

FINAL EVALUATION REPORT

Work package number: 1

Deliverable number: 1.6

Contributors: Dr Kutoma Wakunuma and Mr Tilimbe Jiya

Institutions: De Montfort University

Revision Date: 30/07/2018

Status: Final

Executive Summary	3
Introduction	3
The Evaluation Design	4
Evaluation Findings	4
Partner Interview on Implementation of Toolkit – WP5	5
Successes and challenges of toolkit implementation	5
An assessment of the chosen module activities	6
Input from teenagers regarding toolkit’ activities	7
Outcomes of the toolkit activities implementation	7
Third Party Interviews on Toolkit Implementation - Hubs	7
The successes of the facilitation of toolkit activities.	8
The challenges of the facilitation of toolkit activities.	8
Level of third parties involvement in choosing the toolkit activities	8
Input from teenagers in the toolkit activities	9
Facilitation of the toolkit activities.	9
Observations from the Third Parties	9
Contexts of the observations	9
Focus of the observations	13
Intended outcome(s) of the assigned module	13
Indicator of success of the intended outcome of the modules	14
Effectiveness of the Toolkit guidelines	16
Challenges in ensuring gender inclusiveness during the implementation	17
Level of gender inclusiveness among teenagers on the four different perspectives	18
Individual level	18
Interactional level	19
Institutional level	20
Societal/Cultural level	22
Interviews conducted by Third parties	23
Usefulness of the guidelines on facilitation and gender inclusiveness	24
Risks and challenges encountered while implementing the modules	24
Strategies which were used to implement the activities	27
Improvements that can be made to the modules for future use	29
Conclusion	29
Annexes	32
Annex 1: Evaluation Strategy Document	32
Annex 2: Hypatia Evaluation Interim Report	48
Annex 3: WP5 Interview Questions on implementation of the Toolkit	79
Annex 4: Third party interviews - December, 2017	80
Annex 5: Observation and interview instructions	81

Table 1: A summary of the allocated context per Third party and their respective activities ... 12

Table 2: A summary of the intended outcomes for the modules implemented in different countries 14

Table 3: Module objectives and indicators of success 15

Table 4: Responses from stakeholders that were given to Third parties 26

Executive Summary

In this final evaluation report, the Evaluation team presents interview findings and observations outcomes that were conducted during the implementation phase of the Hypatia project to assess how this process has worked. In total there were 10 interviews conducted : 1 with the WP leader of WP5 - Experimentarium, 4 were conducted between evaluators and third parties from Estonia, Serbia, Spain and Sweden and the last 5 were conducted by third parties themselves with facilitators of modules from Ireland, Austria, United Kingdom, Poland and Greece. Observations were done by third parties from all the nine countries listed in the preceding section in their respective countries. The evaluation interview with the Work Package Leader (WP) 5 responsible for Toolkit implementation was conducted in Month 25 of the project. Furthermore, Third party interviews were conducted to examine the different aspects of the adaptation process of the Hypatia Modules as well as the implementation of the toolkit in these countries. In addition, information is provided in this document based on the detailed findings and the analysis of the observations that were conducted by the Third Parties and served to evaluate the usability of the Hypatia toolkit, its effectiveness and the engagement of teenagers with the toolkit activities. Towards the end of this final evaluation report, de Montfort University provides recommendations based on the evaluation findings.

Introduction

This final evaluation report of the Hypatia project focuses on the implementation of the toolkit modules. It builds on the initial strategy document (see Annex 1) and the interim report which was focused on the development of the toolkit activities (see Annex 2). The evaluation of the Hypatia project, was focused on aspects related to the usability of the toolkit, usefulness of the toolkit on the way science is communicated, participation and interest of teenagers in STEM, gender stereotypes and stakeholders' engagement with the toolkit. Specifically, third parties were asked to look at activities within allocated contexts. For example, Ireland, Poland and Serbia looked at activities in a Museum context while Austria, Greece and Spain looked at activities in a School context. Lastly, the United Kingdom, Estonia and Sweden looked at modules implemented within the Industry context. The allocation of observation context was done in order to have a fair overview of how the modules were being implemented across different parts of Europe.

In the final phase of our evaluation of the Hypatia project, we used interviews and observations to evaluate the progress of the project towards its aim. There were three interviews that were conducted. The first interview was conducted by the evaluator with the leader of Work Package WP5 who was responsible for toolkit implementation to assess and understand the following:

- i. the successes and challenges of toolkit implementation

- ii. the choice of the toolkit activities
- iii. the input from teenagers regarding toolkit' activities
- iv. the major outcomes of the toolkit activities implementation

For detailed interview questions with the WP5 leader please see Annex 3.

The second set of interviews were conducted by the evaluator with four Third parties from the following countries: Estonia, Serbia, Spain and Sweden. The interviews were conducted via Skype. The aim of this second set of interviews was to understand;

- i. the successes and challenges of the implementation of toolkit activities
- ii. the level of Third parties' involvement in choosing the toolkit activities
- iii. the input from teenagers in the toolkit activities
- iv. whether the third parties were happy with the implementation of the toolkit activities

For detailed interview questions with the Third parties please see Annex 4.

The third interviews were conducted by the 5 Third parties themselves with stakeholders /members of their Hubs who participated in the implementation of the toolkit. The questions for these interviews were outlined in the evaluation guidelines (see Annex 5) that were provided to the Third parties by the evaluators. The second and third sets of interviews were part of the observations that were conducted with support from the Third parties. The observations were aimed at understanding firstly, how teenagers were engaging with STEM activities and secondly, how hubs were supporting the toolkit implementation and its subsequent adoption.

The outcomes of all these interviews and observations are presented in the section 0, 0 and 0.

The Evaluation Design

The evaluation design consisted of three semi-structured interviews. Besides interviews observations took place by the Third parties themselves following the evaluation guidelines proposed in the evaluation strategy by de Montfort University (see Annex 1).

As there was a budget constraint, partners and third parties played a key role in implementing the evaluation design in their respective countries. For consistency in the evaluation process, all the nine Third parties were given evaluation guidelines.

Evaluation Findings

Following the interim evaluation in January 2017, Evaluators conducted further evaluation activities between April 2017 and January 2018. The evaluation during the period indicated

included an interview with a WP leader, observations and interviews with Third parties. Thus the evaluation findings following the evaluation are presented. Firstly, in Section 0 we present the results from the interview that we conducted with WP5 leader. Secondly, in Section 0 we present the results from interviews that we the evaluators conducted with four Third parties that were purposively selected for the interviews. The four Third parties were from Estonia, Serbia, Spain and Sweden. Lastly, in Section 0, we present the results from the observations conducted by the Third parties. The results from the observations include findings from the observations on the toolkit implementation in nine countries across Europe including Ireland, Austria, UK, Poland, Greece, Estonia, Serbia, Spain and Sweden. As part of the observations, five out of the nine Third parties conducted additionally interviews with relevant stakeholders to gain an in depth understanding of the toolkit implementation process. These included Third parties from Ireland, Austria, UK, Poland and Greece.

Partner Interview on Implementation of Toolkit – WP5

WP5 involved the national selection and adaptation of modules as well as the facilitation of national seminars involving different stakeholders such as teachers and head teachers working in science. As indicated in the introduction section, to understand the process of the implementation of the toolkit, specifically towards the aim of the WP, the interview focussed on the following aspects:

Successes and challenges of toolkit implementation

From the interview several successes and challenges were mentioned. The successes are as follows:

- i. Experimentarium mentioned that the gender criteria developed by UCPH were one of the biggest successes of the project as they provided the necessary theoretical background to the project work (adaptation of the toolkit activities following a common background for all activities) .
- ii. Teachers were able to approach and observe both girls and boys who were engaged with different/varied set of activities, take for example one of the Modules that included a robotics programming activity based on the four different levels of gender inclusiveness.
- iii. Role model speakers were involved and engaged in some of the toolkit activities particularly with respect to technology. In addition, the involvement of role models in some activities gave the opportunity to better discuss and debunk stereotypes in a more gender inclusive way.

- iv. There was positive feedback coming from different stakeholders/hub members who were involved in the process of adaptation in each country contributing to making changes to the assigned activities so that they would become more gender inclusive and relevant for the context/country.
- v. As a result of the vast experiences and prior knowledge of the partners in the field of learning, the modules chosen for the toolkit were deemed to be relevant to the three contexts of schools, museums and industry.

Despite the above successes, there were challenges during the implementation of the toolkit which included the following:

- i. Some stakeholders found that some of the suggested gender inclusive activities were not relevant for their own specific context or that they were difficult to understand.
- ii. One other challenge was that of the management of activities particularly with respect to language. Although Third Parties managed to change the activities to suit their national contexts and translate them the process of adaptation and translation was not an easy task for some of them.
- iii. One other challenge was management related and referred to the delay in sharing of information among project partners as some countries had summer holidays in July and August. However, the challenge was addressed by extra reporting which was intended to provide the necessary information in advance rather than wait for the summer holidays to end.
- iv. In some countries the school system was strict. In those countries, it was difficult for teachers to choose any teaching method they wanted without explaining the purpose for their choice making the implementation of the Hypatia modules in these countries more time demanding.

An assessment of the chosen module activities

The WP5 leader was also interviewed looking at the aspect of the relevance of the module activities that each partner had chosen. The chosen module activities were distributed across museums, schools and industry context for implementation. To have a fair and wide spread distribution of the type of module activities in each country, every Third party was assigned an equal number of different modules activities to implement having to respect a minimum set of criteria (for example each third party had to have modules fitting all three contexts).

The modules that were chosen were expected to be relevant because of the expert knowledge and experience of the partners that were involved in their choices. Partners could and were encouraged to consolidate in this selection process their Hub members. However, it was noticed

during the process of adaptation that some of the modules were not entirely well-suited in some contexts in certain countries. As a result, the partners who were assigned for their implementation in those countries had to adapt them even further than they originally thought. Furthermore, there was in some cases a low level of collaboration and interaction among the teenagers themselves when conducting certain Modules particularly when they were carrying out activities that focused more on discussing gender issues than activities that had the scientific activity in the centre and the gender aspect integrated into them (see for example difference among activity Find Gender Stereotypes in STEM representations versus types of activities such as Inquiry: Shape and Action.)

Input from teenagers regarding toolkit' activities

With regards to the input from the teenagers, it was established that teenagers played a key role in the development/adaptation and implementation of the toolkit modules. Once the activities were chosen by the partners (that in many cases included their youth panels even in the selection process), teenagers were invited to test and give further feedback. There was a feedback system that was put in place to ensure that the chosen module activities were implemented appropriately. For example, teenagers were given an opportunity to give feedback on whether the teachers had applied gender inclusivity in their teaching or not.

Outcomes of the toolkit activities implementation

Several outcomes were realised as a result of the toolkit implementation, Firstly, there was an increased interest from the teenagers towards different STEM subjects. As the result of toolkit implementation activities, WP 5 leader mentioned that young people started showing interest in taking STEM subjects. However, she mentioned that the change in attitudes was not so huge in Denmark but then with time there is an expectation that the interest will further increase.

Secondly, there was a positive behaviour change among all the Hub members involved in the partner countries. Members were very open to give suggestions on how the activities could be improved in the future, particularly towards gender inclusiveness and how they can effectively support their implementation.

To supplement the interview with the WP partner on the toolkit implementation, the next section presents a further analysis on the interviews that were conducted with Third parties.

Third Party Interviews on Toolkit Implementation - Hubs

This section presents findings from the interviews on toolkit implementation with Third parties. The interviews were conducted by the evaluators with four Third parties from the following countries: Estonia, Serbia, Spain and Sweden. The main purpose of these interviews was to examine the different aspects of the toolkit implementation process which included the following:

The successes of the facilitation of toolkit activities.

In terms of successes of toolkit activities implementation, third parties showed several responses during interviews on the level of success. On the one hand, Sweden and Serbia chose without any issue museums, schools and research institutions for implementing the toolkit activities. In the case of Serbia, it was indicated that the activities helped them especially in their teacher trainings. While in Sweden, the activities were extremely interesting particularly in schools because of the lively discussions that took place and the positive feedback they received from teenagers.

On the other hand, in Estonia and Spain the implementation of toolkit activities faced some challenges. For instance, in Estonia, AHHA considered the implementation of the toolkit activities as challenging in the beginning because the activities were not considered by some stakeholders in their country as novel enough. Therefore, less institutions carried in the beginning of the implementation activities than they originally expected. In Spain, the process of toolkit implementation was delayed by external reasons such as school summer holidays. However, despite these setbacks, the process in both countries later picked up and in the end the implementation was fruitful and impact numbers were achieved.

The challenges of the facilitation of toolkit activities.

In terms of challenges, the third parties from the four different countries shared these different aspects:

- i. The Third party from Estonia indicated that the biggest challenge for them was to get on board institutions who had not dealt with gender equality issues in STEM before. They also added that the challenge was to get them to develop their interest in trying out the activities and getting their feedback.
- ii. The Third party from Sweden mentioned that there was a lot of text in the toolkit which made it challenging to get people or companies to read all the texts that were included in the toolkit.
- iii. The Third party from Serbia mentioned that it was challenging to initiate contact with the big companies in order to arrange workshops and cooperation with them.
- iv. The Third party from Spain stated that engaging teenagers in the implementation of toolkit activities was in some cases challenging.

Level of third parties involvement in choosing the toolkit activities

Third parties played a vital role in choosing the toolkit activities. They had complete freedom to participate in the selection of the activities. The Third parties were very excited about their participation because this gave them the opportunity to choose activities which were most appealing to them. All the Third parties said that they piloted the activities before implementing them. In so doing, they had the opportunity to discuss with hub members and teenagers in order

to get their opinions and ideas about the activities before making a final choice on the most suitable activities.

Input from teenagers in the toolkit activities

Similar to what the WP 5 leader mentioned in section 0, the Third parties also said that there was very good input from the teenagers. In particular, teenagers from Sweden and Spain played a role via youth panels while testing some activities. Their input was taken into consideration and used to improve the activities.

In other countries like Estonia, teenagers had a lot of input towards the customisation of the activities. While in Serbia, teenagers provided feedback during and after the implementation activities. The feedback obtained was useful in fine-tuning the activities. This also helped the third parties to know which parts of the activities the teenagers found easy and which ones they found difficult.

Facilitation of the toolkit activities.

Some third parties from countries like Serbia, Spain and Sweden mentioned that they were very happy with the facilitation of the toolkit activities. According to them, activities were very interesting and rewarding while working with teenagers especially in schools. After the implementation of activities in particular school classes, students from other classes also wanted to arrange those workshops for them which was a very positive response.

However, AHHA from Estonia indicated that they were not as satisfied with the facilitation of the activities because of low levels of participation from teenagers participating in certain activities. The Third party also mentioned that as the original design of activities was done by the main partners, certain aspects remained irrelevant to them.

Observations from the Third Parties

In this section we present the results from the observations that were conducted by the Hypatia Third parties. The observations were done during the toolkit implementation in 9 countries across Europe that included Ireland, Austria, UK, Poland, Greece, Estonia, Serbia, Spain and Sweden. Observations were important in the evaluation because they enabled the development of an understanding of how teenagers were engaging with the activities. In addition, the observations were important in understanding how institutional hubs supported the toolkit implementation.

Contexts of the observations

There were three contexts in which the 9 Third parties conducted at least one observation on the implementation of the toolkit activities. The three contexts were museums, schools and

industry. The museum context included Third parties from Ireland, Poland and Serbia. The school context included Austria, Greece and Spain while the Industry context included Estonia, Sweden and the UK. Table 1 below provides a summary of the allocated context per Third party and their respective activities that emerged from the observations.

Country	Observation Insights
Museum Context	
Ireland	<ul style="list-style-type: none"> • Students were involved in a Kahoot quiz to examine their prior knowledge of stereotypes, differences between sex and gender, the fact that gender is non-binary, and some stats on STEM participation in Ireland. • Mostly students answered correctly, and it gave the facilitators an insight into their existing knowledge in advance of the activity, and also made students aware of stereotypes and bias in advance.
Poland	<ul style="list-style-type: none"> • It was evident that girls participating in the activity do not even consider studying STEM subjects because in their opinion they are only for male students.
Serbia	<ul style="list-style-type: none"> • The discussion about stereotypes in STEM was open and it involved the majority of the participants. There were some divided opinions. However, after the discussion, all participants appeared to recognize gender stereotypes and biases in STEM, and repercussions of those problems
School Context	
Austria	<ul style="list-style-type: none"> • A lesson with a special focus on gender was observed. The teacher chose the module “STEM Women Cooperative Card Game” for the lesson. The students played a memory version of the game in a competitive manner and students were then asked to place the cards in historical order. This was followed by a discussion on aspects of language or Gender-neutral material in the STEM subjects.
Greece	<ul style="list-style-type: none"> • At the beginning of the activity, participation was limited to the pupils who could confidently express their opinion. Progressively, even the pupils who seemed uninterested at first, wanted to speak up. The teacher managed to encourage the students to reflect on their views despite some pupils’ views being persistent compared to others.
Spain	<ul style="list-style-type: none"> • An activity called “Science café” was carried out where female scientists were presented to the students. This gave the students the opportunity to interact with role models who talked about real experiences and daily work with a hope to help them in deciding their professional career.
Industry Context	

Estonia	<ul style="list-style-type: none"> In the University of Tartu Institute Of Chemistry (Chemicum), the researchers introduced their specialities such as electrochemistry and energetics, by giving examples of how the research is done. They further explained how materials that are studied there are then used by the industry. In addition, all 5 researchers (of whom 3 were female and 2 male) were very engaged with the teenagers and they stressed on the importance of studying STEM subjects in making career choices. They further stated that at least in the field of chemistry, there is an expectation of a 50-50 gender balance.
Sweden	<ul style="list-style-type: none"> Different examples were given about jobs and careers in the Chemistry business. The scientists who were implementing the activities mentioned that in chemical business the gender balance is almost 50-50. In addition, good examples were given to encourage both girls and boys to become scientists
UK	<ul style="list-style-type: none"> A 50:50 ratio of male and female students met a wide variety of gender balanced STEM professionals from different stages in career development. The STEM professionals were provided with best practice guidelines to ensure they are considering their use of language, using appropriate and relevant examples, talking about work/life balance & the purpose behind their work (etc.) in improving gender inclusivity in their communication with students. Male role models were encouraged to talk about their female work colleagues. Overall this provided inspiration and careers advice to students who may otherwise not consider a career in STEM.

