked to cultural stakeholders concerning co-creation processes (Annex 4) and an initial participant observation protocol (Annex 5).

10 Phase 2: the performance

The second phase in the evaluation will be the actual performances, which will take place in different moments in the project depending on the trial.

10.1 Specific objectives of this phase

The specific objectives of this phase are to:

- evaluate the value of the technology and tools on the performance experience (T4.2),
- evaluate the experience of the community involved as artists or in other ways in the performance and its ancillary activities (T4.4);
- evaluate the artistic output of the trials through an expert evaluation (T4.4.).

10.2 Summary of methods

It is expected that the main methods at this stage may be:

- Surveys
- Focus groups
- In-depth interviews
- Participant observation

10.3 Protocols

Protocols will be developed in an updated version of this deliverable and included as annexes.

11 Phase 3: after the performance

The third phase in the evaluation will take place after the performances. This is a very broad initial approach, as this will be further developed as the project progresses.

11.1 Specific objectives of this phase

The specific objectives of this phase are to:

- evaluate how well the collected user requirements fit the system design needs for the three trials;
- evaluate the global co-creation and training processes after the actual performance;
- evaluate the social transformation impact of the three trials.

11.2 Summary of methods

It is expected that the main methods at this stage may be:

- Focus groups
- In-depth interviews
- Survey
- Rubrics
- Participant observation

11.3 Protocols

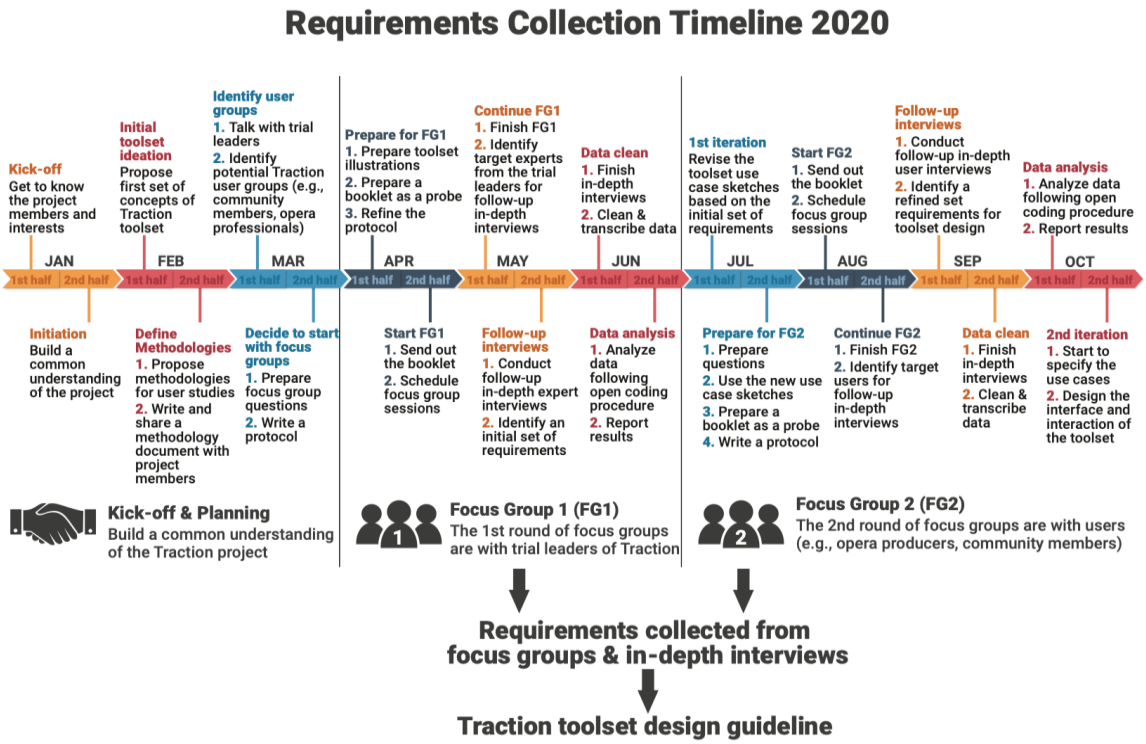
Relevant protocols will be developed, if necessary, in an updated version of this deliverable and will be included as annexes.

