Intelligent Verification/Validation for XR Based Systems

Research and Innovation Action

Grant agreement no.: 856716

D1.2 – Second Project Report

iv4XR – WP1 – D1.2

Version 1.5

December 2021

Project Reference	EU H2020-ICT-2018-3 - 856716
Due Date	31/12/2021
Actual Date	29/12/2021
Document Author/s	Rui Prada, Marta Couto, Pedro Fernandes (INESC-ID), Beatriz Marin, Fernando Pastor, Tanja Vos (UPV), Joseph Davidson (GA), Jeremy Cooke (GWE) Wisnhu Prasetya (UU), Fitsum Kifetew (FBK)
Version	1.5
Dissemination level	Public
Status	Final

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement No 856716

Document Version Control			
Version	**Date**	**Change Made (and if appropriate reason for change)**	**Initials of Commentator(s) or Author(s)**
1.0	06/12/2021	Initial document structure and contents	MC
1.1	08/12/2021	Added content on WP1 and WP4	MC, PF
1.2	09/12/2021	Added content to WP3 and WP6	FP, BM, TV, FK
1.2	13/12/2021	Added content to WP5 and WP6	JD, JC
1.4	21/12/2021	Added content to WP2	WP
1.5	29/12/2021	Final corrections and arrangements for submission	RP

Document Quality Control			
Version QA	**Date**	**Comments (and if appropriate reason for change)**	**Initials of QA Person**
1.4	22/12/2021	Read through and edits	IS
1.4	26/12/2021	Reviews	AS

Document Authors and Quality Assurance Checks		
Author Initials	**Name of Author**	**Institution**
RP	Rui Prada	INESC-ID
MC	Marta Couto	INESC-ID
FP	Fernando Pastor	UPV
TV	Tanja Vos	UPV
BM	Beatriz Marín	UPV
FK	Fitsum Kifetew	FBK
PF	Pedro Fernandes	INESC-ID

JD	Joseph Davidson	GA
JC	Jeremy Cooke	GWE
IS	Ian Saunter	GWE
AS	Angelo Susi	FBK

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable presents the second project report (M15 - M27). WP1 is a management package, therefore, the deliverable focuses on management and the project's overall progress. Since this is a follow-up from D1.1, submitted in December 2020, some information provided here overlaps with the last deliverable as this document was created to be read as a stand-alone document.

1. OVERVIEW OF PROGRESS

1.1 OBJECTIVES

Iv4XR aims to build a novel verification and validation technology for Extended Reality (XR) systems based on AI techniques to provide learning and reasoning over a virtual world. With this technology, XR developers can deploy powerful test agents to automatically explore and test their virtual worlds' correct parameters as they iteratively develop and refine them. Given the importance of user experience for XR systems, we will also develop socio-emotional AI to enable test agents to conduct an automated assessment of User Experience.

1.2 EXPLANATION OF THE WORK CARRIED OUT BY WORK PACKAGE

1.2.1 Work Package 1

WP1 is responsible for all project management activities, covering the project's execution, monitoring progress, controlling the budget, managing risks, coordinating reviews, the consortium agreement, communication between the partners, and liaising with the EC. To facilitate this, we created a management structure at the first consortium meeting, which includes a management board (MB), the quality assurance (QA) team, and the Science and Technology Committee (S&T).

The Management Board meets during the consortium meetings and aids the project coordinator with administrative and financial tasks. So far, we have had five MB meetings: the first and second meetings took place in person in October 2019, Lisbon and January 2020 at Utrecht. The following meetings were virtual due to COVID-19 and took place in July 2020, February 2021 and June 2021.

The Science and Technology Committee focuses on R&D and executive decisions and meets once a month, using teleconference technology (on Skype), to discuss and solve possible conflicts or problems regarding technical issues, identify necessary changes to the direction in which the project is going concerning technical content development, and share progress. The QA team is responsible for ensuring that the project members deliver high-quality outputs in all project phases. All project deliverables are subject to quality control by at least one member of the QA team. We have a series of channels to facilitate communication between the project's members. We have a Slack channel for all members, mailing lists for all members, for the S&T Committee and the Management Board. To increase communication between the members, the project's coordinator also promoted the creation of working groups. Each WP coordinator sets up monthly meetings to discuss their progress and foster collaboration between the different partners. The meetings are open to all iv4XR members who want to join. The current pandemic imposed constraints on most partners. We requested an amendment to extend the project for three months. The amendment was approved in July 2020. All deliverables were submitted on time, taking into consideration the extension approved.