Table 1: A summary of the allocated context per Third party and their respective activities

Focus of the observations

The Third parties used a set of guidelines to conduct the observations (see Appendix). The observations focussed on the following aspects; the intended outcomes of the modules, indication of achieving the intended outcomes, effectiveness of the toolkit guidelines and the level of gender inclusiveness among the four different perspectives.

Third party	Module that was observed
Ireland	<u>Wearable technology</u>
Austria	<u>STEM Women Cooperative Card Game</u>
Estonia	<u>Speed Dating: Encounters between Pupils and Scientists & Engineers</u>
Poland	<u>Wearable Technologies</u>
Serbia	<u>Find Gender stereotypes in STEM representations</u>
Spain	<u>Science Cafe</u>
Sweden	<u>Science ambassadors and ambassadors</u>
UK	<u>Speed dating: Encounters between Pupils and Scientists & Engineers</u>
Greece	<u>What is your opinion?</u>

Table : List of Selected Modules per Third Party

Intended outcome(s) of the assigned module

From the responses from the Third parties, it was found out that the intended outcomes of the module varied according to the countries in which they were being implemented. Table 2 presents a summary of the results on the intended outcomes of the assigned module

Third party's Country	Intended Outcome(s)
Ireland	<ul style="list-style-type: none"> To facilitate bonding between the participants through sharing of tools, materials and ideas.
Austria	<ul style="list-style-type: none"> To get students familiar with famous female scientists. Teenagers were expected to learn more about their achievements, their field of study and their life in general. To raise the question, why famous female scientists were not really known in contrast to their male counterparts. The model offered the opportunity to start an open discussion about gender and the way that our social norms influence our career choices.
Estonia	<ul style="list-style-type: none"> To make young people more aware of the STEM careers available (both locally and internationally).

	<ul style="list-style-type: none"> To give them (especially girls) the opportunity to meet and talk to inspiring female role models so that they would be more willing to consider STEM fields as a potential career path.
Poland	<ul style="list-style-type: none"> To encourage more girls to choose the studies that really interest them rather being coerced by the society to choose studies relating to humanities
Serbia	<ul style="list-style-type: none"> To be aware of suppression of women through history of science and its consequences gender stereotypes in science in present To make female scientists the role models of the students so female teenagers could relate to them
Spain	<ul style="list-style-type: none"> Get to know real and current problems related to gender Debate about gender stereotypes and how to overcome them Bring closer to the students the real work of scientists
Sweden	<ul style="list-style-type: none"> To show the teenagers the connection between the 'real world' of engineering and what they learn at school. To make the teenagers aware of the career opportunities for both teenage girls and boys through testimonial from men and women working in industry
UK	<ul style="list-style-type: none"> Increasing exposure of young people to the many different career opportunities available to them and to many different role models. to raise their confidence and aspirations and allow them to personally interact with STEM professionals with the hope that they will meet someone like them and dispel misconceptions about job opportunities or their own suitability for a career in STEM.

Table 2: A summary of the intended outcomes for the modules implemented in different countries

Indicator of success of the intended outcome of the modules

The Third parties were asked to give an indication of whether they had achieved the intended outcome of their respective modules. To ensure that the observations were relevant, each category of Third Party countries was asked to focus on a specific objective of the Hypatia project. Using the indicators of success, the Third Parties were then able to observe and assess if these were met as shown in Table 3 below.

Third Party Country	Context	Objective	Indicator(s) of success	Outcome (Y/N)
<ul style="list-style-type: none"> • Ireland • Poland • Serbia 	Museum	Understand what sex and gender stereotypes are and identify them	Evidence of an understanding of gender stereotypes e.g. Engineering jobs are mainly for boys	Y Y Y
<ul style="list-style-type: none"> • Austria • Greece • Spain 	School	Increase awareness about the negative impact they can have on their own representations of science, the world of science and technology, and their study/career choice	Evidence of reflection on more positive ways of representing STEM in Schools e.g. Gender neutral STEM materials such language, role models in STEM	Y Y Y
<ul style="list-style-type: none"> • UK • Estonia • Sweden 	Industry	Learn about careers in STEM and develop an interest in them, regardless of their sex.	Evidence of how industry has/is helping with the promotion of STEM careers for both boys and girls e.g giving talks and presentations about STEM careers for both girls and boys	Y Y N/A

Table 3: Module objectives and indicators of success

From the Table above, almost all the Third Parties indicated that the intended outcomes of the engagement activities were achieved. The exception was Sweden. This could be that the objective and indicator of success that the Third Party was assigned to observe was not appropriate for the specific activity.

Effectiveness of the Toolkit guidelines

Third parties were also asked to indicate whether the instructions of the modules were effective as far as facilitation and gender inclusion was concerned.

Ireland - Science Gallery

Science Gallery stated that there were no challenges. In addition, the Irish stated that the guidelines made the facilitators more aware of the issues in advance of the workshop than they would not have been otherwise.

Austria - Science Center Network

In Austria, it was mentioned that they had a few difficulties with the guidelines. For instance, they indicated that there were issues with formatting of the guidelines that referred to the module of the Cooperative Card Game. The dates on the cards were not correctly formatted which affected how the students engaged with the content on the cards. In addition, there was a problem with handling of the rules for the card game particularly when the students were playing against each other.

Estonia - AHHA

In Estonia, the institution (Chemicum) that implemented the module did not seem to have any particular problems with the implementation of the module, but they insisted on doing it “their way”. For example, rather than have a whole group of youngsters, 34 in total in one room, they divided them into 5 groups and put them in separate labs. This proved effective in ensuring that they worked well together and were attentive to each other’s contributions. In addition, Chemicum skipped the Kahoot game because they believed that the discussion would spring up naturally and the youngsters could give feedback on the go. Further, the professionals found the guidelines on gender inclusion useful in giving examples of the gender balance issues.

Greece - NOESIS

In Greece, the guidelines were also helpful especially with respect to time limits which seemed to work well together. However, there was a slight challenge with keeping up with the timeline for the implementation of the activities. Another challenge was related to the rephrasing of the questions in the guidelines so that they would make sense in Greek.

UK - ASDC

In the UK the observation was that the guidelines were long for certain activities. It was noted that STEM professionals were busy and may not have read through all the guidelines.

Spain - CAIXA

In Spain, there were a few challenges with the guidelines. Firstly, it was difficult to engage scientists in the activity but Caixa managed to find a solution with the help from ALBA Synchrotron, a member of the Spanish hub, who volunteered to facilitate female scientists to implement the activity. Secondly, it was not clear from the guidelines who was supposed to prepare the introduction to the activity. Thirdly, it was also not clear which music was supposed to be used during the break. Some options (such as YouTube or Spotify lists) could have been made available through the project website for common use. Lastly, there was a lack of a previous contextualisation on gender and STEM or on the activity to be implemented which would have helped to increase interaction with students from the beginning.

Sweden -Teknikens Hus

In Sweden, the Third party stated that the guidelines were good overall, however, some Hub members indicated that they were maybe a little too long with a lot of text. Additionally, some pointed out that some examples in the guidelines were basic. As such, this proved slightly problematic in contextualising the implementation at the beginning.

Poland - EXPERYMENT and Serbia - CENTER FOR THE PROMOTION OF SCIENCE

In Poland, and Serbia the Third parties stated that there were no challenges in following the guidelines.

Challenges in ensuring gender inclusiveness during the implementation

Following the identification of some of the challenges faced by the Third parties in relation to the guidelines, the third parties went on to give their opinion and understanding on the challenges in ensuring gender inclusiveness while implementing the activities.

Science Gallery in Ireland stated that the challenge was that despite best efforts towards gender inclusiveness, roles and group dynamics did not come down to gender but to personalities. As a result, this could not be accounted for in advance. Some people, irrelevant of their gender, were naturally more shy and reticent in a group scenario while others were more dominating. The Third party saw both genders fit both of those descriptions in the exercise.

In Austria, the view was that the module they decided to observe (Cooperative Card Game) was only based on women. While playing, students kept asking if there were only women. For a gender inclusive game, this was quite hindering as it caused a separation between women and men.

In Estonia on the other hand, the Third party noticed that gender inclusiveness was considered when choosing the researchers who met and talked to the youngsters. Out of five, three were female researchers. In engaging with the children, the researchers did not seem to tilt towards either including more boys or girls. All of the youngsters got the same welcoming treatment.

In Greece, it was observed that the challenges were mainly a combination between gender inclusiveness and class management issues. For instance, the boys kept talking and interfering without asking for permission. The boys further expected, and at times even demanded the attention of the rest of the class when they talked than girls did.

In Poland the challenge was related to stereotyping. Experyment stated that stereotypes were quite strongly rooted in the young generation. They further stated that these stereotypes were perpetuated by the media and the older generation. On the contrary, in Serbia and Spain there were no problems identified regarding gender inclusiveness during the module implementation.

In Sweden Teknikens Hus stated that even if the focus was to be equal, for example giving the same talk time to girls and boys, it was not always easy to do that in practice. They discussed with the responsible researchers or industry representatives that sometimes it was not enough to follow the guidelines for the first time. This was something that people have to “train on” to be good at.

In the UK the challenge was on how different people understood and used the guidelines. ASDC stated that everyone was an individual with individual motivations, opinion and ideas. Therefore, each person had his or her different take home message from the same guidelines. They observed that most people focused on a particular area that they felt they could or wanted to change and ignored other areas of the guidelines.

Level of gender inclusiveness among teenagers on the four different perspectives

The Third parties were asked to assess the level of gender inclusiveness among teenagers on the four different levels of the Hypatia Framework which includes the individual, interactional, institutional and societal levels.

Individual level

In Ireland, the participants were of mixed gender, but there were more boys than girls. Although there was an imbalance with respect to gender, the boys and girls participated in similar ways. In Austria, there were more boys than girls (2 girls, 10 boys, aged 13/14). The boys and girls

participated in a slightly different manner. During the memory game, the girls were not as active with their comments or their engagement in the game like some of the boys as well. In the discussion, one girl was really active and gave in-depth explanations, while the boys mostly kept their answers short.

In the case of Estonia, the participants were of mixed gender (13 girls and 21 boys aged 15 to 16). They participated in similar ways. In Greece, the participants were also of mixed gender. There was however a difference in how the teenagers were participating. Girls seemed to take their engagement in the activity more seriously than boys. The girls tended to be more involved in the group work than the boys. The latter made fun at first, and it took them some time to get engaged in the activity (e.g. the boys kept talking and interfering with no permission from the teacher and the girls took over writing the answers to the questions posed in 3 out of 4 groups). In Greece, the planned activity allowed both boys and girls to interact. However, nothing was mentioned with regards to the ways the teenagers were participating.

In Poland, the participants were of mixed gender as well. However, they participated differently. The girls multitasked and were more spontaneous than the boys. In Serbia, both boys and girls participated in the activity. However, the boys and girls did not participate in a similar way. Girls seemed to be more aware of the stereotypes in STEM than boys and were better informed than boys which led to more open and direct discussions, while boys were a bit shy about discussing gender stereotypes that target girls in most cases.

In the case of Spain, there was mixed gender in the participation, although it was disproportionate as there were 20 girls to 8 boys. The boys and girls participated in a similar way. This was also the case in Sweden. There was a mixed gender in participation. However, the boys and girls did not entirely participate in a similar way. Girls were more laid-back than the boys.

Interactional level

In Ireland the planned activity allowed girls and boys to interact in a variety of ways. It was observed that the boys and girls had a chance to be methodical, logical and inquiring in learning about circuits, and to be creative and artistic in the designing and making of the wearable part. All the participants had to rotate and do all the parts of the activities. In Austria, the boys and girls partly interacted because of the small numbers of girls as two of the three small groups consisted only of boys. The observation in Austria also revealed that there were not so many different roles included in the activity. The main role was taken by the instructor as she facilitated the game and the discussion. In the discussion, she tried to engage different people by addressing them directly. However, the activities did not reflect a variety of gender neutral or gender equal personalities.

In Estonia that planned activity allowed girls and boys to interact in a variety of ways. Both girls and boys were encouraged to ask questions, they could also tell their personal stories and have a

discussion. The facilitators addressed both girls and boys, gave them time to think about the answers and did not assume that one or the other gender would be the first to jump at giving an answer. The activities reflected a variety of gender neutral or gender equal personalities.

In Greece, the planned activity allowed boys and girls to interact through pair discussions, a group discussion, debate and sketching ideas on the board. All through the activity, the teacher was actively helping out and facilitating. Deliberately, though not overtly, the facilitators allocated the different roles, in turn, to different pupils in the same group each time the activity was repeated. However, the activities did not reflect a variety of gender neutral or gender equal personalities.

In Poland, nothing was said if the planned activity allowed girls and boys to interact in a variety of ways or not. With regards to ensuring that the different roles were rotated, nothing was mentioned to that regard. However, it was observed that the facilitators approached participants of the workshop individually and asked them additional questions.

In Serbia, the planned activity allowed girls and boys to interact in a variety of ways. Interaction included debate as well as time for each of the groups to individually present their opinions to others, which seemed to have had impact on the general discussion in a way that led to more open discussion. Each of the groups took turns which lead to equal opportunities for participants to view all sides of the argument. However, it was stated that the activities did not reflect a variety of gender neutral or gender equal personalities.

In Spain, the planned activity did not allow girls and boys to interact in a variety of ways. No reason for this was given. The activities did not reflect a variety of gender neutral or gender equal personalities. Both the scientists and the teachers were females. It could have been interesting to include a male scientist as well.

In Sweden, the planned activities allowed girls and boys to interact in a variety of ways. They observed, listened and looked, and made practical experiments. Everyone was encouraged to do different things. For example, it was not one individual in a group that had the same role the whole time. The activities reflected a variety of gender neutral or gender equal personalities. The activities were adapted to show that being a researcher or engineer was an attractive job for both boys and girls. This ignited an interest for the ones that were not interested in science before.

Institutional level

In Ireland, the mission of the institution had strongly shaped the activity as the Third party pointed out that gender inclusivity was embedded in the design of the activity and the instruction. This was reflected in staff and visiting artists as well as scientists who were introduced to the students during their time in Science Gallery Dublin. The physical learning environment was appropriate for planned activities. Materials available were an assortment of upcycling materials with no specific

associated gender. Examples used to explain the wearable technology design challenge were gender neutral. There was 1 male and 1 female facilitator.

In the case of Austria, gender is a big topic and is often addressed at the school where the toolkit was being implemented. During the activity it became pretty clear that the students are very familiar with the question of gender as it is discussed often in the school context. The game was played on the floor to break away from the normal habits such as sitting order. The physical learning environment was appropriate apart from the fact that only one teacher was there.

In Poland the mission of the institution helped to popularise STEM among young people in a pleasant and interactive way. In addition, the physical learning environment was appropriate for the planned activities.

In Estonia, the mission of the institution shaped the activity. The institution, which is involved in Chemistry, was very eager to show different sides of their facilities and opportunities for the future. In addition, since the gender balance of chemists in the institution is about 50-50, the activity clearly illustrated this fact. The physical space was very neutral, no gender-specific interior design solutions were used. In addition, the researchers showed their offices, which they were able to set up according to their personal preferences.

In Greece, overall, the mission of NOESIS did not shape the activity in a notable way. It was the first time gender inclusiveness issues were at the heart of the activities at the institution.

In Serbia, the aim or mission of the institution shaped the activity. The mission of CPN in general includes promotion of STEM and equality for all genders in those fields so activities and mission were intertwined from the start, as well as the method of implementation. As such, the shape of the activity remained in its intended form. In terms of the physical learning environment, it was appropriate for the planned activity. Design and materials had no specific genders to which they were exclusively appealing to. As for the role models, the facilitators included examples from both genders equally, although the majority of role models were female.

In Spain, the aim or mission of the institution did not shape the activity. The activity was shaped by the Hypatia guidelines that were followed in implementing the module. The physical learning environment was appropriate for the planned activities although it is difficult to say. At the beginning of the activity, both girls and boys were not really interested and not motivated to participate in the activity. The first testimony of one of the scientists captured their attention, although a small group did not participate at all in the activity. The testimony of the daily activities

of the female scientists managed to attract most of the audience. The female scientists were vested in the facilitation of a good dialogue.