12 A roadmap

At this stage of the project, an initial roadmap for T.4.2 and T.4.3 has been planned.

12.1 Roadmap for T4.2

Figure 4: Requirements collection timeline 2020



12.2 Roadmap for T.4.3

Step 1. Defining what is understood by co-creation (and its associated training) through interviews with cultural stakeholders, associations, and trial leaders. September-November 2020.

Step 2. Evaluation of participation in co-creation through an evaluation sheet (Annex 3). This will be tested in exploratory workshops in July (SAMP) and workshops in September-October (INO) and will be actually implemented during Year 2.

Step 3. Based on the definition process and the participation evaluation, focus group discussion on the co-creation process and its effects. Non-professional artists will be involved at this stage, not before, because they may not be familiar with co-creation processes and asking about their expectation may not be productive. Date to be confirmed.

13 Conclusion

Traction aims to have a social transformation impact by means of co-creation processes in three very different settings. Each of the trials has its own specificities and follows its own processes and paths, based on a global set of best practices. Evaluation activities will need to take into account these specificities, but a general framework that is shared across trials will undoubtedly make the evaluation process easier. This has been the aim of this deliverable: establish a rationale for the Traction evaluation methodology and define a clear path of action trying to reply to five central questions: What will be evaluated? Where will the evaluation take place? Who will be the persons involved in the evaluation? How will the evaluation be performed? When will the evaluation take place?

Some questions have a clear answer at this stage of the project. Others will need to be further developed as the project progresses. As indicated at the beginning of this document, our methodological approach is iterative, open, adaptable, and flexible. This future developments in Traction methodology will be documented in D4.6, *Final evaluation methodology* (November 2021).

References

[1] F. Matarasso, *A Restless Art: How participation won, and why it matters*, Lisbon, London, Calouste Gulbenkian Foundation, 2019.

[2] V. Minichiello, R. Aroni, R., T.N. Hays, T.N., *In-depth interviewing: Principles, techniques, analysis*, Pearson Education Australia, 2008.

[3] J.M. Johnson, J. M., In-depth interviewing, *Handbook of interview research: Context and method*, *1*, 2002.

[4] L. Harvey, *Researching the Real World*, 2000. Retrieved Apr 06, 2020 from qualityresearchinternational.com/methodology.

[5] D.R. Thomas, A general inductive approach for analyzing qualitative evaluation data, *American journal of evaluation, 27*(2), pp. 237-246, 2006.

[6] S. Sincero, *Types of Survey*, Sep 21, 2012. Retrieved Apr 06, 2020 from Explorable.com: https://explorable.com/types-of-survey

[7] F.J. Fowler, *Applied Social Research Methods: Survey research methods* (4th ed.), Thousand Oaks, CA, SAGE Publications Inc, 2009. doi: 10.4135/9781452230184

[8] D.L. Morgan, D. L., M.T. Spanish, M. T. Focus groups: A new tool for qualitative research, *Qualitative sociology, 7*(3), pp. 253-270, 1984.

[9] D.L. Morgan, *Focus groups as qualitative research* (vol. 16), Sage publications, 1996.

[10] J. Kitzinger, Qualitative research: introducing focus groups, *Bmj, 311*(7000), pp. 299-302, 1995.

[11] E.N. Sanders, Generative tools for co-designing, *Collaborative design*, pp. 3-12, Springer, London, 2000.

[12] P. Gill, K. Stewart, K., E. Treasure, E., B. Chadwick, B., Methods of data collection in qualitative research: interviews and focus groups, *British dental journal, 204*(6), pp. 291-295, 2008.

[13] J. Nielson, J., Usability Testing 101, 2019. Retrieved Apr 06, 2020 from https://www.nngroup.com/articles/usability-testing-101/

[14] E. Geisen, J.R. Bergstrom, Usability and Usability Testing, In *Usability testing for survey research*, edited by Emily Geisen and Jennifer Romano Bergstrom, Morgan Kaufmann, Boston, pp. 1-19, 2017.