During this second year we worked to meet the second project milestone and reach an intermediate framework with testing agents with mature capabilities and running in the case studies.

WP1 coordinated the preparation of the first project review that happened online on September 17, 2021.

All deliverables for this second year were submitted on time. We asked for a 6-month extension on the market study deliverable (D6.5), and this has been approved by the PO. The reviewers asked for more details on two deliverables (D3.1 and D5.2). We are working on the new versions that will be submitted in the beginning of next year.

Plans for next year

During the third year, we will work towards our 3rd and 4th milestones. MS3 is the full iv4XR framework with all the promised features completed. As part of MS3 we will also have the evaluation study, using the pilots, almost complete, and the deliverables. MS3 was predicted for month 36, however, the extension means that the milestone has been pushed back to month 39. We will also work on MS4 which is an extension of MS3. By the 4th and final milestone, we will

have completed the evaluation and all deliverables will be submitted. We will coordinate the work to reply to the statements on the review report, received in November 2021, to guarantee the high quality of the final deliverables.

1.2.2 Work Package 2

WP2 is responsible for the construction of the infrastructure/core part of the iv4xr Framework. This includes, for example, basic agents, the agent runtime system, an information collection system, and an interface scheme with the XR system under test. In the first year we delivered the first prototype of the Framework (in D2.1). In the second year WP2 has focused its effort on improving the already implemented features and adding new features towards the end product that is to be delivered at the project's conclusion. Existing features are improved by evaluating them on various case studies (those provided by WP5 as well as various other internal lab case studies) and discussing them with other work packages and case study partners. This leads to e.g., extending the Domain Specific Language (DSL) of the Framework agent programming with new constructs or to the restructuring of the architecture of certain parts of the Framework.

An example of a new feature added in the second year is information collection. Advanced collection schemes have been added to allow not only test verdicts (pass/fail) to be collected, but also execution traces which can be subjected to post-mortem data analyses, or even to carry out reverse engineering of models. Another example of a new feature is multi-agency that would allow multiple test-agents to be deployed, for example to test the multi-user features of an XR system. An initial implementation of multi-agency was present in the first-year prototype, but it was tested and finalized in the 2nd year and is now ready for use.

A more detailed status of WP2 can be found in D2.2.

Plan for next year. To allow partners to work distributed across different tasks within different WPs, we allow tasks to work on separately managed sub-projects to develop their software modules. This has enabled task progress to proceed with minimal delays and without getting distracted by breaking experimental changes each task may commit (which would happen very often, as is common in a large software project). However, in the end, features produced by the tasks will have to be integrated into the Framework. The focus of the 3rd year is in preparing an architecture for such integration, and to support other WPs to integrate their features into the

Framework. For example, WP3 develops explorative and model-based testing modules, and WP4 develops emotional agents. In the 3rd year we will focus on integrating these modules into the Framework.

1.2.3 Work Package 3

Work Package 3 focuses on the development and integration of Functional Test Agents (FTAs) intended to verify if the behavior of the XR systems responds correctly according to the interaction of users. A second prototype of FTAs allows the navigation in the Systems Under Tests (SUT) of the WP5 use cases and the interaction with different WOM entities while verifying the correctness of the functionality. Document D3.3 - “2nd prototype of Functional Test Agents (FTAs)” includes an explanation of the second prototypes resulting from WP3.

The implementation of a Test Specification Language (TSL) in Task 3.1 allows developers to declaratively express a testing task. This specification language is used to indicate to iv4XR Functional Test Agents (FTAs) what entities and properties to test and which goal structure to follow to achieve the testing goal objective (see D3.1 - “Test Specification Language” and D3.3, task 3.1).

In the iv4xr framework we mainly distinguish two types of FTAs (Task 3.2):

1. FTA focused on SUT Goal Solving
   These agents use the TSL developed in Task 3.1 to take deliberations about which strategies will allow them to navigate and interact with different entities to solve their goals.

2. FTA focused on exploration of the SUT
   This type of agent is intended to explore and test the robustness of the SUT. Decision making that executes non-sequential actions allows detecting abnormal and unexpected SUT behaviors. Model-based Testing and Reinforcement learning can be used to explore in a more guided way.