In Sweden, the activity was shaped by the mission of the institution. For instance, despite having a female facilitator, the students met lot of different staff and researchers who were both women and men from different countries.

Societal/Cultural level

In the case of Ireland, it was difficult to conclude whether the activity caused participants to question gender norms that may be established at the societal/cultural level. Although this was the case, the Third party stated that it could be argued that it challenged the idea that boys are logical and better suited to engineering-type challenges and girls are creative and better at artistic or design activities. This was challenged because both genders had to do both types of activities and all succeeded well in achieving them. The activity was straightforward and did not pose any problems or clash with any societal or cultural norms.

In Austria, the activity did not cause participants to question gender norms. The discussion was more about gender in general, but not about detailed examples. Students rather described examples, in which the gender norms have been already broken up (e.g. boys are into fashion). It felt as if they did not experience it as a big problem. The activity clashed with societal or cultural norms and gave rise to difficulties with respect to achieving the intended outcomes. Students were really surprised by the game mode, which they had to play against each other as they are used to act as a team. This caused some arguments and complaints as some students helped others.

In the case of Estonia, the activity did not cause participants to question gender norms that may be established at the societal/cultural level. The facilitators stressed that studying STEM is for everyone. However, they also stressed that it is a personal choice based on preferences. The fact that the youngsters saw very motivated female and male facilitators, helped to reinforce the notion of STEM being for everyone. This is something that resonates with the common rhetoric in the Estonian society. The activity did not clash with societal or cultural norms since the institution's practices were not revolutionary in the cultural sense.

In Greece, the activity caused participants to question gender norms that may be established at the societal/cultural level. It was the girls that questioned the established gender norms and started the related discussion. For instance, on one hand, one girl asked if it was "sexist to have these sort of discussions, because it was an indicator that there was actually a gender issue because the discussions were initiated by girls". On the other hand, a boy, insisted throughout the activity, that there are biological reasons why boys are better in STEM careers. The activity clashed

with societal or cultural norms. The questions, both oral and written, posed, and induced some sort of reaction from a number of pupils. The reaction ranked from funny comments to opposition. The latter was both a teenage attitude and in a smaller percentage an opposite view.

In Poland the activity caused participants to question gender norms that may be established at the societal/cultural level. It appeared that young people have noticed that society is conditioned by the role that girls and young people play in their lives. The activity clashed with societal or cultural norms. This gave rise to difficulties with respect to achieving the intended outcome because breaking stereotypes is always difficult and requires constant work.

In Serbia, the activity caused participants to question gender norms. The discussion in the activity was the main goal of the workshop in which the main topic was in fact stereotypes with the emphasis on destroying those stereotypes. The activity clashed with societal or cultural norms which led to the rethinking of the examples that were given during the activity.

In the case of Spain, the activity caused participants to question gender norms that may be established at the societal/cultural level. The motives that can condition the selection of the professional career both for boys and girls were debated, such as capacitation, attitude, social conditionings among others. The students were interested in knowing the research currently carried out at the ALBA Synchrotron and in the fact that the Operations Manager was a woman. The activity did not clash with societal or cultural norms.

In Sweden it was difficult to tell whether the activity caused participants to question gender norms that may be established at the societal/cultural level or not. The fact that the facilitators were females showed who the role models were. But discussions about gender norms were not raised during the activities. In addition, the activity did not clash with societal or cultural norms.

Interviews conducted by Third parties

In order to get an in-depth understanding of the implementation and adoption of the toolkit, Third parties interviewed relevant stakeholders in their countries in addition to making observations. The expectation was that aside from making observations with teenagers, the Third parties had to conduct an interview each with a stakeholder of their choice from the institution that was implementing the toolkit. Such a stakeholder could be a head teacher, museum coordinator or industry representative. While all Third parties were tasked to make the observations, only five of the Third parties were chosen to conduct the interviews (for specific questions see Annex 5). The interviews focused on the following areas:

Usefulness of the guidelines on facilitation and gender inclusiveness

Countries including Ireland, Austria, UK and Poland considered the guidelines useful because they gave the opportunity to think about gender stereotypes. This allowed the evaluation of teaching and facilitation styles. The Third parties appreciated that the description of modules was very good as it contained all the information needed while not going too much in detail. Further, they considered the guidelines helpful in terms of giving practical advice to carry out the workshops effectively.

In the case of the UK, the guidelines assisted in the creation of the framework for training sessions for industry professionals to go into schools or to deliver events and workshops at their workplaces. Gender inclusive activities and useful links were utilised alongside other physical resources available in the UK such as the 'People Like Me' WISE training pack.

In Ireland, the guidelines were very useful. Facilitators stated that the guidelines gave them the opportunity to think about stereotypes and to question their own biases and teaching/facilitation styles. The interviews also revealed that the guidelines made them think about the content of the other modules and workshops they would be delivering during the week and to ensure that the tools from these guidelines were used for all the other scenarios.

Similarly, in Austria the Third party also found the guidelines useful. They stated that the guidelines were very good as was the method and the description. The Third party further revealed that while not going too much into detail, the guidelines contained all the necessary information. This was especially in relation to the different version of the activity which helped to get some ideas for the right adaption for the class. Also, the idea to set up a leash was great. It was a nice change to the normal activities that they did in the class. The same was also the case for the Third party from Poland that revealed that the guidelines were very helpful because they gave practical advice and helped the facilitator to carry out the workshop effectively.

The rest of the four countries, namely; Estonia, Serbia, Spain and Sweden did not provide any response to explain whether they found the guidelines useful or not.

Risks and challenges encountered while implementing the modules

In terms of challenges, few countries like Ireland and Austria indicated that there were no particular challenges or major risks while conducting the modules as the staff was well trained in delivering workshops and covered safety guidelines at the start of the workshop. In schools, the Third parties were told that the staff were very open to new formats.

Table 4, below indicates the responses from facilitators that were given to Third parties in the different countries where the toolkit was implemented.

Country	Risks and challenges
Ireland	<p>No major risks - staff were well trained in delivering workshops and covered safety guidelines at the start of the workshop. For example, one student in the group was on the Autism Spectrum and another had suffered an anxiety attack on the previous day, but the facilitators were aware of these issues and ready to intervene were any issues to arise during delivering of their module. The student who was on the Autism Spectrum did approach the design challenge slightly differently than directed as he wanted to show the circuits of his creation and not cover it with any materials. The facilitators were accommodating and did not push him outside of his comfort zone.</p> <p>A challenge was in keeping a large group in sync so that the group would be able to do the task at the same time and with the same part of the task at any given time. However, some students naturally rushed ahead.</p>
Austria	<p>There were almost no challenges as the activity did not require any special resources. As the school was open to new formats, there were no complications within the school setting. The only problem that occurred was that the discussion was not as good as expected. This was because the students did not have a clear opinion which made it difficult to start a discussion. As such, it was probably conceivable to think of a new approach.</p> <p>However, it was difficult to uncover the different roles or gender typical stereotypes as the game only included women. It seemed to be a rather feministic game. The stakeholder thought that there should have been stronger emphasis on the accomplishments of those women and on the inventions they discovered instead of only explaining what they became. Additionally, due to time limitation, the discussion was rushed.</p>
Estonia Serbia Spain Sweden	No responses

Greece	
Poland	It was a challenge to influence facilitators' opinions.
UK	<p>The challenges were personal, particularly for female industry role models. The challenge had to do with whether individuals felt empowered to think that what they do would make a difference or that whether what they had to say would be interesting to an audience. There was also the fear that they would put some people off because some people do not naturally feel confident that they can be an inspiring speaker when put in front of a group of teenagers.</p> <p>Although a great number of STEM ambassadors were men, as they often put themselves forward, those that came to the training sessions on gender inclusion were often the women. This was not exclusively the case, but it was the female role models who were motivated to act. However, a huge change in attitude occurred in the men that attended.</p> <p>Since the stakeholders were training cohorts of professionals to implement gender inclusivity in their interactions with the public and school students, the challenge was to keep the training and CPD for thousands of industry professionals at a high quality without funding to do so. Despite that, it was considered of high priority and as a result was done without funding.</p>

Table 4: Responses from stakeholders that were given to Third parties

With regards to the risks that were identified in the different countries, as presented in Table 4 above, it became apparent that different stakeholders had put some measures in place to minimise those risks. The measures that were put in place in countries where responses were provided are outlined below:

- In Ireland, the risk was mitigated by being prepared to brief students in advance on safety risks. In addition, the risks were mitigated by being aware of students differing learning needs and by being ready to respond accordingly and appropriately.
- In Austria, the risk was mitigated by formulating more provocative statements in order to get a reaction from students. In addition to this, students were also asked directly about their opinion. However, despite these measures, the students were not so engaged or willing to give more than just a short answer.
- There was nothing said for Estonia, Serbia, Sweden, Greece and Spain with regards to risk mitigation.
- In Poland, the facilitators had enough time to prepare themselves for workshops and activities.
- In the case of the UK, public speaking training helped the role models with confidence and making sure they were simply suggesting little changes that could be reframed in order to make a difference. This was put in place to mitigate against putting some people off from becoming industry role models altogether. The training provided by Graphic Science in terms of gender inclusivity for STEM industry professionals was one of the most popular training sessions. This was important because it helped people to notice that there was a problem and this provided something concrete that could be done about it.

Strategies which were used to implement the activities

The Third parties also asked the facilitators on the different strategies that they used to implement the toolkit activities. Despite not having indications for strategies used in Estonia, Spain, Serbia, Greece and Sweden, the following strategies were given from the other countries:

- i. In Ireland, the stakeholders used Kahoot Quiz focussing on sex and gender, gender stereotypes, and gender statistics in STEM for the country. The facilitators also used gender neutral instructions and examples. In addition, they employed an open ended creative challenge which allowed students to think for themselves, test and modify activities. Lastly, they encouraged peer-to-peer sharing of tools, idea and materials among the students.
- ii. In the case of Austria, the facilitators played the memory game in order to get students to be familiar with the content. This was used as an introduction to the topic. Afterwards, the students were divided into three smaller groups for the task. It was thought that within the small groups, the students could effectively engage in discussions.

- iii. In Poland, there were no instructions on how to perform “step-by-step” experiments. Rather the focus was on the creativity of the participants. The strategy was very important because it promoted the teenagers’ ability to independently experiment and draw conclusions.
- iv. In the UK, those attending the activity were scientists and professionals. As such, it was very important to know who was being trained. Many were from very high level academic backgrounds, so it was important to provide the academic background to the guidelines, which were critiqued before they were accepted to provide evidence for a change in practice. The STEM professionals ensured that they provided evidence for ‘why’ these guidelines were important. Thus, they used examples that they were very confident with and often based evidence in psychology and social science. Having a firm grounding in the theoretical background was therefore an important strategy as well as being able to understand and justify the analysis. Lastly, to make sure that the resources were as engaging as possible, it was also important to include biographies, real-life examples and role models for those in the training.

The Third parties then went on to ask the facilitators who participated in the modules that were chosen for the evaluation if the strategies that were implemented worked well. Apparently, the facilitators from Ireland, Poland and the UK said they worked well while those from Austria stated that they did not. As a follow up, they were further asked if there were any challenges associated with the strategies and how they overcame them. There were no responses from six countries apart from Austria, Poland and the UK.

To begin with, in Austria the challenge was that the students were not focused. On top of that, there was not enough time to go through the order of the cards. Therefore, the presentation of the cards was skipped, and the discussion was started right way. In the case of Poland, there was an issue with space for the independent work of the workshop participants. Despite this, the facilitator, asked questions rather than instructing the participants. This strategy provided a lot of freedom for experimentation and the ability to draw own conclusions.

Lastly, in the UK, the challenge was around grabbing the attention from the participants and making the training effective. This was overcome by giving lots of opportunity for pauses that allowed the STEM professionals to ask challenging questions to each other and to the facilitator. This was important because it allowed for the professionals to talk about themselves differently by discussing their work-life balance, their personal stories and skills instead of focussing on talking just about what they do. The facilitators ensured that the training that was provided in gender inclusivity in STEM teaching was comfortable enough to share thoughts and materials, encourage listening skills and thinking about the different levels of approach (from individual to cultural). At times they had to agree to disagree within the group, but it was important to listen and allow

everyone to work together. Something as simple as the lay out of the room (avoiding a lecture space and arranging it as small groups around tables) made a difference to the response to training.

Improvements that can be made to the modules for future use

The last focus of the interviews was on suggestions for future improvements that at the end were mostly related to aspects of the certain activities/modules they implemented. A number of suggestions were made which are as follows:

- The inclusion of hairclips in the materials for the Wearable Technology module is NOT gender inclusive as they are very much associated with girls. Therefore, this needs a rethink for the future
- In the cooperative card game images could be used that show women as teens, so that the students could better connect with the women.
- More time was needed for certain activities. Maybe a second phase could have been added in which the students could have had the opportunity to learn more about an invention.(relevant to the Cooperative Card Game)
- More examples of activities could have been used during workshops.

Conclusion

This evaluation report presented the findings from the interviews and observations that were conducted during toolkit implementation. The findings reveal that the different toolkit activities were framed with gender inclusiveness in mind despite some challenges that were encountered during their implementation. The findings also show that there was a good level of stakeholder engagement in both adaption and implementation of the toolkit activities. Also, there was great emphasis put on ensuring that teenagers had an active role in the project. From both interviews with the Third parties and project partners responsible for the piloting and implementation of the toolkit activities, it is clear that there was an active participation of teenagers, for instance, through feedback during process. However, although this was generally the case, there were some challenges that were encountered during the process such as delay in sharing of information among project partners and school systems being strict in some countries. In addition, a few stakeholders and Third parties found the guidelines either too long or difficult to follow, despite them being generally helpful.

Likewise, there were some contextual differences across the nine countries, in which the toolkit was implemented, with regards to attitudes on gender inclusiveness at individual, interactional, institutional and societal levels. Some of these differences were pertaining to attitudes and stereotyping towards gender and STEM careers, the structural systems within institutions and

representativeness of teenagers and facilitators from both genders. However, despite these differences, facilitators did a great job in creating different suitable strategies for implementation for their respective countries. The implementation strategies were adapted by taking into account the local factors of where the toolkit activities were being implemented. There was fluidity in finding solutions to some of the issues that were encountered during the implementation process. Some of the solutions were geared to address issues relating to guidelines and the environment in which the toolkit activities were being facilitated. For example, there were some changes made to class set ups, language and format of activities.

The evaluation also has revealed that the Hypatia project has made some notable impacts through the implementation of the toolkit. For example, this was evident in the national seminars conducted by WP 5 which reached out to the targeted head teachers and teachers. In addition, the project's aim of developing 15 toolkit activities was realised and saw the activities adapted and translated in the 14 countries that the Hypatia project focussed on. An additional positive impact of the project is the fact that there has been a heightened awareness of gender inclusivity in STEM in the institutions that took part and were affiliated to the project. This awareness has been evident through institutional websites, events and newsletters.

Further, the evaluation, through third party interviews revealed that although the third parties were accepting of the activities assigned to them, some of them would have preferred to begin the adaptation process with the basis of their own activities rather than a selection of modules prepared by other organisations. Certain third parties felt that the allocated activities were slightly restrictive and as such would have preferred to have more time to work with them in order to have more flexibility and adapt them even further. Despite this, the evaluation noted that the reason why the activities were selected in this format rather than have third parties choose at the second year any activity they wanted was done for various reasons. First in order to give the maximum time to integrate the Hypatia framework to a set of a limited number of activities and second to have them thoroughly piloted and tested. Finally it was done to ensure that all settings (informal organisations, schools, industry) were targeted . A more flexible scenario would have resulted in a less coherent project.

It was clear from the evaluation that teenagers played a key role in the project. This is evident from the fact that teenagers were invited to test and give feedback on the activities chosen. There was great emphasis put on ensuring teenagers play an active role in implementing the module activities. It is also clear that teenagers proactively participated in the toolkit development. Additionally, there was a feedback system put in place to ensure that the chosen module activities were implemented in the right way. The different activities were framed with gender inclusiveness in mind. The inclusion and framing of gender inclusiveness criteria in the project was one of the biggest successes as it provided theoretical background for the project work.

The evaluation has also revealed that the Hypatia project through its partners and third parties has developed various national and European networks which is positive for sustainability. This can be seen from the local action plans that have been developed by the respective partners. The local action plans list which partners the Hypatia partners and third parties will keep working with once the project comes to an end and what elements of the project they will keep working with. This is a positive outcome in as far as sustainability is concerned beyond the life-time of the Hypatia project. Such action plans are also good for ensuring extensive dissemination and education so that there is a wider reach for teenagers that could consider undertaking STEM related subjects locally and nationally now and in the future.