[15] J. Nielson, How to Conduct a Heuristic Evaluation, 1994. Retrieved Apr 06, 2020 from https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/

[16] J. Nielson, 10 Usability Heuristics for User Interface Design, 1994. Retrieved Apr 06, 2020 from from https://www.nngroup.com/articles/ten-usability-heuristics/

[17] L. Schankman, Peer evaluation, In *Teaching Improvement Guide*, University of Wisconsin at La Crosse Center for Advancing Teaching and Learning, 2015. Retrieved Apr 06, 2020 from http://www.uwlax.edu/catl/teaching-guides/teaching-improvement-guide/how-can-i-improve/peer-evaluation/.

[18] E. Panadero, Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In Handbook of Human and Social Conditions in Assessment, edited by G. T. L. Brow, L.R. Harris, pp. 247-266, New York, Routledge, 2016.

[19] F.S. Visser, P.J. Stappers, R. Van der Lugt, E.B. Sanders, E. B. Context mapping: experiences from practice, *CoDesign, 1*(2), pp. 119-149, 2005.

[20] C. Naranjo-Bock, Creativity-based Research: The Process of Co-Designing with Users. *UX Magazine*, 820, 2012. Retrieved Apr 06, 2020 from https://uxmag.com/articles/creativity-based-research-the-process-of-co-designing-with-users

[21] S. Shahi, *Planning a co-design session*, 2018. https://wiki.fluidproject.org/display/fluid/Planning+a+Co-Design+Session

[22] C. Wang, E.N. Geelhoed, P.P. Stenton, P. Cesar, P., Sensing a live audience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1909-1912, 2014, April.

[23] N. Shirzadian, J.A. Redi, T. Röggla, A. Panza, F. Nack, F., P. Cesar, Immersion and Togetherness: How Live Visualization of Audience Engagement Can Enhance Music Events. *International Conference on Advances in Computer Entertainment*, pp. 488-507, Springer, Cham, 2017.

[24] E. Gedik, L. Cabrera-Quiros, C. Martella, G. Englebienne, H. Hung, Towards Analyzing and Predicting the Experience of Live Performances with Wearable Sensing*, IEEE Transactions on Affective Computing*, 2018.

[25] S. Gashi, E. Di Lascio, S. Santini, S., Using Unobtrusive Wearable Sensors to Measure the Physiological Synchrony Between Presenters and Audience Members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(1), pp. 1-19, 2019.

[26] M. Feidakis, A Review of Emotion-Aware Systems for e-Learning in Virtual Environments, *Formative Assessment, Learning Data Analytics and Gamification*, pp. 217–242, 2016.

[27] M. Wong, Emotion assessment in evaluation of affective interfaces, *Neuron, 65*(3), 293, 2006

[28] L.D. Liao, C.T. Lin, K. McDowell, A.E. Wickenden, K. Gramann, T.P. Jung, J.Y. Chang, Biosensor technologies for augmented brain–computer interfaces in the next decades, *Proceedings of the IEEE, 100(Special Centennial Issue)*, pp. 1553-1566, 2012.

Annexes

Annex 1. Focus group protocol

What is a focus group?

A focus group is "a form of group interview in which: there are several participants (in addition to the moderator/facilitator); there is an emphasis in the questioning on a particular fairly tightly defined topic; and the accent is upon interaction within the group and the joint construction of meaning" (Bryman, 2012: 502)[1].

First, we describe the moderator's role, the number of participants, the study setup, and the data collection process. Then we describe the required preparation for a focus group, and the focus group protocol.

Focus groups are run by a moderator (also called a facilitator). They are expected to guide each session but not be too intrusive. In Traction, the focus groups will be conducted by a moderator and an assistant. The moderator should be a member of a trial or evaluation team, and will be given specific instructions and assistance so that they are adequately prepared. During the session, the moderator ensures that every member of the focus group contributes equally to the discussion, and prevents some participants from dominating the conversation.

Focus groups can range between 4 and 10 participants depending on the type of activity. Focus groups that include design requirement generation and co-creation typically range between 4-6 people. In Traction, we aim to include 4-6 participants per focus group. We plan to invite 6 to 7 per focus group in the case there are no-shows.

The expected length of a focus group is 60-90 minutes.