XR systems may contain hazardous entities (Task 3.3) that can sabotage an agent’s progress or even cause it to fall into an inescapable stuck state. Solving goals to test specific tasks, means that dealing with hazardous elements (Task 3.3) will not be very different from dealing with other XR application elements. Consequently, the goal solving FTA can already handle these

hazardous elements which completes Task 3.3 and its inclusion in the FTA prototype. This task has finished.

Task 3.4 is intended to measure how good FTAs are in exercising the SUT and to develop test generation strategies that increase the coverage obtained from the tests. The notion of coverage is defined on models that abstractly represent the range of possible interactions to be carried out in the SUT.

Based on these models, state and transition coverage is used to measure the space of interactions reached by the FTAs in the SUT.

The development of a model-based testing (MBT) tool that employs different search-based heuristics, allows the generation of test cases with TSL tactics that FTAs use to cover all possible targets in a given model of the SUT.

The iv4XR framework has been extended with a new module that allows the execution and communication of multiple agents (Task 3.5). This enables verification of the correct interaction of multiple users in the same environment and improves the entity-search and exploration performance by coordinating a group of agents.

In addition to the new multi-agent module, task 3.5 is also investigating the parallel execution of multiple FTA instances that share knowledge of the SUT to speed up the RL training procedure and the exploratory process.

Finally, Task 3.6 exposes the overall integration of all WP3 tasks together with the WP2 core framework and WP5 use-cases integration.

Plans for next year

Over the next year, the integration of FTAs on the XR systems of industrial partners will begin. Also, it is planned to research and develop a second prototype that will include:

- Reinforcement learning (RL) algorithms and strategies to diversify executions on existing FTAs. We will consider RL rewards that are general and can improve exploration as well as domain specific RL rewards for optimal goal solving (This is conform with Task 3.2).
- Evaluate coverage effectiveness on iv4XR use cases

- Finish multi agent support in FTAs tools for iv4XR use cases
- Work on a final integration of all the components from WP2, WP3 and WP4.

1.2.4 Work Package 4

Work Package 4 focuses on developing socio-emotional test agents (SETAs) to aid the systematic assessment of user experience of XR systems while minimizing the manual effort.
We currently have two implementations of the SETA integrated with the framework (task 4.1), one based on the two-dimensional core affect theory, modelling the pleasure and arousal emotional dimensions (PAD), and one based on the OCC theory of emotion. We have decided to pursue these two different implementations as each has its own advantages for different testing and developmental scenarios. The OCC model is better adapted to event-based systems whereas the PAD dimensional model is better adapted to continuously evolving systems, as well as being a better fit for machine learning models. This implementation will be improved and validated in the following months through user-testing.

Other than the two emotional models, WP4 has studies in progress to build intelligent agents capable of simulating other components of user-experience during interaction with XR systems, including:

1. Testing a model to predict the cognitive load imposed by a game. This is the first step towards developing a model that can tell designers the amount of cognitive load that their systems are imposing on the user.

2. The difficulty progression of a series of levels can have a significant impact on the user experience and learnability of a game. We are developing methods to rank a series of levels in terms of the difficulty.

3. Exploring the use of clustering techniques to classify game-play traces in terms of play-style personality and using Inverse Reinforcement Learning to create agents that play according to those different play-style personas.

Two implementations of the SETA have already been integrated with the iv4XR Framework (Task 4.5) and used to test several maps of the Lab Recruits game. Further details on this integration are provided in document D4.2- “2nd prototype of SETAs” deliverable.

The current pandemic imposed serious constraints to our user studies, namely those that required the collection of physiological data. We conducted user studies online to inform the emotional model, the personas, and the cognitive load model. These constraints have delayed the work of task 4.2, which is nonetheless near completion as can be seen in D4.2 - 2nd prototype of SETAs. The task should be completed in the coming couple of months.

Both emotional models being used for the SETAs, the OCC and PAD, have been implemented with temporal progression in mind, which means Task 4.4 already has a solid foundation and is near completion. Effort from this task can thus be directed to highlight the progression perspective of the testing modules.

A definition of coverage for UX is being pursued (task 4.3), which will guide the development of self-motivated SETA. The work on task 4.2 has already created agents with parameterized motivations, which will be a bedframe for the work of task 4.3. Details of said agents can be found in document D4.2- “2nd prototype of SETAs” deliverable.