Annexes

Annex 1: Evaluation Strategy Document



Inception Report: Evaluation Strategy for Horizon 2020 Project HYPATIA

Dr Kutoma Wakunuma

Tilimbe Jiya

November, 2015

Contact details for corresponding authors:

Dr Kutoma J. Wakunuma

Centre for Computing and Social Responsibility

De Montfort University

The Gateway

LE1 9 BH, Leicester, United Kingdom

kutoma@dmu.ac.uk

Mr Tilimbe Jiya

Centre for Computing and Social Responsibility

De Montfort University

The Gateway

LE1 9 BH, Leicester, United Kingdom

tilimbe.jiya@dmu.ac.uk

Contents

1 Executive Summary 34

2 Background 34

3 Introduction 34

4 Objectives of the Evaluation 36

4.1 Assessing Impact of the Project 36

4.2 Looking at Evaluation Criteria 36

4.3 Gauging Indicators of Success against Expected Impact 36

5 Evaluation 36

5.1 A Formative Approach 36

5.2 A Summative Approach 37

6 Methodology 37

6.1 The Evaluation Design and Guidelines 37

6.1.1 Evaluation Guidelines 37

6.1.2 Evaluation Framework 38

6.2 Evaluation Approaches in Detail 43

6.2.1 Observations 43

6.2.2 Interviews 44

6.2.3 Translation and Transcription 45

6.2.4 Document Analysis specifically through use of Basecamp 46

6.3 Data Analysis 46

7 Activities Timeline 46

7.1 Interim Report in Month 18 46

7.2 Final Report in Month 36 46

8 Conclusion 46

9 References 47

- **Executive Summary**

It is well documented that there are more boys than girls taking up STEM subjects. However, it is now a general consensus that in order to overcome societies' challenges, both boys and girl ought to be given and have the same opportunities when it comes to education, jobs and other economic empowerment activities, which is consonance with Hypatia's aim and expected impact. In order to ensure that this aim is met and the expected impact is realised effectively, the project needs to be evaluated. The evaluation is important not only to provide evidence about the efforts and quality of the Hypatia's activities, but also to improve outcomes in future. This inception report describes the evaluation strategy for evaluating the Hypatia project. The evaluation strategy will outline the processes of evaluation which will be used in order to assess the objectives of the project e.g. implementing modular toolkits that promote engagement of teenagers in STEM in a gender inclusive manner. Ultimately, the proposed strategy will assess whether the expected impact of the project has been reached. The strategy involves two evaluation approaches which include both formative and summative approaches. These will be employed to meet the objectives of the strategy which involve assessing the impact of the project by determining how toolkits are being implemented, looking at evaluation criteria and gauging indicators of success in terms of the way science is communicated. The expectation is that this will go a long way towards looking at indicators of success that will be used to take into account the expected impacts which include:

- increase in participation and interest of girls in STEM
- encouragement of sustainable collaboration among schools, science museums, research centres on gender equality and science
- contribution towards European Research Area objectives and
- Innovation Union Objectives.

As part of the strategy, the inception report also presents an evaluation methodology that consists of an evaluation design that will include a set of evaluation guidelines to be used in conducting the evaluative activities proposed in this evaluation strategy. Furthermore, the evaluation design will include an evaluation framework which will be used in assessing findings from the evaluation. In addition, the evaluation design will also include approaches that will be employed in data collection such as observation, interviews and document analysis. Further, the inception report touches on a data analysis approach for analysing results from the evaluation activities. The report then concludes with an outline of activities that are planned for Hypatia's evaluation within a predefined timeline.

- **Background**

The evaluation team is part of the Centre for Computing and Social Responsibility (CCSR) at De Montfort University (DMU), Leicester. The centre is one of the leading research centres in the UK and its research areas include; responsible research and innovation (RRI), civil society organisations in research, stakeholder engagement and involvement, ethics (computer and applied ethics), information communication technologies (ICTs), gender in technology and project evaluation. The centre has been involved in a number of EU funded projects at an international level such as Network Analysis of CSO Participation in Research Framework Programmes, Responsible-Industry, Responsibility, GREAT (Governance for Responsible Innovation), CONSIDER (Civil Society Organisations in Designing Research Governance), ETICA (Ethical Issues of Emerging ICT Application) and SATORI (Stakeholders Acting Together On the ethical impact assessment of Research and Innovation) project. The team has wide experience in evaluation methodology and implementation.

- **Introduction**

Over the years research has shown that the way science is communicated to young people in and out of school is not yet gender inclusive. As a result, this has left both boys and girls having little knowledge and awareness of the range of careers in science, technology, engineering and mathematics (STEM), and the skills that are relevant for those career paths. This has affected the number of teenagers who are pursuing STEM related careers. STEM subjects are important drivers and cornerstones for development therefore, increasing women's as well as men's pursuit of STEM-related jobs is considered critical for any nation to remain competitive in the global economy[1]. Although this is an important aspect, girls remain underrepresented in many STEM fields. This is down to social influences which include the relative degrees of encouragement that girls may experience to do well in STEM and non-STEM subjects. In addition to that, personal influences such as gender-related variations in self-schemata and attitudes shape girls' motivation in STEM or non-STEM domains [3]. Without science and technology, social and economic development would be difficult to achieve. This aspect is highlighted in several H2020 programmes that indicate that in order to overcome societal challenges, STEM must be at the heart of Europe [4].

As part of the EU's development agenda for a better Europe, the need to change the state of affairs and with regards to STEM and gender inclusivity has become more vital than ever before. It is therefore essential to ensure that STEM careers become increasingly oriented towards society's needs and opening up new dimensions in terms of the skillsets required in a gender inclusive way. In a study by Wilson (2009), it was found that STEM fields constitute 9 of the top 10 college degrees leading to the highest paying occupations. The study suggests that if girls and boys are equally encouraged to pursue STEM related subjects, they can equally achieve equal economic empowerment for both which is positive for overall societal benefit [5]. Therefore, considering proportion of teenage girls pursuing STEM subjects, it is fundamental to expose more girls to the variety of STEM-related careers, empower them to make connections that develop their lives and their own skills.

In order to realise this, there is a need to engage teenagers in STEM in a gender-inclusive way, and addressing the attitudes of STEM education professionals towards more gender-inclusive practices. This is the mission that Hypatia is aiming to achieve by bringing together groups of stakeholders from science centres and museums, schools, research institutions and industry to collaborate with gender experts and teenagers in hubs and co-develop the content and co-organise activities that promote girls awareness of STEM careers. Work being carried out by Hypatia is important in ensuring that there are equal opportunities for both boys and girls in getting enough information of the variety of STEM career pathways they can follow. This is in line with Europe's agenda on RRI of which one of its pillars is 'gender inclusivity'[2]. In line with this, the project has therefore planned to develop and implement expert toolkits in schools, museums, research institutions and industry and conduct a set of seminars across 14 European countries bringing together head teachers, museum professionals, researchers, industry professionals and other related stakeholders to discuss how the toolkits should be implemented. Hypatia aims at ensuring an effective adoption of the toolkit of activities that will focus on ways of communicating STEM to empower teenage girls in exploring the range of skills needed for a variety of STEM studies and careers. The project will also develop guidelines for engaging teenagers in STEM in a gender inclusive way.

Considering the efforts and goal(s) of Hypatia, it is important that the project is evaluated. The project needs to be evaluated against its aim and objectives in order to ensure that the results have been reached. This is where DMU as evaluators will implement the strategy proposed in this inception report. The evaluation strategy is directly tailored to Hypatia and focuses on the methodology for evaluating the outcomes and impact of the project. The evaluation strategy will ensure that Hypatia's work is meeting its intended objectives of bringing lasting change to attitudes towards STEM and how it is communicated in museums, schools and industry through engaging with teenage girls in STEM across Europe. The evaluation that we propose to undertake for the Hypatia project will be both formative and summative. Using the formative evaluation we will be able to contribute to the refinement of toolkit implementation by identifying weaknesses or areas

for improvement [8] through feedback. On the other hand, use of summative evaluation will enable us to evaluate the quality of the project's outputs and outcomes in order to assess its success [8] for example in meeting its stated objectives and expected impacts against a set of success indicators that are laid down in a pre-defined evaluation framework.

In this inception report we will introduce the objectives of the evaluation strategy then we will move on to a brief discussion of evaluation, touching on the two evaluation approaches that will be employed in Hypatia's evaluation. Following this, we will then introduce the methodology that will be used in the evaluation, consisting of the evaluation design proposed for Hypatia, where we will cover the evaluation guidelines that will assist in directing the evaluation process and our proposed evaluation framework. Subsequently, we will then discuss the evaluation approaches in detail before concluding with the timeline for the next evaluation activities.

- **Objectives of the Evaluation**

In order to effectively evaluate Hypatia, the evaluation strategy will aim at meeting the following 2 objectives:

- **Assessing Impact of the Project**

This will entail assessing Hypatia's expected impact in order to understand whether the proposed activities are meeting the impact or not.

- **Looking at Evaluation Criteria**

This will cover elements that will assist in determining whether the expected impact has been met.

- **Gauging Indicators of Success against Expected Impact**

This will involve using a set of indicators of success to gauge whether Hypatia has reached its expected impact or not.

- **Evaluation**

The term evaluation has many meanings depending on context and purpose of the evaluation, however all the definitions have elements of credibility[9]. Evaluation involves learning new knowledge through gathering information, making credible conclusions or judgements that can be used in decision making and communicating the findings to an audience[10]. Evaluation is paramount because it acts as a control mechanism [11] that ensures that strategic benefits of undertaking project are realised. Broadly speaking, evaluation can be defined as "the process of determining the merit, worth and value of things" [12], and can be used to describe many "different kinds of judgments, from informal assessment that relies on intuition or opinion, to well-defined and systematic research that makes use of social science research methods" [13]. In light of Hypatia project, evaluation will focus on the "design, implementation and effectiveness" of the project's toolkit activities [14], as well as outputs where we will specifically be looking at the such elements as adoption of the toolkits e.g. from reports/deliverables and attainment of expected impacts [15].

In evaluating Hypatia, we will use both formative and summative approaches to evaluation. These 2 distinct approaches are discussed in the following sections.

- **A Formative Approach**

A formative evaluation approach examines the quality of procedures or methods of stakeholder engagement while they occur and their outcomes contribute to the refinement of the engagement processes by identifying weaknesses or areas for improvement prior to the project's conclusion, thereby acting as a feedback mechanism or 'double loop' to refine project activities [18]. In relation to Hypatia, with this approach we will look at assessment of deliverables and adoption of modular toolkits among others. For instance, we will assess where and how the third parties are implementing the toolkits in relation to the guidelines. This will give us an understanding of the usability of the toolkits, usefulness and effectiveness of the toolkit as well as who is engaging with the toolkits in terms of gender inclusivity.

- **A Summative Approach**

A summative evaluation approach assesses the outcomes of procedures or methods [7], [19]–[21], evaluates the quality of engagement project's outputs and outcomes in order to evaluate its success [7]. In terms of Hypatia, we will use this approach to evaluate whether the project is or has met its stated objectives, expected impacts against a pre-defined evaluation framework [see section 6.1.2]. For example, we will evaluate if the activities have been well adopted and are established within their respective hubs. This will help us determine the overall impact (value or significance) of the Hypatia project.

- **Methodology**

- **The Evaluation Design and Guidelines**

The evaluation design will include a set of evaluation guidelines that will be used in conducting the evaluative activities proposed in this strategy. It will also include an evaluation framework which will be useful in analysing our findings. In addition to these, the evaluation design includes evaluation approaches that will be used in data collection and how the data collected will be analysed.

As there is a budget constraint, partners and third parties will play a key role in implementing the evaluation design in their respective countries. This will be with the guidance of the evaluators who will provide evaluation guidelines described in the following section.

- **Evaluation Guidelines**

The evaluation guidelines expand on the above described evaluation design. The guidelines are mainly a guide to partners of the project who as indicated in the preceding section, will play a pivotal role in implementing the evaluation design. The use of the term 'partners' in this document denotes the 5 principal partners of the Hypatia project with specific work packages and the third parties. We use partners in this document because we view all parties concerned in Hypatia as playing a relevant role in the evaluation of the project. The Hypatia project is a very interesting and important project, particularly in encouraging teenage girls to get involved in STEM. As such, there are very interesting, important and diverse elements that need and could be evaluated during the course of the project such as:

- feedback on development of toolkit activities to partners
- the implementation of the toolkit activities
- assessing the project impact and providing partners with a better idea of the value of their participation by tracking influence on the process
- improving the design of future related activities

However, as is indicated in section 6, due to the limited resources allocated for the evaluation of the project, evaluating the many diverse elements was always going to be a challenge. As such, it is important to prioritise what can be evaluated, how it can be evaluated, where and when. During

the kick-off meeting held in Amsterdam between 2-4 November 2015, the evaluation team was fortunate enough to have received very useful feedback from members of the project that could be incorporated into the evaluation, especially in terms of what we might evaluate, how we might evaluate, where and when we might evaluate. The main feedback which members of Hypatia were keen to realise with regards to evaluation, is for a concerted effort to have European wide evaluation results. This means that the evaluation of the Hypatia project will need to extend to as many European countries as possible while considering the resources available. In order for this to be a reality, Hypatia partners will have to assist with the implementation of the evaluation. Specifically, the most ideal group to assist in this evaluation would be the third parties who are assigned with the implementation of the STEM toolkits. This is because the third parties will be implementing the toolkits in hubs in their respective countries. In light of this, the evaluation team will send detailed guidelines to third parties in month 18. Month 18 has been chosen because the evaluation team has into consideration 2 facts:

- i) That the production of the toolkits will be between month 15 and 18.
- ii) That the expected implementation of the said toolkits will between months 23 to 36.

As such, sending of the observation and interview guidelines in month 18 is ideal because it will give third parties enough time (5 months) to sort out logistics as described in the timeline. Elements of what will be covered in the guidelines are discussed in more detail in section 6.2. In the meantime, below is a discussion of the elected evaluation framework.

- **Evaluation Framework**

The evaluation framework (see Table 3) has four specific elements that focus on the following:

- ***Expected Impact***

The Hypatia consortium answered the GERI-1-2014 call which has specific impacts that the project has to achieve. These expected impacts and how Hypatia intends to achieve them are outlined in Table 1 below:

Expected impact	How Hypatia will achieve this impact
1. Change the way science is communicated	The project will reach schools, informal learning organisations, research institutions and industry activities. It will produce sets of activities and guidelines about how to communicate science with their audiences ensuring the gender aspect is taken on board. The project will measure the number of institutions that adopt the new tools.
2. Increase the participation and interest of girls in STEM	The activities that will be used will be based on existing good practices related to gender in STEM. Now these activities will be implemented in much larger scale.

<p>3. Encourage sustainable collaboration among schools, museums, research centres on gender equality and science</p>	<p>The creation of hubs in each participating country and the existing connection among partners will ensure a long-term collaboration, supported through the online community Scientix. The wealth of existing activities and the importance given to advocacy and sustainability will ensure a long and strong collaboration among stakeholders.</p>
<p>4. Contribute towards European Research Area objectives (increase female researchers in Europe)</p>	<p>By giving a different perspective on STEM careers to both young boys and girls today we contribute to the increase of female researchers in the future.</p>
<p>5. Contribute to the Innovation Union Objectives (better match skills to available jobs)</p>	<p>Discovering young people's skills in relation to current and future STEM careers is a particular focus of the proposal, with a direct relevance to gender. A set of activities will be directly devoted to this, including the Find Your Skills card game.</p>

Table 1: Expected impacts and how Hypatia proposes to meet them

In order to ensure that the expected impacts have been achieved, an evaluation framework becomes imperative. This is important in order to understand whether the expected impacts in the call have been met or not. As such, the evaluator's will use the listed expected impacts as a check list to assess whether Hypatia has achieved its goal of meeting the GERI-1-call expected impacts.

- **Suggested Approach**

These are the approaches that will be employed in the evaluation process. In this case, the proposed evaluation approaches will be in the countries where the toolkits will be implemented by third parties. These approaches will include observations and interviews. These are discussed in more detail in sections 6.2.1 and 6.2.2.

- **Evaluation Criteria**

Evaluation criteria is the benchmark upon which Hypatia's work and agenda with regards to attitudes and gender inclusion of STEM will be assessed during the evaluation. Table 2 below describes the evaluation criteria by looking at what criteria will be used, its description and which expected impact is the criteria targeting.

Criteria	Description	Targeted expected impact from Table 1
Activities	What toolkit activities and guidelines were developed and delivered and to whom	1 and 5

Representativeness	How many boys and girls are using and engaging with the toolkits	2 and 4
Geographical coverage	How many countries Hypatia's activities have reached and how effective have they been implemented	3 and 4
Usability, usefulness and effectiveness	How usable are the toolkits and how useful and effective have the activities been in reducing the gender gap in STEM	1, 2 and 3

Table 2: Criteria for evaluating targeted expected impact

- **Indicators of Success**

During evaluation, it will be necessary to have indicators of success which will show whether Hypatia has reached its impact or not. Using Table 1 which tabulates how Hypatia intends to achieve the expected impacts, the evaluation will endeavour to assess the outcomes of the tabulated potential achievements as indicators of success. Specifically, the evaluation will seek to understand;

- Whether Hypatia's toolkits have been adopted
- Who has adopted them (boys and girls/research institutions)
- How well the tools have been received by those that have adopted them
- What sort of activities have been developed
- How the activities have been received

Below is Table 3 summarising the evaluation framework.