Focus groups are traditionally conducted face-to-face. However, online focus groups can also provide valuable and insightful data. During the Traction project, we will prioritise face-to-face focus groups, but online focus groups will be conducted when necessary. In face-to-face focus groups, the room needs to be arranged so that participants feel comfortable and see each other.

Focus groups are a qualitative method. The session can be recorded and subsequently transcribed, or an assistant can take notes during the session. In all cases, explicit consent has to be given for any data collection (notes, artefacts, audio recording, video recording) to take place. For Traction, we suggest that a note-taker is present during each session. A good practice is to provide a summary of the focus group at the end of the session. The summary should not exceed one page. This summary should highlight a set of ideas or conclusions the participants agree on.

Participants can be briefed before the focus group and, where relevant, they can be asked to reply to some individual questions. However, these individual questions are only recommended for certain focus groups –mainly with professionals– and care should be

[1] Bryman, Alan (2012). Social research methods. Oxford: OUP, 4th edition.

taken not to ask too much from participants, as this increases the change of no-shows. Unless necessary, this must be avoided.

Before the focus group

For each focus group, the following decisions need to be made:

- Goal: what is the goal of the focus group.
- Participant pool: who will participate, and how many people should be included in each focus group.
- Recruitment process: how will participants be contacted.
- Timeline: when the focus group will take place, and when the materials should be ready.
- Place: where the focus group will take place (e.g. physical space, online)
- Language: what language(s) the test will take place in, and whether translation support will be included.
- Materials: what physical/digital materials should be ready.
- Preparatory work: whether and what prep materials should be done by participants.
- Guiding questions: what questions and activities will guide the session.

The focus group leader needs to fill in a protocol using the template.

During the focus group

Focus groups in Traction should be conducted using the following steps:

1. Introduction
    a. Moderator introduces themselves and thanks participants for coming to the study
    b. Moderator briefly explains the project, the goal and procedure of the focus group
    c. Moderator explains participants the three ethical forms: information sheet, consent form, and disclosure.
        i. Moderator describes the type of data that will be collected (e.g. audio, video, notes), that this data will be stored confidentially, and that all data that is collected and published will not directly identify participants
        ii. Moderator describes the format in which the information will be published or available (e.g. article, report, publication, website).

    Participants are asked to read the form and ask questions, and then are asked to consent to the study. Forms should be provided in an accessible format and language, and participants should be given enough time to

read and understand the form before consent. Participants should also be told that they can withdraw consent at any time during the study.

d. Moderator asks participants to fill in the demographic questionnaire. Focus group participants should be provided with help where needed.

e. Participants introduce their names and give a brief background about their involvement with the project.

f. Moderator thanks participants, and asks them to speak one at a time during the session. The moderator emphasizes that there are no right or wrong opinions, and that everyone's views are important.

2. Focus group discussion

    a. The moderator facilitates the discussion using a set of topics and open-ended questions and/or activities. Depending on the goals of the focus group, there should be a limited number of topics, and a set of relevant questions that may or may not be answered during discussion, depending on the direction of the conversation. The moderator directs the discussion, encouraging participants to interaction, and to discuss their opinions in a friendly and comfortable tone. As participants discuss each topic, the moderator remains as unobtrusive as possible.

3. Focus group conclusion

    a. The moderator wraps us discussion by reviewing the session tasks.

    b. If participants agree on a set of ideas or conclusions during the session (e.g. design requirements), the moderator reads these aloud. At the end of the session, participants make final edits and reach an agreement about these suggestions.

After the focus group

Focus group responsible fills in the focus group report template.

Template for focus group protocol

FOCUS GROUP ID	METHODOLOGY

1. General information

- Goal:
- Participant pool:
- Recruitment process:
- Timeline:
- Place:
- Language:
- Materials:
- Preparatory tasks:
- Focus group responsible:

2. Protocol: guiding questions or activities

(Please make sure this protocol follows the structure indicated in the protocol.)

Template for focus group report

FOCUS GROUP ID	REPORT

1. General information

- Goal:
- Date:
- Place:
- Number of participants:
- Focus group responsible:

2. Participant profile

(Summarise the participant profile of the participants)

3. Approved conclusions

(Include the main conclusions from the focus group, which have been agreed by participants)

4. Researcher observations

(Provide any relevant information about the development of the focus group and its results)

Annex 2. Interview protocol

What is an interview?