We have taken efforts to mitigate further impacts of the pandemic by adapting our user studies and are confident all work of WP4 will be finished by the end of the project deadline.

WP4 meets on the last Monday of the month to discuss the socio-emotional test agents (SETA). Partners present the ongoing work and share research plans or results, which are then discussed with the attendees (these discussions have been linked to tasks 4.1 and 4.2). The second part of the meeting is usually about test coverage and what it means for SETA (affective coverage - task 4.3). These discussions are fruitful not only for the project development but also because they allow students to get more feedback on their work.

Plans for next year

For next year we will focus on integrating the different modules into a User Experience (UX) model. We will continue to the last integration phase (Task 4.5) where we integrate software

modules developed in this WP into the iv4XR Framework from WP2. We will also bundle and document the Application Programming Interfaces (APIs) of these modules to become part of iv4XR SDK.
We are also working towards submitting 2 journal articles, an article detailing our first Cognitive Load experiment and a second one as a position paper detailing our UX model.

1.2.5 Work Package 5

Work Package 5 focuses on the preparation of pilot programs by the industrial partners in order to showcase how iv4XR can be used. These pilots are a 3D game, a simulation scenario depicting the infiltration of a nuclear power plant, and a sensor suite for the monitoring of civil engineering projects. In this second year, both the second and final stages of the integration between the framework and the pilots were completed.

Task 5.1 - “Preparing the Pilots” concerns the preparation of the pilots - this effort is primarily reflected in the progress towards deliverables D5.1 - D5.3 reporting on the integration of the pilots with the framework. In this period, both of the deliverables have been achieved which means that the integration procedure is complete and that each pilot supports the connection and actions of an external testing agent. The interfaces between the pilots and the iv4XR framework also supports the querying of properties of the pilots to allow for the testing framework to verify the results of agent interaction. The interfaces are publicly available on the project’s github page[1].

Task 5.2 - “Pilot Environment” is about building the testing scaffold for the pilots so that an iv4XR test suite can be executed with minimal human intervention. As of the full integration deliverable, the interfaces allow agents to interact with the pilots, and the pilot environment can wrap this process by loading scenarios and verifying testing results.

Task 5.3 - “Demos” is about packaging the pilots into a form suitable for the public. The interfaces and environments for each pilot have been published in repositories linked to the project as free software. If an interested party owns Space Engineers, both the interface and tests are available to them to install and use. LabRecruits, developed by UU is also a pilot system for the framework which is wholly free and modifiable by those who wish to experiment.

[1] https://github.com/iv4xr-project

Plans for next year

This year marked the completion of the majority of the technical work on connecting the iv4xr framework with the pilots. Although the interfaces are “feature complete” as far as functional testing is concerned, further work may be needed to implement more high level control of the pilots (e.g. navigation via a dynamic nav mesh for Space Engineers) to reduce the amount of boilerplate behaviour required for the agents. As the agents of WP4 get more advanced, the interface developers will add more functionality as needed to help the SETA agents reach a level of usefulness to the pilot developers.

One of the industrial partners (GA) is currently in discussion with the developers of Space Engineers in order to integrate the iv4XR results into the current testing workflow. This means that the interface will be adjusted, and agents developed to implement test cases which the developers currently perform manually.

One of the final efforts on WP5 is the execution of pilot studies that assess the potential impact of the tools developed on the project.

1.2.6 Work Package 6

The tasks in WP6 have been executed according to the original plan:

- The website is in place https://iv4xr-project.eu/ (D6.1)
- The data management plan has been delivered (D6.2). It includes an assessment of the various data generated and maintained in the project by all the partners and outlined the project’s management plan with respect to the storage and dissemination of the various data produced in the scope of the project.
- A number of dissemination and communication activities (Task 6.1) have been undertaken in the second year, as planned in D6.3:
    - Organized a public workshop on the Future of XR, in collaboration with the ARETE project.
    - Organized the 12th A-TEST workshop on automated testing, as planned in the dissemination and exploitation plan.

    - Participated in the VRDays Europe Immersive Tech Week event and presented ongoing work.
- The 1st and the 2nd Dissemination and exploitation plan has been delivered (D6.3 and D6.4). The 1st Dissemination and exploitation plan describes more detailed plans for dissemination and the results for the first year for Tasks 6.1, 6.2 and 6.3. The 2nd Dissemination and exploitation plan updates the information of the first plan and provides details of the results obtained in the first and second year for Tasks 6.1, 6.2 and 6.3. Moreover, we start working in exploitation.