Expected Impact	Suggested Approaches	Evaluation criteria	Indicator for success
Change in the way science is communicated	<ul style="list-style-type: none"> • Observations • Interviews • Document analysis 	<ul style="list-style-type: none"> • Activities • Representativeness (gender, geographical location) • Quality of information (gender sensitivity, relevance of information to realise gender inclusivity, use of language that is easy to understand so that STEM is not seen to be intimidating or seen to be too technical such that it 'scares' away young girls) 	<ul style="list-style-type: none"> • Adoption of toolkit (number of institutions and stakeholders using new tools). • Usefulness of Hypatia's proposed guidelines (what do the guidelines cover, who was involved in coming up with the guidelines, do guidelines address gender issues and how to overcome them in STEM)
Increase in the participation and interest of girls in STEM	<ul style="list-style-type: none"> • Observations • Interviews • Document analysis 	<ul style="list-style-type: none"> • Representativeness (gender proportionality) • Openness (What activities are open and inviting to girls to encourage involvement in STEM) 	<ul style="list-style-type: none"> • Degree of involvement (extent to which girls are integrated into STEM processes e.g. Hypatia's research activities) • Information accessibility (extent to which girls access appropriate information and materials on STEM) • Inclusion (extent to which girls are included and represented in STEM activities)
Encouragement in sustainable collaboration among schools, science museums, research centres on gender equality and science	<ul style="list-style-type: none"> • Interviews • Document analysis 	<ul style="list-style-type: none"> • Participation (who is collaborating, how representative is the collaboration, how relevant do collaborating partners feel their participation is and/or appreciated) • Resource accessibility (Exchange of information, information accessibility, contribution to required information) • Stakeholder representation) 	<ul style="list-style-type: none"> • Creation of hubs (How many hubs have been created, where have the hubs have been created and how are they impacting gender inclusivity in STEM, how are the hubs encouraging collaboration)

<p>Contribution towards European Research Area objectives (increase female researchers in Europe)</p>	<ul style="list-style-type: none"> • Interviews 	<ul style="list-style-type: none"> • Representativeness (gender representation) 	<ul style="list-style-type: none"> • Look at Hypatia's proposed different perspectives on STEM careers (what are the different perspectives, how different are they from traditional perspectives e.g. attitudes to STEM)
<p>Contribution to the Innovation Union Objectives (better match skills to available jobs)</p>	<ul style="list-style-type: none"> • Interviews 	<ul style="list-style-type: none"> • Accessibility of information with regards to available skills 	<ul style="list-style-type: none"> • Look at Hypatia's proposed activities that have a gender focus on building STEM skills (what were the activities and how were they directly relevant to gender)

Table 3: Evaluation Framework

Deliverable

○ Evaluation Approaches in Detail

As touched on in section 6.2.2, this section gives more details of the approaches that have been adopted for the evaluation. As evaluators, we suggest two approaches that can be used easily, efficiently and effectively for evaluating and therefore meeting the expected evaluation outcomes. These two approaches should not be too onerous to implement particularly by partners and third parties who are expected to play a significant role in implementing the evaluation process. The two elected approaches are observations and interviews. As the development and implementation of STEM toolkits is central to this project, particularly in terms of engaging teenagers in STEM, it becomes imperative to observe how teenagers are engaging with STEM activities in addition to how institutional hubs are supporting the toolkit implementation and subsequent adoption.

To get an in-depth understanding of the implementation and adoption of the toolkits, interviewing relevant stakeholders in addition to observations becomes a necessity. As evaluators, we recommend that each third party conducts at least 1 observation at a hub of their choice in their respective country. In addition to the observations and as a way of complementing and strengthening the observations, an additional set of interviews will be carried out by the third parties. These interviews will be with any of the following stakeholders involved with ensuring that the implementation of the toolkits comes to fruition in their respective institutions: head teachers, teachers, museum coordinators, industry representatives. Further to this set of interviews will be another set of interviews specifically with some of Hypatia's partners and third parties. These will be conducted by the evaluators via Skype.

▪ Observations

The suggested observations will be used to assess the usability of toolkits, toolkit user numbers, the effectiveness of the toolkits and the engagement with the toolkit activities. Therefore, the observation will look at the following;

i) How the toolkits will be used

This will help us to understand the usability of the toolkits. The main aspect that will be looked at here will be understanding whether the toolkits are usable for the intended users and by default, whether the users actually find the toolkits usable. This will indicate whether they are being effective for the purposes for which they are intended.

ii) Who is using the toolkits

This will help us to gauge whether the intended users are actually using the toolkits. For instance, we would like to know whether the selected institutional hubs such as museums and schools are using them and whether by extension, teenage girls are engaging with the toolkits

iii) How often are toolkits used

This will help us develop our knowledge of how often the users use the toolkits since being implemented.

iv) How effective the toolkits are

This will help us assess how effective the toolkits are in encouraging teenagers to engage with STEM activities.

With regards to conducting observations there are different types of observations. One is where the observer is detached from the subject matter being observed while the other is where the observer is more involved in the subject matter being observed. In the detached one, the observer merely looks at activities taking place without any meaningful interaction



with the observed. In the other type where the observer is more involved there is meaningful interaction with the observed. This means that the observer will talk to the subject, perhaps be hands on with the activities in an attempt to cultivate more understanding of a situation. In our case, we recommend the latter in that as the third parties conduct the observations, they will also have some level of interaction with the subject being observed.

With the above in mind, we recommend that observations be conducted in the countries being represented by third parties. However, if a country has two representatives, the evaluators will only select one representative to implement the evaluation design. In addition, if there is more than one hub created in a country, we will leave it to the third party implementing the toolkits to decide the most ideal hub to conduct the observations. This is because the third parties will have the relevant knowledge of where the hub will have been created as well as how easily accessible it will be in their respective country. As evaluators, we would however recommend that third parties make observations in museums. This is because we feel that it would be easier to access participants who come to museums merely because museums are usually more open and do not necessarily need lengthy and protracted accessibility permissions compared to schools for instance. As a recap and in addition to the four areas that will need to be observed which are outlined above, third parties are expected to observe the following:

- i) How the toolkits are implemented in relation to the toolkit guidelines: e.g. third parties will look at whether the guidelines were adequate enough for toolkit implementation to be robust
- ii) How teenagers are engaging with toolkit activities
- iii) The number of girls and boys engaging with the toolkit activities over a period of time. Preferably this observation should be from the implementation month up to two months following the implementation. So for example, if a third party implements a toolkit in month 23, the recommendation is that there should be an observation period of two months which will end in month 25. During the observation period, partners are expected to make weekly observations of the usability of the toolkit in relation to whether there is an upsurge or dwindling of usage of the toolkit by both girls and boys. This will help us assess whether there are any gender differentials in the uptake of the toolkit and indeed whether the toolkit is proving effective and of interest. The observations will also help us gauge how the toolkits are influencing how science is communicated.

▪ Interviews

There will be two sets of interviews, one during the implementation of the toolkits to support the observations. The other set of interviews will be undertaken in order to understand some aspects of the development process as well as that of implementation.

• *First Set of Interviews*

In order to support the observations, we suggest that the first set of interviews be conducted by 5 out of the 9 third parties with a view to cultivating more in-depth knowledge of how users are engaging with the toolkits. The third parties will conduct interviews with stakeholders from the hubs that they will be observing. As indicated in section 6.2, the third parties have a choice of stakeholders who might include head teachers, teachers, museum coordinators, industry representatives. In this case, we propose that the first set of interviews by the third parties be from the following countries: Ireland, Austria, UK, Poland and France (Universcience).

The expectation is that aside from conducting observations with teenagers, third parties will conduct interviews with the said select stakeholders from the same hubs where the observations will be conducted. The aim of the interviews will be to:

- i. Look at the success or challenges of the toolkits implementation and adoption.

- ii. Assess whether the toolkits are improving how girls view science and other STEM subjects.
- iii. Understand how toolkits are improving the way girls view science, assuming that this will be the case.
- iv. Understand the level of usage by teenage girls and boys.
- v. Evaluate the activities of the hubs for instance by comparing and contrasting the activities of the hubs in order to assess whether certain activities draw more girls and encourage their engagement with STEM.
- vi. Gauge the usefulness of the activities that are implemented within the hubs.
- vii. Gauge the usefulness of the guidelines that will be provided to toolkit users.
- viii. Ascertain the interests of the teenagers by looking at how the activities affect the interests of the teenagers.
- ix. Draw the interviewees into sharing what has been working well within their hubs and what could be improved and how.

- **Second Set of Interviews**

The second set of interviews will be conducted by the evaluators. These interviews will be conducted with 1 partner from WP4 (Toolkit development) and another partner from WP5 (Toolkit implementation). In addition, we will conduct interviews with the remaining 4 third parties from the following countries: Estonia, Serbia, Spain and France (L'Oreal).

The second set of interviews will be during the toolkit development and implementation phases. The aim of this second set of interviews will be to;

- i) Understand the successes and challenges of toolkits development.
- ii) Understand the successes and challenges of the toolkits implementation.
- iii) Assess the chosen toolkit activities in order to understand why they were chosen.
- iv) Consider whether there was any input from teenagers in the toolkit activities and whether such input was taken into consideration.
- v) Gauge how toolkit developers went about engaging the teenagers in toolkit activities.

The aims of both sets of interviews outlined in sections 6.2.2.1 and 6.2.2.2 should be seen as indicative interest areas from which the interview questions will be developed. Therefore, the specific question and observation guidelines will be sent to the respective third parties in month 18. This is because as indicated in section 6.1.1 the toolkit production would have started in month 15 and between this month and the expected start of the implementation in month 23, third parties would have had adequate time to sort out implementation logistics. As such, sending the specific questions at that time will mean that third parties will have enough time to prepare and subsequently implement the implementation process.

- **Translation and Transcription**

Since the observations and interviews will be conducted in different countries, we understand that they may be conducted in the local language of respective countries. Therefore, the expectation is that the third parties will translate the observation and interview results to English so that the evaluators are able to analyse the results and give appropriate feedback and recommendations.

In addition, it is expected that all observation and interview data will be transcribed by third parties prior to sending the collected data to the evaluators.

- **Document Analysis specifically through use of Basecamp**

We will use document analysis to gain an understanding of the multiple aspects of the Hypatia project such as background information and progress of the project through looking at some reports/deliverables.

- **Data Analysis**

Data analysis will be a mix of qualitative interpretive and quantitative analysis. Responses from questions will be subjected to a thematic data analysis where words and phrases with similar meaning will be grouped together into themes and presented narratively. As part of the analysis we will use Nvivo data analysis software to analyse qualitative data.

- **Activities Timeline**

- **Interim Report in Month 18**

i) Document analysis

- *Specifically focussing on aspects to do with hub creation, where the hubs will be, number of hubs.*

ii) Interviews by evaluators with 1 partner on toolkit development in **Month 12**.

- *This will be during the piloting phase of the development task*

iii) Send out the observation and interview protocol guidelines in **Month 18**

- *This will give third parties enough time to sort out logistics e.g. who they will interview, where they will observe.*

iv) Submission of Interim Report in **Month 18**

- *The report will include;*

- *Findings from document analysis and 1 partner interview*

- *Observation and interview protocols with detailed guidelines and questions*

- **Final Report in Month 36**

i) Third parties begin observations in **Month 25 to 27**

- *Third parties to observe on a weekly basis resulting in 8 observations each.*

ii) Third parties conduct Interviews (see selected 5 third parties in section 6.2.2.1) between **Month 25 to 27**

iii) Third parties send evaluators collected observation and interview data in **Month 29**

iv) Submission of Final Report in **Month 35**

- **Conclusion**

This report has covered the evaluation strategy for the Hypatia project. It has described the objectives of the evaluation strategy which has touched two evaluation approaches. This has been in addition to evaluation methodology covering the evaluation design and evaluation guidelines. In addition to this, the report has given a detailed evaluation framework covering such elements expected impact, suggested approaches, evaluation criteria and indicators of success. Finally, the report has given a timeline for the next evaluation activities that are part of the evaluation strategy.

- **References**

- [1] Campaign for Science and and Engineering (CaSE), “CaSE Diversity in STEM report,” King College London, 2014.
- [2] European Commission, “Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation,” Directorate -General for Research and Innovation, Brussels, 2012.
- [3] C. Leaper, T. Farkas, and C. S. Brown, “Adolescent Girls’ Experiences and Gender-Related Beliefs in Relation to Their Motivation in Math/Science and English,” *J. Youth Adolesc.*, vol. 41, no. 3, pp. 268–282, Mar. 2012.
- [4] OECD, *The OECD Innovation Strategy Getting a Head Start on Tomorrow: Getting a Head Start on Tomorrow*. OECD Publishing, 2010.
- [5] R. Wilson, “Wilson, R. 2009. The Demand for STEM Graduates: some benchmark projections,” Council for Industry and Higher Education., London, 2009.
- [6] Research Councils UK, “Evaluation: Practical Guidelines - A guide for evaluating public engagement Activities.” [Online]. Available: <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/publications/evaluationguide.pdf>. [Accessed: 17-Nov-2014].
- [7] G. Rowe and L. J. Frewer, “Public Participation Methods: A Framework for Evaluation,” *Sci. Technol. Hum. Values*, vol. 25, no. 1, pp. 3–29, Jan. 2000.
- [8] C. Chess and B. B. Johnson, “Organizational Learning about Public Participation: ‘Tiggers’ and ‘Eeyores,’” *Hum. Ecol. Rev.*, vol. 13, no. 2, p. 182, 2006.
- [9] D. L. Stufflebeam and A. J. Shinkfield, *Evaluation Theory, Models, and Applications*. San Francisco: Jossey-Bass, 2007.
- [10] J. Bennett, *Evaluation Methods in Research*. Bloomsbury Publishing, 2003.
- [11] X. Song, N. Letch, L. Xu, and K. Huang, “The utilisation of IS evaluation a HIS case study,” in *Information Technology in Medicine and Education (ITME), 2012 International Symposium on*, 2012, vol. 2, pp. 960–964.
- [12] M. Scriven, *Evaluation thesaurus*. Sage, 1991.
- [13] S. Joss, “Evaluating consensus conferences: Necessity or luxury,” in *Public participation in science: The role of consensus conferences in Europe*, 1995, pp. 89–108.
- [14] A. Tuominen, T. Järvi, K. Hyytinen, E. Mitsakis, M. E. Lopez-Lambas, L. L. Paix, J. van der Waard, A. Binsted, and A. Sitov, “Evaluating the achievements and impacts of EC framework programme transport projects,” *Eur. Transp. Res. Rev.*, vol. 3, no. 2, pp. 59–74, Jul. 2011.

- [15] E. Arnold, "Understanding long-term impacts of R&D funding: The EU framework programme," *Res. Eval.*, vol. 21, no. 5, pp. 332–343, Dec. 2012.
- [16] Research Councils UK, "Evaluation: Practical Guidelines - A guide for evaluating public engagement activities," 2014.
- [17] G. Rowe and L. J. Frewer, "Evaluating public-participation exercises: a research agenda," *Sci. Technol. Hum. Values*, vol. 29, no. 4, pp. 512–556, 2004.
- [18] J. Abelson, P.-G. Forest, J. Eyles, P. Smith, E. Martin, and F.-P. Gauvin, "Deliberations about deliberative methods: issues in the design and evaluation of public participation processes," *Soc. Sci. Med.*, vol. 57, no. 2, pp. 239–251, 2003.
- [19] J. Abelson and F.-P. Gauvin, *Assessing the impacts of public participation: Concepts, evidence and policy implications*. Canadian Policy Research Networks Ottawa, 2006.
- [20] F. Merx, I. van der Weijden, A.-M. Oostveen, P. van den Besselaar, and J. Spaapen, "Evaluation of Research in Context A Quick Scan of an Emerging Field," *Den Haag Rathenau InstituutERIC*, 2007.
- [21] P. Mickwitz, "A framework for evaluating environmental policy instruments context and key concepts," *Evaluation*, vol. 9, no. 4, pp. 415–436, 2003.
- [22] J. Fitzpatrick, J. R. Sanders, and B. R. Worthen, *Program Evaluation: Alternative Approaches and Practical Guidelines* ., 4th ed. San Francisco: Allyn and Bacon, 2010.

Annex 2: Hypatia Evaluation Interim Report

Deliverable

Work package Evaluation
number:

Deliverable number: 2

Contributors: Kutoma Wakunuma and Tilimbe Jiya

Institutions: De Montfort University

Revision Date: 22/02/2017

Status: Final

Executive Summary: 49

Introduction: 49

Observation and Interview Guidelines: 50

Interim Evaluation Results: 50

Interview Analysis 50

Document Analysis 52

Feedback 53

Our plans 53

Annex 54

Annex 1: Hypatia Evaluation- Observation and Interview Guidelines 54

Annex 2: Inception Report: Evaluation Strategy for Horizon 2020 Project HYPATIA 62

i. Executive Summary:

In this interim evaluation report, we present the observation and interview guidelines that will be used by Hypatia partners in assessing the effectiveness of the toolkit that will be used by the different partners and Third parties across Europe. In addition, we present the findings from an evaluation interview that was conducted with a representative from the partner responsible for Work Package (WP) 4 - Toolkit development (Museum of Science and Technology Leonardo da Vinci). The interview was conducted in month 15 of the project. Furthermore, we present findings from a documents analysis. The document analysis focuses on the deliverables and reports relating to toolkit development and hub implementation. Specifically, the analysis looked at Deliverables 1.2, 3.2, 3.3, 4.1 and 4.2. The document analysis was conducted in order to have a holistic understanding of the toolkit development and hub implementation. The outlined Deliverables went some way in giving an insight into aspects of the said toolkit development and hub implementation. Towards the end of this interim evaluation report, we provide feedback and remedial action based on the evaluation findings thus far and information about the steps to come in the next months.

ii. Introduction:

Evaluation involves learning new knowledge through gathering information, making credible conclusions or judgements that can be used in decision making and communicating the findings to an audience. It includes looking at the quality of the content, the delivery process and the impact of the activity on the participants. Evaluation provides an opportunity to determine whether the aims and objectives of an undertaking have been achieved or not. It also provides an in-depth understanding of the potential impact of the project, therefore, allowing the project to reflect as well as improve on the processes and outcomes should need be.