In research, the aim of the interviewer is to "elicit from the interviewee or respondent [...] all manner of information: interviewees' own behaviour or that of others; attitudes; norms; beliefs; and values" (Bryman 2012:209).

In Traction we plan to conduct "**in-depth interviews**", which encourage a detailed exchange of information with a participant.

There are main types of in-depth interviews, depending on different factors. In Traction we will consider two aspects: the number of interviewees, and the structure.

Depending on the number of interviewees, we differentiate between **individual interviews** (just one interviewee) and **group interviews** (more than one). In Traction both types of interviews are possible, depending on the aim of the interview.

It is important to differentiate between focus groups (see focus group protocol) and group interviews. In Traction we will conduct focus groups when our goal is to understand group interaction dynamics and joint construction of meaning on a very focused topic. In contrast, we will conduct group interviews when focusing on broader topics without directly considering interaction. The number of participants in a focus group in Traction will typically be 4-6, whereas the number of participants in a group interview can be lower.

Depending on the structure, three main types of interviews can be differentiated:

- A **structured interview**, or a standardized interview, "entails the administration of an interview schedule by an interviewer" (Bryman 2012: 210). Any interviewee will be given exactly the same questions with the same phrasing and in the same order. Questions are generally specific, and often include a closed list of possible answers the interviewee can choose from. This type of interview is generally chosen because it reduces error due to the interviewer variability and because it allows to process data more easily. It is primarily a one-way data collection technique. In group interviews, all interviewees reply to all questions.
- A **semi-structured interview** is an interview in which the interviewer has a list of questions in what can be called an "interview schedule", but s/he can vary the order of the questions, adapt them or add questions in response to previous replies. Questions are more open than in structured interviews. It is mainly a one-way information flow from the interviewee, but the interviewer may contribute to it where relevant. In group interviews, any question can be answered by any group interviewee.
- An **unstructured interview** is an interview that employs a list of topics (an interview guide) but the interview is not clearly structured. The style is usually informal and the phrasing and order of questions varies from interview to interview. It is more an open conversation in which the interviewer may contribute to the conversation as much as the interviewee.

In Traction we prioritise semi-structured interviews because they allow us to gather input on an interview schedule, and to make changes where necessary. The contexts in which the different trials take place seem to favour this approach as the best way to gather

qualitative –rather than quantitative- feedback. Where relevant, unstructured interviews led by co-participants, which get involved as co-researcher, may be considered. In this case, the interview schedule will just be a general “aide memoire” or agenda.

Interviews can be conducted face-to-face, on the phone, or can conducted through video conferencing software. the phone. In Traction we will prioritise face-to-face interviews where relevant. However, we will also consider conducting online interviews where necessary, such as during the Covid-19 situation.

There can be one interviewer or more than one interviewer. In Traction, both options are possible, taking into account that interviewers may not be experienced in conducting interviews. Thus, we advise having one lead interviewer and one support interviewer that can advise each other if a question is missed or formulated in a wrong way. If having two interviewers present feels intimidating for certain interviewees, then only one interviewer will be present.

Interviews are a qualitative method. Each session can be recorded and subsequently transcribed, or a note-taker can be in the room. In any case, explicit consent has to be given. In Traction we suggest recording the interviews, with previous consent, unless it is considered too intimidating for certain interviewees. Please notice that some online video conferencing systems already include a recording and transcription system that can facilitate this process, with a subsequent human editing process.

The interview should take place in a comfortable and quiet setting and with the adequate recording equipment. Please make sure the tools work.

The expected length of an interview is 30-60 minutes.

Participants can be briefed before the interview, but it is advised not to overwhelm them with preparatory activities unless absolutely necessary.

Before the interview

For each interview, the following decisions need to be made:

- Goal: what is the goal of the interview.
- Interview responsible.
- Interviewer.
- Type of interview.
- Participant(s): who will participate.
- Recruitment process: how participants will be contacted.
- Timeline: when the interview and any preparatory actions will take place, when materials (if any) should be ready.
- Place: where the interview will take place (e.g. physical space, online).
- Language of the interview and, if necessary, translation support.
- Materials needed.