- To deal we some delays in the exploitation activities and plans, T6.2 leadership has been assigned to GWE and fortnightly conference calls have been instigated to review progress within the task towards D6.5. In order to facilitate the market research requirement a questionnaire has been created and shared with agreed key XR market sectors: Aviation & Automotive, Education & Training, Games, Construction, Software Design & Testing, Healthcare, Retail, Manufacturing & Logistics. More than 200 surveys have been sent out to targeted individuals and to date 30 responses have been received. These qualified respondents input data will be assessed to inform market opportunity via D6.5 (Market research report).

Plans for next year

For next year, the task will continue to carry out dissemination and communication activities as outlined in D6.3 and we will continue working on exploitation. One of the important tasks next year will be the organization of the second project workshop.

1.2.6 Work Package 7

Work Package 7 has been included to address ethical questions and guideline development of the project. Its main goal was to develop deliverable D7.1 - “POPD - Requirement No. 1” and assure that its principles and guidelines are followed.

We created the Ethics Advisory Board (POPD requirement No.1) to maintain information on all studies conducted with humans within the project. The Ethics Advisory Board (EAB) is responsible for ensuring that all studies that involve humans are submitted to an Ethical Review Committee

(ERC) for approval. The EAB will also help the partners compile the required materials for ERC submissions, inform all partners about the decisions, and help with any modifications requested by the ERC's. This Advisory Board will also be tasked with assisting all partners in complying with the principles defined in D7.1. The user studies performed so far were submitted to the a ERC.

2. Financial Summary

The tables below show the personnel and direct costs from M1 to M21 as the next reporting period covers months 22 to 29.

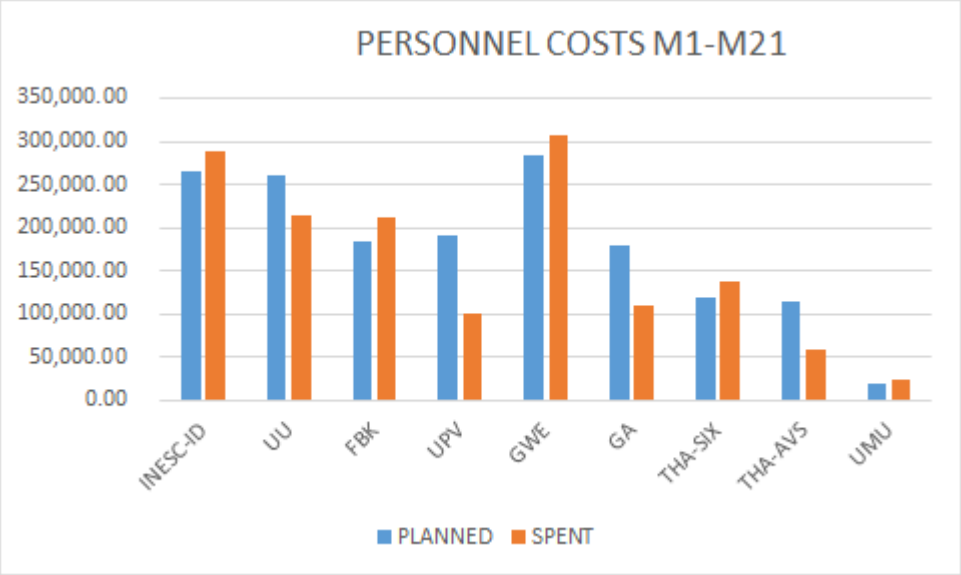

Figure 1 – Personnel costs: planned vs spent.