In light of the Hypatia project, the evaluation process will include aspects related to the usability of the toolkit, toolkit user numbers, the effectiveness of the toolkit on the way science is communicated, on increasing participation and interest of teenagers in STEM, on contributing to the promotion of careers in STEM involving boys and girls, on encouraging collaboration among

different stakeholders as well as on changing gender stereotypes. In addition, the evaluation process will include a look at elements of engagement with the toolkit. In this phase of our ongoing evaluation of the Hypatia project, we developed observation and interview guidelines which will be used by the third parties that were identified in our strategy document (see Annexe of D7.1). We focus on third parties because the third parties are assigned with the implementation of the STEM toolkit modules whereas the main museum partners have additionally developed the first versions of the modules of the toolkit. The fact that the third parties will be implementing the toolkit modules in their countries makes them ideal entities to carry out observations and interviews.

Additionally, during the same evaluation period, we conducted an interview with a representative of MUST (WP4 leader) to look at the progress of toolkit development. The interview focused on two tasks within WP4 involving the development of and piloting of modules. The interview was focussed on the following aspects;

- a. Consideration of whether there was any input from teenagers in the development of the toolkit modules.
- b. Understanding the successes and challenges of module development
- c. Assessing appropriateness of the chosen modules.

Equally important, we conducted a document analysis that was focussed on the deliverables and reports of activities that were within the timeline between the inception report and the interim report, specifically the documents related to toolkit development and hub coordination and stakeholder engagement. The deliverables that were analysed included;

- a. Deliverable 1.2: 1st Periodic report Hypatia
- b. Deliverable 3.2: Strategic plan for stakeholder engagement
- c. Deliverable 3.3: An outline of the distribution and composition of the hub
- d. Deliverable 4.1: Set of Developed Modules
- e. Deliverable 4.2: Pilot Reports

iii. Observation and Interview Guidelines:

The observation and interview guidelines include several sections which will be used to assess the usability of the toolkit, toolkit user numbers, the effectiveness of the toolkit and stakeholders engagement with the toolkit modules. The observation and interview guidelines are included in Appendix D7.1 of this interim report. As indicated in our Strategy document in Section 6.2.1, the observation and interview guidelines should be sent to some of the third parties implementing the toolkit.

iv. Interim Evaluation Results:

In the interim, we conducted an interview with the partner who was responsible for Work Package 4 - Toolkit Development. In line with the project timeline, this meant that we were able to interview the partner on the following two tasks: Task 4.1 - Development of modules and Task 4.2 - Piloting of modules. We present the findings from the interview in Section 4.1 below.

- a. Interview Analysis

The partner responsible for the tasks was MUST (Italy). The aim of this work package was to develop a toolkit consisting of 15 different modules. The 15 modules were developed across three different contexts (5 for schools, 5 for museums and 5 for industries). They were subsequently divided into three categories, covering STEM content, STEM professions and Gender Inclusiveness. Although the content was not focussed on specific areas of STEM, it covered aspects related to gender and

STEM careers. Therefore, the STEM content was designed to deal with gender inclusivity among teenagers as well as to influence the choice of STEM studies which may subsequently lead to STEM careers. STEM professions involved the encouragement of STEM careers. The modules covered STEM professions through posters as well as games while gender inclusiveness covered guidelines on how facilitators dealing with the toolkit could look at ways of including gender in the modules. All museum partners collaborated with MUST in developing the modules for the toolkit and developed each 3 modules (1 for respectively industry, museum, school)

In addition they were supposed to give feedback and refine 3 modules originally developed by another museum partner. The final output from the tasks in WP 4 was 15 modules. -

In order to understand the process of the development of these modules, specifically towards the aim of the WP, the interview with MUST focussed on four aspects. An analysis of the interview about the four aspects is presented below.

i. Consideration of whether there was any input from teenagers in the development of the toolkit modules.

It is evident from the interview conducted that teenagers played a key role in testing the modules. Firstly the activities for the modules were chosen and then teenagers were invited to test and give feedback. It appears that the different activities were chosen with gender inclusiveness in mind. Initial indications suggest that teenagers were keen to meet professionals in STEM. If this is upheld, it will contribute to meeting some of Hypatia's expected impacts including an *Increase in the participation and interest of girls in STEM* and in *Contributing to the Innovation Union Objectives (better match skills to available jobs)*. In addition, they were happy to discuss issues related to gender.

However, what the interview suggests is that although teenagers were happy to discuss matters related to gender, there was a visible challenge with some facilitators who had no educational background in gender awareness to easily discuss gender issues. This means that the more facilitators were aware of gender, the easier it was to incorporate and discuss it with teenagers. However, the less aware facilitators were about gender, the more difficult it was to incorporate and discuss it with teenagers. The implication is that this could have an impact on the promotion of the activities when it came to gender inclusiveness.

Secondly,, the partner mentioned the use of an observation grid that was used as part of the engagement process. This observation grid looked at some areas which included;

- Emotions that were shown by participants during the activities.
- The cooperation between boys and girls during the activities.
- The level of involvement in plenary work and discussions during the activities.
- The facilitation of the activities looking at setting up conversations.
- The degree of clarity in the activity guidelines.

ii. Understanding the successes and challenges of module development

In measuring the success of the modules, the partners developed a feedback system i to see whether the chosen modules worked or not. Each partner tested their own 3 modules as well as 3 modules from other developing partners. All partners used the observation grid mentioned above covering a range of aspects from emotions, boys and girls working together, discussions with professionals as well as observations on gender stereotypes and where necessary was transferred

into local languages. The essence of the grid was to collect feedback and then fine-tune the module. However, the successes of the modules were not without challenges. These included:

- a. Lack of experience in developing modules that relates to industry context compared to the school and museum contexts.
- b. It was difficult to decide on the nature of modules to include in the toolkit.
- c. There was a misunderstanding and lack of clarity on what was suitable STEM content. The relationship between the module content and STEM professions or careers was not clear.
- d. The aspect of developing a module that was gender balanced and inclusive without the need to address gender directly or indirectly.

Despite the above challenges, from the interview, it is evident that there is good collaboration between the partners who have keenly shared their experiences in getting the modules up and running.

iii. Assessing the chosen module activities in order to understand why they were chosen

The interviewee mentioned that 15 modules were developed under WP4. These were divided into three categories across the contexts which are Museums, Schools and Industry. A list of modules that were developed was provided by the interviewee and this partner later sent documentation detailing the chosen modules. It was established from the interview that the modules were chosen because of;

- the experience of the partners that were involved with their development
- the relevance of the module to the three different contexts and categories that the project was dealing with

Having looked at the interview results, it is evident that there was great emphasis put on ensuring teenagers active role in testing the modules. It is also clear that teenagers proactively participated in the toolkit development. Additionally, there was a robust feedback system put in place to ensure that the chosen module activities were relevant. Further, the interview revealed that the modules were chosen mainly due to the experience of the partners involved in their development as well as to the relevance of the modules to the three categories of museum, school and industry. Having looked at the above, a further analysis was conducted through document analysis outlined below.

b. Document Analysis

The third element of this interim report involves a document analysis. We used document analysis to gain an understanding of the progress of the project through looking at some reports and deliverables that specifically relate to the development and implementation of the toolkit. The deliverables that we looked at were;

- Deliverable 1.2: 1st Periodic report Hypatia

The deliverable afforded us overview of the projects progression between the period of August 2015 to July 2016. The indication was that WPs between that period were progressing well.

- Deliverable 3.2: Strategic plan for stakeholder engagement

This deliverable covered aspects of stakeholder engagement with particular focus on the establishment of hubs and the recruitment of relevant stakeholders.

- Deliverable 3.3: An outline of the distribution and composition of the hub

Deliverable 3.3 gave an insight into national hub members outlining main partners and third partners. With regard to this deliverable the main role of the partners was to give guidance and recommendations on the establishment of national hubs which third parties could then use to establish the hubs in their respective localities.

- Deliverable 4.1: Set of Developed Modules

The deliverable highlights the modules developed by the 5 developing Museum partners (BSMJ, Experimentarium, MUST, NEMO and Universcience)

- Deliverable 4.2: Pilot Reports

From April to September 2016, the 5 museum partners conducted tests on the viability of the modules that were developed. This deliverable outlined outcomes of the conducted tests.

From the analysis of the documents, we established that there is a broad range of expertise and stakeholders that the partners involved and intends to involve especially in the distribution and composition of hubs. From the documents, the project is involving a range of stakeholders from industry, local and national authorities, research institutes, schools, informal education institutes and gender experts. In addition, the roles or positions of the people involved in the hubs is diverse which includes heads of section, consultants, professors, engineers, researchers, teachers and industry representatives. This is an encouraging development towards achieving the objectives of the project and meeting the impacts that Hypatia is working towards.

With regards to piloting, looking at D4.2, the partners seem to have followed a rigorous process although we could not locate some vital information such as distribution of boys and girls that were involved in the activities. The documents have given the target audience include age, the number of participants, the number of facilitators and type of audience but we missed an important insight in the proportions of boys and girls involved in the activities to gauge gender inclusiveness.

v. Feedback

- i. There should be on-going consultation with industry players on what to include in the toolkit with regards to STEM careers or professionals. Although, the toolkit is final, we would recommend on-going consultation to reflect on future developments. STEM is not static therefore in order to have a long-term sustainable impact, the toolkit should be reviewed and amended accordingly.
- ii. There should be a link between the modules that are offered in the three different contexts and a successful uptake of STEM careers. This can be done by ensuring that all the modules consistently address the importance of taking up STEM subjects which have a direct impact on STEM careers
- iii. It is encouraging to learn that partners have recognised the need for a toolkit that consists of gender balanced and inclusive activities. This fits well with the overall aims of the Hypatia project.
- iv. The project is doing well in involving a wide range of stakeholders from different backgrounds. This will have a positive impetus on the work that Hypatia is doing towards meeting its aims.
- v. From the document analysis, it is clear that during the hub creation, a well thought out strategy was in place to ensure that there was evident teenage/gender consideration put in place.

vi. Our plans

With regards to our ongoing evaluation of the project we will interview the partner responsible for toolkit implementation (WP5 Leader, EXPERIMENTARIUM) and some of the third parties. For more details please see attached evaluation strategy in Annex 2.

vii. Annex

a. Annex 1: Hypatia Evaluation- Observation and Interview Guidelines



viii. Observation and interview instructions

This observation and interview guideline is intended to be used by third party members of the project. The total number of third parties conducting the observations will be nine and include Ireland, Austria, UK, Poland, Greece, Estonia, Serbia, Spain and Sweden. We recommend that each third party conducts at least 1 observation at one of the hubs they are hosts at. In addition, we recommend that the third parties choose a figure who can be able to conduct the observation in a more neutral way in order to avoid bias. Further to conducting observations and as a way of complementing and strengthening the observations, an additional set of interviews will be carried out by at least 5 third parties who will include Ireland, Austria, UK, Poland and Greece. These interviews will be with any of the following stakeholders involved with ensuring that the implementation of the toolkit comes to fruition in their respective institutions: head teachers, teachers, museum coordinators, industry representatives. In this respect, the third parties will choose who they prefer to interview.

1 Introduction

Evaluation involves learning new knowledge through gathering information, making credible conclusions or judgements that can be used in decision making and communicating the findings to an audience. It includes looking at the quality of the content, the delivery process and the impact of the activity on the participants. Evaluation provides an opportunity to determine whether the aim of the activity was achieved or not. For instance, in light of the Hypatia project, the evaluation process will assess aspects related to the usability of toolkits, toolkit user numbers, the effectiveness of the toolkits and the engagement with the toolkit activities, thus the need for the evaluation guidelines. The guidelines include several sections as covered below. The expectation is that these sections ought to be covered by the parties undertaking the evaluation. A description of what ought to be included is highlighted in the respective sections.

2 Objectives of the Module

Provide an indication of whether the objectives in the module have been met or not? If yes, please state how they have been met and if not please give insight into why they have not been met.

3 Conducting Observations

The collection of data based on observations is an example of qualitative evaluation tool which involves watching, recording and analysis of behaviour as it occurs in a ‘natural’ setting. Observation enables the evaluators to understand the participant (s) engagement with specific tasks and define the key issues that may be followed up in the interviews. As the development and implementation of STEM toolkits is central to this project, particularly in terms of engaging teenagers in STEM, it becomes imperative to observe how teenagers are engaging with engaging with the activities in addition to how institutional hubs are supporting the toolkit implementation and subsequent adoption. As such, a further guideline would be the completion of the following sections specifically related to observed elements within the hub.

3.1 Intended outcome of the module

Give an indication of the intended outcome of the module

3.2 Toolkit guidelines

Please give an indication of the effectiveness of the toolkit guidelines provided as far as facilitation and gender inclusion is concerned. Were there any challenges in:

- a. *The facilitation of the guidelines? If they were, please highlight what these challenges were.*

- b. *Ensuring gender inclusiveness? If they were, please highlight what these challenges were*

3.4 Activities

Please outline the module activities that were used at the hub.

3.5 Indicator of success

Please use the following success criteria and indicators of success to assess the modules and indicate the outcome in the results column:

Deliverable

Success criteria	Indicator of success	How to measure	Result
Gender Inclusiveness	The level of inclusiveness on the four different perspectives of gender inclusiveness: the individual level, the interactional level, the institutional level and the societal/cultural.	<p>Individual level:</p> <ul style="list-style-type: none"> i. Number of boys and girls ii. How many boys and girls were engaging with the activity iii. Were boys and girls participating in different or similar ways? iv. If different, give examples of the different ways boys and girls were engaging with the activity v. <p>Interactional level</p> <ul style="list-style-type: none"> vi. Ensure that the activity has a balanced approach to participants' learning preferences, i.e. includes thinking tasks, motor skill tasks, and value-related tasks. vii. Did the hub apply different forms in which teenage boys and girls could interact? viii. If they did, please outline the different forms of interaction that were observed 	

		<ul style="list-style-type: none"> ix. Were there different roles taken by boys and girls during the activity and did these have equal status and rotated between girls and boys? x. If No, please outline the specific roles that were taken by boys and those that were taken by girls xi. Were the different roles given equal status in order to avoid othering or subordination xii. xiii. Do the activities reflect a variety of gender neutral or gender equal personalities e.g. were facilitators at the hub of equal gender balance or did the game activities reflect a gender balanced protagonists xiv. <p>Institutional level</p> <ul style="list-style-type: none"> xv. xvi. Was the physical learning environment appropriate for the planned activities i.e. was it appealing to both girls and boys in 	
--	--	--	--

		<p>terms of design, materials, availability of role models.</p> <p>xvii. If yes, please give examples of how this was achieved</p> <p>Societal/Cultural level</p> <p>xviii. After having observed the activities, do you think there are any discernible societal/cultural impacts that may result?</p> <p>xix. If yes, please describe them</p> <p>xx. If no, please indicate why you think there are no discernible impacts</p>	

4 Conducting Interviews

Interviews are another type of evaluation tool which is used to understand and obtain useful information from the participant's experiences. To get an in-depth understanding of the implementation and adoption of the toolkits, interviewing relevant stakeholders in addition to observations becomes a necessity. The expectation is that aside from conducting observations with teenagers, third parties conduct one interview with a stakeholder of their choice from the HUB involved with the coordination or facilitation of the hub. While all third parties are conducting the observations, only 5 of the third parties (Ireland, Austria, UK, Poland and Greece) conduct the interviews. .

4.1 Interview Questions

- ix. How useful were the guidelines on facilitation and gender inclusiveness in ensuring that the activities were effectively implemented?
- x. Can you give an indication of how many stakeholders such as head teachers, teachers, museum coordinators, industry representatives were involved in the implementation of the toolkit and ensuring that it came to fruition?
- xi. Can you give an indication of how they came to be involved?
- xii. What was the initial response of teenagers to the activities i.e. first impression of the activities?
- xiii. In your point of view, how do you think these activities affect the interests of the teenagers?
- xiv. Do you think these activities have encouraged girls to engage with STEM and have improved their view on STEM subjects? If yes, please elaborate on how their views are improved?
- xv. What risks/complications/challenges did you encounter while conducting these activities e.g. availability of resources, availability and understanding of the staff members, overall management of the activity?
- xvi. If there were any risks involved, how were they tackled? What measures were put in place to minimise those risks?
- xvii. Were there any challenges faced when these activities were conducted or during implementation or adoption?
- xviii. Describe the strategies which were used to implement the activities?
 - a. Did they work well?
 - b. If there were any challenges, kindly highlight them while describing how they were overcome?
- xix. What improvements can be made to the activities for future use?
- xx. Give an account of teenagers' reflection on awareness and attitude towards STEM. This could include feedback on STEM information accessed during the activity.
- xxi. At the end of the activity take a feedback survey on the stakeholders' expectation, learning outcomes and willingness to participate again if the activities were used again.