- Preparatory work to be done by interviewee, if any.
- Interview schedule (i.e. list of questions).

The interview schedule needs to be carefully designed. Some advice in semi-structured interviews with an interview schedule is the following:

a) Identify the topics you are interested in and formulate questions. Revise the questions, where possible with external input.
b) Create a certain order on the topic areas, but be flexible to alter the order if necessary.
c) Formulate questions that do not trigger "yes/no" answers. Instead, prompt interviewees to provide comprehensive details which respond to your research questions.
d) Do not ask leading questions.
e) Some participants need time to think. Manage silence wisely, but avoid discomfort. An important ingredient in an interview is listening.
f) Try to use a language that is easy to understand.
g) If necessary, use support material (visual representations, photographs, etc.).

The interviewer should be fully familiar with the interview schedule, so that the possible stress experienced during the interview does not cause the interviewer to miss questions or ask the wrong ones.

The interview responsible needs to fill in a protocol using the template.

During the interview

Interviews in Traction will be developed as follows:

1. Introduction

    a. The interviewer thanks interviewee(s) for coming and introduces himself/herself.

    b. The interviewer briefly explains the project, the aim of the interview and how it will be developed.

    c. The interviewer explains participants the three ethical forms: information sheet, consent form, and disclosure.

        i. The interviewer describes the type of data that will be collected (e.g. audio, video, notes), that this data will be stored confidentially, and that all data that is collected and published will not directly identify participants.

        ii. The interviewer describes the format in which the information will be available (e.g. article, report, publication, website).

    Participants are asked to read the form and ask questions, and then are asked to consent to the study. Forms should be provided in an accessible format and language, and participants should be given enough time to read and understand the form before consent. Participants should also be told that they can withdraw consent at any time during the study.

d. The interviewer asks interviewee to fill in the demographic questionnaire. Interviewee should be provided with help where needed.

e. In a group interview, interviewees present themselves.

f. The interviewer thanks interviewee(s), and, in group interviews, asks them to speak one at a time during the session. The moderator emphasizes that there are no right or wrong opinions, and that everyone's views are important.

2. Interview development

   a. The interviewer follows the interview schedule, trying to engage the participant. The interviewer avoids discomforts and remains as little intrusive as possible, dealing carefully with sensitive issues. The interviewer does not lead interviewees.

3. Interview conclusion

   a. The interviewer thanks the interviewee(s) and indicates where more information about the project, and this specific interview, will be found.

   b. Once the interviewee has left, it is advisable that the interviewer writes a few notes on how the interview went.

After the interview

Interview responsible transcribes the interview and fills in the interview report template provided by UAB.

Interview responsible makes sure the information is available as indicated during the interview (report, article, etc.).

Template for interview protocol

INTERVIEW ID	METHODOLOGY

General information

- Goal:
- Participant(s):
- Interview type: structured/unstructured.
- Recruitment process:
- Timeline:
- Place:
- Language:
- Materials:
- Preparatory tasks:
- Interview responsible:

Interview protocol and schedule

(Write down the list of questions and how the interview will be developed. Please take into account the recommendations in the main body of the document. If an unstructured interview is chosen, please suggest an "aide memoire" or agenda.)

Template for interview report

INTERVIEW ID	REPORT

General information

- Goal:
- Date:
- Place:
- Type of interview: (individual/group). If group, number of participants.
- Interview responsible:

Participant profile

(Summarise the profile of the participant or participants. Personal data need to be anonymous.)

Researcher's observations

(Include a summary of the researchers' notes taken after the interview.)

Main findings

(Include the main findings from the interview.)

Conclusions

(Summarise in a maximum of two paragraphs the main findings.)

Annex

(Provide the interview transcript. If a speech-to-text system is used, please make sure you revise the transcript.)