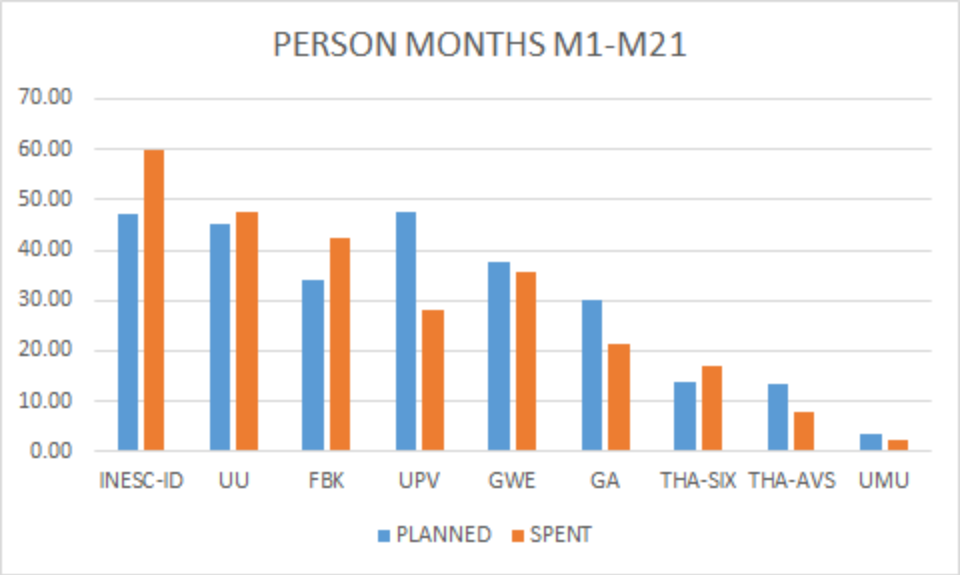

Figure 2 – Effort reported by partners: planned vs spent.

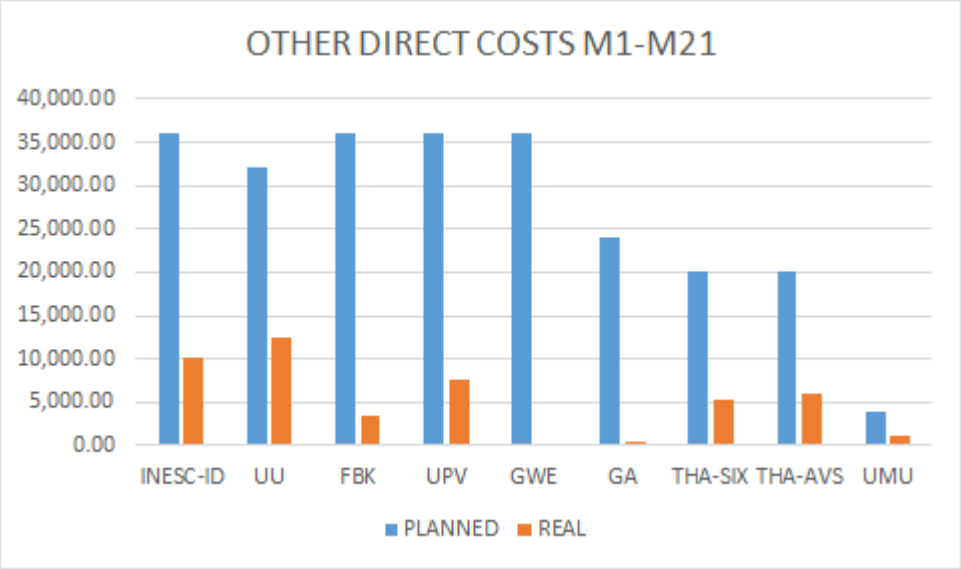

Figure 3 – Other direct costs: planned vs spent.

Overall, the actual costs are lower than the estimation. This difference arises from the fact that one project meeting was held online due to the current pandemic. Participation in conferences and workshops was also moved online, which decreased a lot of travel costs. Aside from these minor issues, the budget is on track, and we do not have, or expect, significant deviations to report. Eventual deviations will be detailed in the next period report.

3. CONCLUSIONS

This deliverable provides an overview of the work carried out in each work package throughout the second year. The project is in general on track and all deliverables have been submitted on time. For the next period we will address and accommodate the reviewers comments.

The focus on the next period will be to finalise the software packages and to produce materials and evidence that support the quality of the outputs of the project.

The outputs from the project can be found on Zenodo[2], Github[3], Project Website[4] and on Social Media: Twitter[5] and Facebook[6].

[2] https://zenodo.org/communities/iv4xr-project/?page=1&size=20
[3] https://github.com/iv4xr-project
[4] https://iv4xr-project.eu/
[5] https://twitter.com/iv4xr
[6] https://www.facebook.com/iv4xr