- a. Annex 2: Inception Report: Evaluation Strategy for Horizon 2020 Project HYPATIA



Inception Report: Evaluation Strategy for Horizon 2020 Project HYPATIA

Dr Kutoma Wakunuma

Tilimbe Jiya

November, 2015

Contact details for corresponding authors:

Dr Kutoma J. Wakunuma

Centre for Computing and Social Responsibility

De Montfort University

The Gateway

LE1 9 BH, Leicester, United Kingdom

kutoma@dmu.ac.uk

Mr Tilimbe Jiya

Centre for Computing and Social Responsibility

De Montfort University

The Gateway

LE1 9 BH, Leicester, United Kingdom

tilimbe.jiya@dmu.ac.uk

1 Executive Summary

It is well documented that there are more boys than girls taking up STEM subjects. However, it is now a general consensus that in order to overcome societies' challenges, both boys and girls ought to be given and have the same opportunities when it comes to education, jobs and other economic empowerment activities, which is consonance with Hypatia's aim and expected impact. In order to ensure that this aim is met and the expected impact is realised effectively, the project needs to be evaluated. The evaluation is important not only to provide evidence about the efforts and quality of the Hypatia's activities, but also to improve outcomes in future. This inception report describes the evaluation strategy for evaluating the Hypatia project. The evaluation strategy will outline the processes of evaluation which will be used in order to assess the objectives of the project e.g. implementing modular toolkits that promote engagement of teenagers in STEM in a gender inclusive manner. Ultimately, the proposed strategy will assess whether the expected impact of the project has been reached. The strategy involves two evaluation approaches which include both formative and summative approaches. These will be employed to meet the objectives of the strategy which involve assessing the impact of the project by determining how toolkits are being implemented, looking at evaluation criteria and gauging indicators of success in terms of the way science is communicated. The expectation is that this will go a long way towards looking at indicators of success that will be used to take into account the expected impacts which include:

- increase in participation and interest of girls in STEM
- encouragement of sustainable collaboration among schools, science museums, research centres on gender equality and science
- contribution towards European Research Area objectives and
- Innovation Union Objectives.

As part of the strategy, the inception report also presents an evaluation methodology that consists of an evaluation design that will include a set of evaluation guidelines to be used in conducting the evaluative activities proposed in this evaluation strategy. Furthermore, the evaluation design will include an evaluation framework which will be used in assessing findings from the evaluation. In addition, the evaluation design will also include approaches that will be employed in data collection such as observation, interviews and document analysis. Further, the inception report touches on a data analysis approach for analysing results from the evaluation activities. The report then concludes with an outline of activities that are planned for Hypatia's evaluation within a predefined timeline.

2 Background

The evaluation team is part of the Centre for Computing and Social Responsibility (CCSR) at De Montfort University (DMU), Leicester. The centre is one of the leading research centres in the UK and its research areas include; responsible research and innovation (RRI), civil society organisations in research, stakeholder engagement and involvement, ethics (computer and applied ethics), information communication technologies (ICTs), gender in technology and project evaluation. The centre has been involved in a number of EU funded projects at an international level such as Network Analysis of CSO Participation in Research Framework Programmes, Responsible-Industry, Responsibility, GREAT (Governance for Responsible Innovation), CONSIDER (Civil Society Organisations in Designing Research Governance), ETICA (Ethical Issues of Emerging ICT Application) and SATORI (Stakeholders Acting Together On the ethical impact assessment of Research and Innovation) project. The team has wide experience in evaluation methodology and implementation.

3 Introduction

Over the years research has shown that the way science is communicated to young people in and out of school is not yet gender inclusive. As a result, this has left both boys and girls having little knowledge and awareness of the range of careers in science, technology, engineering and mathematics (STEM), and the skills that are relevant for those career paths. This has affected the number of teenagers who are pursuing STEM related careers. STEM subjects are important drivers and cornerstones for development therefore, increasing women's as well as men's pursuit of STEM-related jobs is considered critical for any nation to remain competitive in the global economy[1]. Although this is an important aspect, girls remain underrepresented in many STEM fields. This is down to social influences which include the relative degrees of encouragement that girls may experience to do well in STEM and non-STEM subjects. In addition to that, personal influences such as gender-related variations in self-schemata and attitudes shape girls' motivation in STEM or non-STEM domains [3]. Without science and technology, social and economic development would be difficult to achieve. This aspect is highlighted in several H2020 programmes that indicate that in order to overcome societal challenges, STEM must be at the heart of Europe [4].

As part of the EU's development agenda for a better Europe, the need to change the state of affairs and with regards to STEM and gender inclusivity has become more vital than ever before. It is therefore essential to ensure that STEM careers become increasingly oriented towards society's needs and opening up new dimensions in terms of the skillsets required in a gender inclusive way. In a study by Wilson (2009), it was found that STEM fields constitute 9 of the top 10 college degrees leading to the highest paying occupations. The study suggests that if girls and boys are equally encouraged to pursue STEM related subjects, they can equally achieve equal economic empowerment for both which is positive for overall societal benefit [5]. Therefore, considering proportion of teenage girls pursuing STEM subjects, it is fundamental to expose more girls to the variety of STEM-related careers, empower them to make connections that develop their lives and their own skills.

In order to realise this, there is a need to engage teenagers in STEM in a gender-inclusive way, and addressing the attitudes of STEM education professionals towards more gender-inclusive practices. This is the mission that Hypatia is aiming to achieve by bringing together groups of stakeholders from science centres and museums, schools, research institutions and industry to collaborate with gender experts and teenagers in hubs and co-develop the content and co-organise activities that promote girls awareness of STEM careers. Work being carried out by Hypatia is important in ensuring that there are equal opportunities for both boys and girls in getting enough information of the variety of STEM career pathways they can follow. This is in line with Europe's agenda on RRI of which one of its pillars is 'gender inclusivity'[2]. In line with this, the project has therefore planned to develop and implement expert toolkits in schools, museums, research institutions and industry and conduct a set of seminars across 14 European countries bringing together head teachers, museum professionals, researchers, industry professionals and other related stakeholders to discuss how the toolkits should be implemented. Hypatia aims at ensuring an effective adoption of the toolkit of activities that will focus on ways of communicating STEM to empower teenage girls in exploring the range of skills needed for a variety of STEM studies and careers. The project will also develop guidelines for engaging teenagers in STEM in a gender inclusive way.

Considering the efforts and goal(s) of Hypatia, it is important that the project is evaluated. The project needs to be evaluated against its aim and objectives in order to ensure that the results have been reached. This is where DMU as evaluators will implement the strategy proposed in this inception report. The evaluation strategy is directly tailored to Hypatia and focuses on the methodology for evaluating the outcomes and impact of the project. The evaluation strategy will ensure that Hypatia's work is meeting its intended objectives of bringing lasting change to attitudes towards STEM and how it is communicated in museums, schools and industry through engaging with teenage girls in STEM across Europe. The evaluation that we propose to undertake for the Hypatia project will be both formative and summative. Using the formative evaluation we will be able to contribute to the refinement of toolkit implementation by identifying weaknesses or areas

for improvement [8] through feedback. On the other hand, use of summative evaluation will enable us to evaluate the quality of the project's outputs and outcomes in order to assess its success [8] for example in meeting its stated objectives and expected impacts against a set of success indicators that are laid down in a pre-defined evaluation framework.

In this inception report we will introduce the objectives of the evaluation strategy then we will move on to a brief discussion of evaluation, touching on the two evaluation approaches that will be employed in Hypatia's evaluation. Following this, we will then introduce the methodology that will be used in the evaluation, consisting of the evaluation design proposed for Hypatia, where we will cover the evaluation guidelines that will assist in directing the evaluation process and our proposed evaluation framework. Subsequently, we will then discuss the evaluation approaches in detail before concluding with the timeline for the next evaluation activities.

4 Objectives of the Evaluation

In order to effectively evaluate Hypatia, the evaluation strategy will aim at meeting the following 3 objectives:

1. Assessing Impact of the Project

This will entail assessing Hypatia's expected impact (as described in the DoA) in order to understand whether the proposed activities are meeting the impact or not.

2. Looking at Evaluation Criteria

This will cover elements that will assist in determining whether the expected impact has been met.

3. Gauging Indicators of Success against Expected Impact

This will involve using a set of indicators of success to gauge whether Hypatia has reached its expected impact or not.

5 Evaluation

The term evaluation has many meanings depending on context and purpose of the evaluation, however all the definitions have elements of credibility[9]. Evaluation involves learning new knowledge through gathering information, making credible conclusions or judgements that can be used in decision making and communicating the findings to an audience[10]. Evaluation is paramount because it acts as a control mechanism [11] that ensures that strategic benefits of undertaking project are realised. Broadly speaking, evaluation can be defined as "the process of determining the merit, worth and value of things" [12], and can be used to describe many "different kinds of judgments, from informal assessment that relies on intuition or opinion, to well-defined and systematic research that makes use of social science research methods" [13]. In light of Hypatia project, evaluation will focus on the "design, implementation and effectiveness" of the project's toolkit activities [14], as well as outputs where we will specifically be looking at the such elements as adoption of the toolkits e.g. from reports/deliverables and attainment of expected impacts [15].

In evaluating Hypatia, we will use both formative and summative approaches to evaluation. These 2 distinct approaches are discussed in the following sections.

5.1 A Formative Approach

A formative evaluation approach examines the quality of procedures or methods of stakeholder engagement while they occur and their outcomes contribute to the refinement of the engagement

processes by identifying weaknesses or areas for improvement prior to the project's conclusion, thereby acting as a feedback mechanism or 'double loop' to refine project activities [18]. In relation to Hypatia, with this approach we will look at assessment of deliverables and adoption of modular toolkits among others. For instance, we will assess where and how the third parties are implementing the toolkits in relation to the guidelines. This will give us an understanding of the usability of the toolkits, usefulness and effectiveness of the toolkit as well as who is engaging with the toolkits in terms of gender inclusivity.

5.2 A Summative Approach

A summative evaluation approach assesses the outcomes of procedures or methods [7], [19]–[21], evaluates the quality of engagement project's outputs and outcomes in order to evaluate its success [7]. In terms of Hypatia, we will use this approach to evaluate whether the project is or has met its stated objectives, expected impacts against a pre-defined evaluation framework [see section 6.1.2]. For example, we will evaluate if the activities have been well adopted and are established within their respective hubs. This will help us determine the overall impact (value or significance) of the Hypatia project.

6 Methodology

6.1 The Evaluation Design and Guidelines

The evaluation design will include a set of evaluation guidelines that will be used in conducting the evaluative activities proposed in this strategy. It will also include an evaluation framework which will be useful in analysing our findings. In addition to these, the evaluation design includes evaluation approaches that will be used in data collection and how the data collected will be analysed.

As there is a budget constraint, partners and third parties will play a key role in implementing the evaluation design in their respective countries. This will be with the guidance of the evaluators who will provide evaluation guidelines described in the following section.

6.1.1 Evaluation Guidelines

The evaluation guidelines expand on the above described evaluation design. The guidelines are mainly a guide to the consortium of the project who as indicated in the preceding section, will play a pivotal role in implementing the evaluation design. The use of the term 'consortium' in this document denotes the main partners of the Hypatia project and the third parties. We use both main partners and third parties in this document because we view all parties concerned in Hypatia as playing a relevant role in the evaluation of the project. The Hypatia project is a very interesting and important project, particularly in encouraging teenage girls to get involved in STEM. As such, there are very interesting, important and diverse elements that need and could be evaluated during the course of the project such as:

- feedback on development of toolkit activities to partners
- the implementation of the toolkit activities
- assessing the project impact and providing partners with a better idea of the value of their participation by tracking influence on the process
- improving the design of future related activities

However, as is indicated in section 6, due to the limited resources allocated for the evaluation of the project, evaluating the many diverse elements was always going to be a challenge. As such, it is important to prioritise what can be evaluated, how it can be evaluated, where and when. During the kick-off meeting held in Amsterdam between 2-4 November 2015, the evaluation team was fortunate enough to have received very useful feedback from members of the project that could

be incorporated into the evaluation, especially in terms of what we might evaluate, how we might evaluate, where and when we might evaluate. The main feedback which members of Hypatia were keen to realise with regards to evaluation, is for a concerted effort to have European wide evaluation results. This means that the evaluation of the Hypatia project will need to extend to as many European countries as possible while considering the resources available. In order for this to be a reality, Hypatia partners will have to assist with the implementation of the evaluation. Specifically, the most ideal group to assist in this evaluation would be the third parties who are assigned with the implementation of the STEM toolkits. This is because the third parties will be implementing the toolkits in hubs in their respective countries. In light of this, the evaluation team will send detailed guidelines to third parties in month 18. Month 18 has been chosen because the evaluation team has into consideration 2 facts:

- iii) That the production of the toolkits will be between month 15 and 18.
- iv) That the expected implementation of the said toolkits will between months 23 to 36.

As such, sending of the observation and interview guidelines in month 18 is ideal because it will give third parties enough time (5 months) to sort out logistics as described in the timeline. Elements of what will be covered in the guidelines are discussed in more detail in section 6.2. In the meantime, below is a discussion of the elected evaluation framework.

6.1.2 Evaluation Framework

The evaluation framework (see Table 3) has four specific elements that focus on the following:

6.1.2.1 Expected Impact

The Hypatia consortium answered the GERI-1-2014 call which has specific impacts that the project has to achieve. These expected impacts and how Hypatia intends to achieve them are outlined in Table 1 below:

Expected impact	How Hypatia will achieve this impact
6. Change the way science is communicated	The project will reach schools, informal learning organisations, research institutions and industry activities. It will produce sets of activities and guidelines about how to communicate science with their audiences ensuring the gender aspect is taken on board. The project will measure the number of institutions that adopt the new tools.
7. Increase the participation and interest of girls in STEM	The activities that will be used will be based on existing good practices related to gender in STEM. Now these activities will be implemented in much larger scale.

8. Encourage sustainable collaboration among schools, museums, research centres on gender equality and science	The creation of hubs in each participating country and the existing connection among partners will ensure a long-term collaboration, supported through the online community Scientix. The wealth of existing activities and the importance given to advocacy and sustainability will ensure a long and strong collaboration among stakeholders.
9. Contribute towards European Research Area objectives (increase female researchers in Europe)	By giving a different perspective on STEM careers to both young boys and girls today we contribute to the increase of female researchers in the future.
10. Contribute to the Innovation Union Objectives (better match skills to available jobs)	Discovering young people's skills in relation to current and future STEM careers is a particular focus of the proposal, with a direct relevance to gender. A set of activities will be directly devoted to this, including the Find Your Skills card game.

Table 1: Expected impacts and how Hypatia proposes to meet them

In order to ensure that the expected impacts have been achieved, an evaluation framework becomes imperative. This is important in order to understand whether the expected impacts in the call have been met or not. As such, the evaluator's will use the listed expected impacts as a check list to assess whether Hypatia has achieved its goal of meeting the GERI-1-call expected impacts.

6.1.2.2 Suggested Approach

These are the approaches that will be employed in the evaluation process. In this case, the proposed evaluation approaches will be in the countries where the toolkits will be implemented by third parties. These approaches will include observations and interviews. These are discussed in more detail in sections 6.2.1 and 6.2.2.

6.1.2.3 Evaluation Criteria

Evaluation criteria is the benchmark upon which Hypatia's work and agenda with regards to attitudes and gender inclusion of STEM will be assessed during the evaluation. Table 2 below describes the evaluation criteria by looking at what criteria will be used, its description and which expected impact is the criteria targeting.

Criteria	Description	Targeted expected impact from Table 1
Activities	What toolkit activities and guidelines were developed and delivered and to whom	1 and 5

Representativeness	How many boys and girls are using and engaging with the toolkits	2 and 4
Geographical coverage	How many countries Hypatia's activities have reached and how effective have they been implemented	3 and 4
Usability, usefulness and effectiveness	How usable are the toolkits and how useful and effective have the activities been in reducing the gender gap in STEM	1, 2 and 3

Table 2: Criteria for evaluating targeted expected impact

6.1.2.4 Indicators of Success

During evaluation, it will be necessary to have indicators of success which will show whether Hypatia has reached its impact or not. Using Table 1 which tabulates how Hypatia intends to achieve the expected impacts, the evaluation will endeavour to assess the outcomes of the tabulated potential achievements as indicators of success. Specifically, the evaluation will seek to understand;

- Whether Hypatia's toolkits have been adopted
- Who has adopted them (boys and girls/research institutions)
- How well the tools have been received by those that have adopted them
- What sort of activities have been developed
- How the activities have been received

Below is Table 3 summarising the evaluation framework.