Annex 3. Evaluation log for co-creation and training activities

Aim: to keep track of the co-creation and training activities in Task 4.3. by means of a self-evaluation performed by evaluation coordinators (or trial leaders) participating in co-creation/training activities. **Method**: evaluation sheet/log to be filled in after each co-creation or training session. **Format**: online form

1. Activity code: *Use CO for co-creation and TR for training. Use INO, SAMP or LICEU to indicate the trial. Then put a number. For instance, CO-INO-1, CO-INO-2 and TR-INO-1, TR-INO-2.*

2. Date of the co-creation or training activity: Day-Month-Year.

3. Length of activity: Less than two hours/ Two to four hours/ More than four hours.

4. Trial: INO/LICEU/SAMP.

5. Location of the activity: Arts venue/ Community venue/ Neutral venue (other venue, not an arts venue nor a community venur).

6. Number of participants according to profile:

Professionals:

- Artistic and creative team: _____
- Technical and production team: _____
- Traction-related professionals (with a professional relationship with the project, such as Consortium members, funders and Advisory Board members): _____

Non-professionals:

- Artistic and creative team: _____
- Technical and production team: _____

Others (any participant who enables the participation of non-professionals, such as psychologists, prison staff, foundations and community associations, to name a few. Their role is to accompany the process, but they are not actively involved in the co-creation): ____

7. Has the aim of the activity been fulfilled? Yes/Partially/No.

8. Has the activity been engaging for participants? Yes, for all./ For some./ No, it was not.

9. Add your personal observations on this activity. This is a field open to any type of comments. For instance, you may want to comment on the aim of the activity, on the co-creation and participation process, engagement, artistic value, impact, skills, change, etc.

Annex 4. Protocol for interviews with cultural stakeholders

This is a first ongoing version that will be adapted to the format in Annex 2 as the project evolves.

Aim: to define what is understood by co-creation by cultural stakeholders, associations and trial leaders across trials and understand how they imagine their social role.

Method: online interview.

List of questions to guide the interview

1. Can you describe your association/cultural venue/organisation, and its purpose?

2. What is the social responsibility policy of your association/cultural venue/organisation?

3. One of Traction's main objectives is to co-create an opera with the participation of non-professional artists. Do you have any experiences in participatory approaches to art, in co-creation in your association/cultural venue/organisation? Please explain.

4. If you use the term "co-creation", how would you define it and what do you believe it means? If you use an alternative term, please say which, what it means to you and why you prefer it to co-creation.

5. How would you assess the success of a co-creation process in the arts? What indicators would you use?

Annex 5. Protocol for participant observation

This is a first ongoing version of the participant observation protocol. Templates are in the process of being discussed.

1. What is participant observation?

In participant observation –also referred to as ethnography—, "the participant observer /ethnographer immerses him- or herself in a group for an extended period of time, observing behaviour, listening to what is said in conversations both between others and with the fieldworker, and asking questions" (Bryman, 2012: 432)[2].

Kawulich (2005),[3] whose paper we have used extensively together with Bryman (2012) and Guest, Namey and Mitchell (2013),[4] to write this summary, points out that participant observation has been used as a collection method in anthropological studies for more than a century, and is currently being used in other areas of social research. The settings that the researcher observes can be closed or open/public. Gaining access to a setting and keeping this access ongoing are challenges faced by the researcher.

According to Guest, Namey and Mitchell (2013: 78), participant observation is different from direct observation in that direct observation does "not inherently require any interaction between the observer and those being studied". It is "about observable behaviour and is typically associated with research objectives that require some sort of ordinal data or purely factual description" (*ibid*). On the contrary, participant observation is qualitative and interactive and relatively unstructured. To achieve this interaction, having some basic fluency in the language of the participants is needed. Interpreters can also be hired, but one must consider the effect on the data.

Bryman (2012: 441-444) differentiates different types of field roles and participation:

- Covert full members: full membership of a group. Researcher's status as a researcher is unknown.
- Overt full member: full membership. Researcher's status is known.
- Participating observer: participates in group's core activities but not as a full member.
- Partially participating observer: same as participant observer but observation is not necessarily the main data source (interviews and documents can be equally or more important).
- Minimally participating observer: observes but minimally participates in group's core activities. Again, interviews and documents can be equally or more important.
- Non-participating observer with interaction. Observes but does not participate in group's core activities. Interaction occurs, generally through interviews.

Kawulich (2005) refers to three types of processes:

- Descriptive observation: one observes anything and everything.

[2] Bryman, Alan (2012). *Social research methods*. Oxford: OUP, 4th edition.

[3] http://www.qualitative-research.net/index.php/fqs/article/view/466/996 (accessed 5 July 2020)

[4] https://www.sagepub.com/sites/default/files/upm-binaries/48454_ch_3.pdf (accessed 5 July 2020)

- Focused observation: observation is supported by interviews which guide the researcher's decisions on what to observe.
- Selective observation: the focus is put on different types of activities and the differences in those activities.

Researchers take notes based on their observations: they are detailed summaries and can also include reflections. Bryman (2012: 450) differentiates between:

- Mental notes, when it is not appropriate to take notes.
- Jotted notes or scratch notes on pieces of paper, to be used when writing about certain events later.
- Full field notes, to be written at the end of the day or as soon as possible. They are the main data source.

Fieldwork involves more than data gathering. It can involve interviews and conversations, but also questionnaires or surveys. Visual materials can also be useful, especially for the so-called photo-elicitation, "whereby the researcher uses the images as a springboard for discussion with the producers of the photographs" (Bryman 2012: 455).

2. Before conducting participant observation

To know what to observe, it is important to know the research purpose. A first step is to decide the role of the researcher (see different profiles above) and the type of observation. In Traction we will prioritise a focused observation in which the researcher is a minimally participating observer, but other possibilities may arise as the project evolves. We will also try to involve participants as co-researchers (participating observer).

Regarding ethics, the community must know that one's purpose for observing is to document their activities. This is why in Traction all participants in co-creation will sign an informed consent form. Anonymity of the participants will be preserved both in the final reports and in field notes, and individuals will be described in a way that community members cannot be identified. Health and safety measures need to be taken into account.

Researchers need to become familiar with the setting and the social organisation. They need to gain entry and establish a rapport (build trust) with the participants. Community leaders who are respected and viewed as neutral can ease this process. Language issues must also be considered.

According to Guest, Namey and Mitchell (2013: 94), a process model template "helps lay out sequenced or stepwise processes and decision pathways".

3. During participant observation

Rapport-building implies active listening, showing respect, empathy and commitment. Reciprocity and confidentiality are also key. Although prolonged interaction with the community is important, there are also shorter participants observations in which rapid assessment techniques are used (Kawulich 2005, *apud* Bernard 1994). Community members become research partners to get the questions answers.

Kawulich (2005) gathers different tips from various authors for conducting observation. We reproduce some of those that may be applicable to Traction:

- be unobtrusive, adaptable and flexible, keeping aside your preconceptions;
- become familiar with the setting before beginning to collect data;
- keep the observations short at first to keep from becoming overwhelmed;
- actively observe, attending to details one wants to record later, and pay attention, shifting from a "wide" to a "narrow" angle perspective, focusing on a single person, activity, interaction, then returning to a view of the overall situation;
- look for key words in conversations to trigger later recollection of the conversation content;
- during breaks in the action, mentally replay remarks and scenes one has observed;
- look at the interactions occurring in the setting, including who talks to whom, whose opinions are respected, how decisions are made. Also observe where participants stand or sit, particularly those with power versus those with less power or men versus women;
- counting persons or incidents of observed activity is useful in helping one recollect the situation, and
- practice reciprocity in whatever terms are appropriate.

4. After the participant observation

Observations become data when they are transferred into field notes (written/recorded). Jotting notes, spatial and interaction maps or process flows can be useful during the participant observation, but it is important to keep an up-to-date observation record, describing what is observed, including casual conversations and interviews with participants and records of activities. It is advisable to keep track of day, date and time, together with a coding system to emphasise information, and make sure researchers' thoughts are separated from actual observations. Some aspects that are generally gathered in the field notes according to Merriam (1998), apud Kawulich (2005), are:

- Physical environment
- Participants (using pseudonyms to keep confidentiality)
- Activities and interactions that occur
- Frequency and duration of the activities/interactions and other subtle factors (for instance, conversations: who speaks to whom, who listens, silences, etc.)

Guest, Namey and Mitchell (2013: 94) suggest keeping a reporting summary template to summarise key aspects from lengthy notes, recordings or transcripts. When analysing these notes and writing ethnography, different approaches can be used, from dispassionate third-person accounts to personalised tales (Bryman 2012: 462-464).