Expected Impact	Suggested Approaches	Evaluation criteria	Indicator for success
Change in the way science is communicated	<ul style="list-style-type: none"> • Observations • Interviews • Document analysis 	<ul style="list-style-type: none"> • Activities • Representativeness (gender, geographical location) • Quality of information (gender sensitivity, relevance of information to realise gender inclusivity, use of language that is easy to understand so that STEM is not seen to be intimidating or seen to be too technical such that it 'scares' away young girls) 	<ul style="list-style-type: none"> • Adoption of toolkit (number of institutions and stakeholders using new tools). • Usefulness of Hypatia's proposed guidelines (what do the guidelines cover, who was involved in coming up with the guidelines, do guidelines address gender issues and how to overcome them in STEM)
Increase in the participation and interest of girls in STEM	<ul style="list-style-type: none"> • Observations • Interviews • Document analysis 	<ul style="list-style-type: none"> • Representativeness (gender proportionality) • Openness (What activities are open and inviting to girls to encourage involvement in STEM) 	<ul style="list-style-type: none"> • Degree of involvement (extent to which girls are integrated into STEM processes e.g. Hypatia's research activities) • Information accessibility (extent to which girls access appropriate information and materials on STEM) • Inclusion (extent to which girls are included and represented in STEM activities)
Encouragement in sustainable collaboration among schools, science museums, research centres on gender equality and science	<ul style="list-style-type: none"> • Interviews • Document analysis 	<ul style="list-style-type: none"> • Participation (who is collaborating, how representative is the collaboration, how relevant do collaborating partners feel their participation is and/or appreciated) • Resource accessibility (Exchange of information, information accessibility, contribution to required information) • Stakeholder representation) 	<ul style="list-style-type: none"> • Creation of hubs (How many hubs have been created, where have the hubs have been created and how are they impacting gender inclusivity in STEM, how are the hubs encouraging collaboration)

<p>Contribution towards European Research Area objectives (increase female researchers in Europe)</p>	<ul style="list-style-type: none"> • Interviews 	<ul style="list-style-type: none"> • Representativeness (gender representation) 	<ul style="list-style-type: none"> • Look at Hypatia’s proposed different perspectives on STEM careers (what are the different perspectives, how different are they from traditional perspectives e.g. attitudes to STEM)
<p>Contribution to the Innovation Union Objectives (better match skills to available jobs)</p>	<ul style="list-style-type: none"> • Interviews 	<ul style="list-style-type: none"> • Accessibility of information with regards to available skills 	<ul style="list-style-type: none"> • Look at Hypatia’s proposed activities that have a gender focus on building STEM skills (what were the activities and how were they directly relevant to gender)

Table 3: Evaluation Framework

6.2 Evaluation Approaches in Detail

As touched on in section 6.1.2, this section gives more details of the approaches that have been adopted for the evaluation. As evaluators, we suggest two approaches that can be used easily, efficiently and effectively for evaluating and therefore meeting the expected evaluation outcomes. These two approaches should not be too onerous to implement particularly by partners and third parties who are expected to play a significant role in implementing the evaluation process. The two elected approaches are observations and interviews. As the development and implementation of STEM toolkits is central to this project, particularly in terms of engaging teenagers in STEM, it becomes imperative to observe how teenagers are engaging with STEM activities in addition to how institutional hubs are supporting the toolkit implementation and subsequent adoption.

To get an in-depth understanding of the implementation and adoption of the toolkits, interviewing relevant stakeholders in addition to observations becomes a necessity. As evaluators, we recommend that each third party conducts at least 1 observation at one of the hubs they are hosts at when one of the modules is implemented. In addition to the observations and as a way of complementing and strengthening the observations, an additional set of interviews will be carried out by the third parties. These interviews will be with any of the following stakeholders involved with ensuring that the implementation of the toolkits comes to fruition in their respective institutions: head teachers, teachers, museum coordinators, industry representatives. In this respect, the third parties will choose who they prefer to interview. Further to this set of interviews will be another set of interviews specifically with some of Hypatia's partners and third parties. These will be conducted by the evaluators via Skype.

6.2.1 Observations

The suggested observations will be used to assess the usability of toolkits, toolkit user numbers, the effectiveness of the toolkits and the engagement with the toolkit activities. Therefore, the observation will look at the following;

- ii) How the toolkits will be used

This will help us to understand the usability of the toolkits. The main aspect that will be looked at here will be understanding whether the toolkits are usable for the intended users and by default, whether the users actually find the toolkits usable. This will indicate whether they are being effective for the purposes for which they are intended.

- vi) Who is using the toolkits

This will help us to gauge whether the intended users are actually using the toolkits. For instance, we would like to know whether the selected institutional hubs such as museums and schools are using them and whether by extension, teenage girls are engaging with the toolkits

- vii) How often are toolkits used

This will help us develop our knowledge of how often the users use the toolkits since being implemented.

- viii) How effective the toolkits are

This will help us assess how effective the toolkits are in encouraging teenagers to engage with STEM activities.

With regards to conducting observations there are different types of observations. One is where the observer is detached from the subject matter being observed while the other is where the observer is more involved in the subject matter being observed. In the



detached one, the observer merely looks at activities taking place without any meaningful interaction with the observed. In the other type where the observer is more involved there is meaningful interaction with the observed. This means that the observer will talk to the subject, perhaps be hands on with the activities in an attempt to cultivate more understanding of a situation. In our case, we recommend the latter in that as the third parties conduct the observations, they will also have some level of interaction with the subject being observed.

With the above in mind, we recommend that observations be conducted in the countries being represented by third parties. However, if a country has two representatives, the evaluators will only select one representative to implement the evaluation design. In addition, if there is more than one hub created in a country, we will leave it to the third party implementing the toolkits to decide the most ideal hub to conduct the observations. This is because the third parties will have the relevant knowledge of where the hub will have been created as well as how easily accessible it will be in their respective country. As evaluators, we would however recommend that third parties make observations in museums. This is because we feel that it would be easier to access participants who come to museums merely because museums are usually more open and do not necessarily need lengthy and protracted accessibility permissions compared to schools for instance. As a recap and in addition to the four areas that will need to be observed which are outlined above, third parties are expected to observe the following:

- iv) How the toolkits are implemented in relation to the toolkit guidelines: e.g. third parties will look at whether the guidelines were adequate enough for toolkit implementation to be robust
- v) How teenagers are engaging with modules
- vi) The number of girls and boys engaging with the toolkit activities over a period of time. Preferably this observation should be from the implementation month up to two months following the implementation. So for example, if a third party implements a module in month 23, the recommendation is that there should be an observation period of two months which will end in month 25. During the observation period, partners are expected to make weekly observations of the usability of the toolkit in relation to whether there is an upsurge or dwindling of usage of the toolkit by both girls and boys. This will help us assess whether there are any gender differentials in the uptake of the toolkit and indeed whether the toolkit is proving effective and of interest. The observations will also help us gauge how the toolkits are influencing how science is communicated.

6.2.2 Interviews

There will be two sets of interviews, one during the implementation of the toolkits to support the observations. The other set of interviews will be undertaken in order to understand some aspects of the development process as well as that of implementation.

6.2.2.1 First Set of Interviews

In order to support the observations, we suggest that the first set of interviews be conducted by 5 out of the 9 third parties with a view to cultivating more in-depth knowledge of how users are engaging with the toolkits. The third parties will conduct interviews with stakeholders from the hubs that they will be observing. As indicated in section 6.2, the third parties have a choice of stakeholders who might include head teachers, teachers, museum coordinators, industry representatives. In this case, we propose that the first set of interviews by the third parties be from the following countries: Ireland, Austria, UK, Poland and Greece

The expectation is that aside from conducting observations with teenagers, third parties will conduct interviews with the said select stakeholders from the same hubs where the observations will be conducted. The aim of the interviews will be to:

- x. Look at the success or challenges of the toolkits implementation and adoption.

- xi. Assess whether the toolkits are improving how girls view science and other STEM subjects.
- xii. Understand how toolkits are improving the way girls view science, assuming that this will be the case.
- xiii. Understand the level of usage by teenage girls and boys.
- xiv. Evaluate the activities of the hubs for instance by comparing and contrasting the activities of the hubs in order to assess whether certain activities draw more girls and encourage their engagement with STEM.
- xv. Gauge the usefulness of the activities that are implemented within the hubs.
- xvi. Gauge the usefulness of the guidelines that will be provided to toolkit users.
- xvii. Ascertain the interests of the teenagers by looking at how the activities affect the interests of the teenagers.
- xviii. Draw the interviewees into sharing what has been working well within their hubs and what could be improved and how.

6.2.2.2 Second Set of Interviews

The second set of interviews will be conducted by the evaluators. These interviews will be conducted with work package leader from MUST (Toolkit development) and a with work package leader from Experimentarium (Toolkit implementation). In addition, we will conduct interviews with the remaining 4 third parties from the following countries: Estonia, Serbia, Spain and Sweden.

The second set of interviews will be during the toolkit development and implementation phases. The aim of this second set of interviews will be to;

- vi) Understand the successes and challenges of module development.
- vii) Understand the successes and challenges of the module implementation.
- viii) Assess the chosen module activities in order to understand why they were chosen.
- ix) Consider whether there was any input from teenagers in the toolkit activities and whether such input was taken into consideration.
- x) Gauge how toolkit developers went about engaging the teenagers in toolkit activities.

The aims of both sets of interviews outlined in sections 6.2.2.1 and 6.2.2.2 should be seen as indicative interest areas from which the interview questions will be developed. Therefore, the specific question and observation guidelines will be sent to the respective third parties in month 18. This is because as indicated in section 6.1.1 the toolkit production would have started in month 15 and between this month and the expected start of the implementation in month 23, third parties would have had adequate time to sort out implementation logistics. As such, sending the specific questions at that time will mean that third parties will have enough time to prepare and subsequently implement the implementation process.

6.2.3 Translation and Transcription

Since the observations and interviews will be conducted in different countries, we understand that they may be conducted in the local language of respective countries. Therefore, the expectation is that the third parties will translate the observation and interview results to English so that the evaluators are able to analyse the results and give appropriate feedback and recommendations.

In addition, it is expected that all observations and interview data will be transcribed by third parties prior to sending the collected data to the evaluators. To standardise the process, we will send a template that will detail the format of how data should be sent back the evaluators in month 18.

6.2.4 Document Analysis specifically through use of Basecamp

We will use document analysis to gain an understanding of the multiple aspects of the Hypatia project such as background information and progress of the project through looking at some reports/deliverables that specifically relate to the development and implementation of the toolkits.

6.3 Data Analysis

Data analysis will be a mix of qualitative interpretive and quantitative analysis. Responses from questions will be subjected to a thematic data analysis where words and phrases with similar meaning will be grouped together into themes and presented narratively. As part of the analysis we will use Nvivo data analysis software to analyse qualitative data.

6.4 Summary of Evaluation Approaches

Approach	Who is doing it	Where	When
Observations	All 9 third parties	In their respective countries	Month 25 - 27
First set of interviews (5 third parties conduct interviews with stakeholders of their choice from the hubs that they will be observing)	Science Gallery	Ireland	Month 25 - 27
	NOESIS	Greece	Month 25 - 27
	ScienceCenter--- Netzwerk (SCN)	Austria	Month 25 - 27
	The UK Association for Science and Discovery Centres (ASDC)	UK	Month 25 - 27
	Experyment	Poland	Month 25 - 27
Second set of interviews (Evaluators conduct interviews with 2 partners and the remaining 4 third parties)	Museo Nazionale della Scienza e della Tecnologia (MUST)	Skype	Month 15
	Experimentarium	Skype	Month 25
	Science Centre AHHAA Foundation	Skype	Month 25

	LA CAIXA Foundation	Skype	Month 26
	Center for the Promotion of Science (CPS)	Skype	Month 26
	Teknikens Hus	Skype	Month 26

6.5 Ethical Consideration

Considering the nature of the project and what ethical implications could arise, we will expect all parties involved in the evaluation to ensure that the ethical approval form (Appendix 1) has been duly completed prior to any evaluation activities that are part of this evaluation strategy. It is inevitable that the ethical approval process will vary from country to country and may also depend on the institutional context of where the evaluation will be conducted, therefore all parties involved should follow the process that is applicable and feasible for them.

7 Activities Timeline

7.1 Interim Report in Month 18

- v) Document analysis
 - *Specifically focussing on aspects to do with hub creation, where the hubs will be, number of hubs.*
- vi) Interviews by evaluators with work package leader for MUST in **Month 15**.
 - *This will be during the piloting phase of the development task*
- vii) Send out the observation and interview protocol guidelines in **Month 18**
 - *This will give third parties enough time to sort out logistics e.g. who they will interview, where they will observe.*
- viii) Submission of Interim Report in **Month 18**
 - *The report will include;*
 - *Findings from document analysis and 1 partner interview*
 - *Observation and interview protocols with detailed guidelines and questions*

7.2 Final Report in Month 36

- v) Interviews by evaluators with work package leader for Experimentarium in **Month 25**
- vi) All 9 Third parties begin observations in **Month 25 to 27**
- vii) Third parties conduct Interviews (see selected 5 third parties in section 6.2.2.1) between **Month 25 to 27**
- viii) Evaluators conduct interviews with Third parties (see selected 4 in section 6.2.2.2) between **Month 25 -27**
- ix) Third parties send evaluators collected observation and interview data in **Month 29**

- x) Submission of Final Report in **Month 35**

8 Conclusion

This report has covered the evaluation strategy for the Hypatia project. It has described the objectives of the evaluation strategy which has touched two evaluation approaches. This has been in addition to evaluation methodology covering the evaluation design and evaluation guidelines. In addition to this, the report has given a detailed evaluation framework covering such elements expected impact, suggested approaches, evaluation criteria and indicators of success. Finally, the report has given a timeline for the next evaluation activities that are part of the evaluation strategy.

9 References

- [1] Campaign for Science and and Engineering (CaSE), "CaSE Diversity in STEM report," King College London, 2014.
- [2] European Commission, "Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation," Directorate -General for Research and Innovation, Brussels, 2012.
- [3] C. Leaper, T. Farkas, and C. S. Brown, "Adolescent Girls' Experiences and Gender-Related Beliefs in Relation to Their Motivation in Math/Science and English," *J. Youth Adolesc.*, vol. 41, no. 3, pp. 268–282, Mar. 2012.
- [4] OECD, *The OECD Innovation Strategy Getting a Head Start on Tomorrow: Getting a Head Start on Tomorrow*. OECD Publishing, 2010.
- [5] R. Wilson, "Wilson, R. 2009. The Demand for STEM Graduates: some benchmark projections," Council for Industry and Higher Education., London, 2009.
- [6] Research Councils UK, "Evaluation: Practical Guidelines - A guide for evaluating public engagement Activities." [Online]. Available: <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/publications/evaluationguide.pdf>. [Accessed: 17-Nov-2014].
- [7] G. Rowe and L. J. Frewer, "Public Participation Methods: A Framework for Evaluation," *Sci. Technol. Hum. Values*, vol. 25, no. 1, pp. 3–29, Jan. 2000.
- [8] C. Chess and B. B. Johnson, "Organizational Learning about Public Participation: 'Tiggers' and 'Eeyores,'" *Hum. Ecol. Rev.*, vol. 13, no. 2, p. 182, 2006.
- [9] D. L. Stufflebeam and A. J. Shinkfield, *Evaluation Theory, Models, and Applications*. San Francisco: Jossey-Bass, 2007.
- [10] J. Bennett, *Evaluation Methods in Research*. Bloomsbury Publishing, 2003.
- [11] X. Song, N. Letch, L. Xu, and K. Huang, "The utilisation of IS evaluation a HIS case study," in *Information Technology in Medicine and Education (ITME), 2012 International Symposium on*, 2012, vol. 2, pp. 960–964.
- [12] M. Scriven, *Evaluation thesaurus*. Sage, 1991.
- [13] S. Joss, "Evaluating consensus conferences: Necessity or luxury," in *Public participation in science: The role of consensus conferences in Europe*, 1995, pp. 89–108.

- [14] A. Tuominen, T. Järvi, K. Hyytinen, E. Mitsakis, M. E. Lopez-Lambas, L. L. Paix, J. van der Waard, A. Binsted, and A. Sitov, "Evaluating the achievements and impacts of EC framework programme transport projects," *Eur. Transp. Res. Rev.*, vol. 3, no. 2, pp. 59–74, Jul. 2011.
- [15] E. Arnold, "Understanding long-term impacts of R&D funding: The EU framework programme," *Res. Eval.*, vol. 21, no. 5, pp. 332–343, Dec. 2012.
- [16] Research Councils UK, "Evaluation: Practical Guidelines - A guide for evaluating public engagement activities," 2014.
- [17] G. Rowe and L. J. Frewer, "Evaluating public-participation exercises: a research agenda," *Sci. Technol. Hum. Values*, vol. 29, no. 4, pp. 512–556, 2004.
- [18] J. Abelson, P.-G. Forest, J. Eyles, P. Smith, E. Martin, and F.-P. Gauvin, "Deliberations about deliberative methods: issues in the design and evaluation of public participation processes," *Soc. Sci. Med.*, vol. 57, no. 2, pp. 239–251, 2003.
- [19] J. Abelson and F.-P. Gauvin, *Assessing the impacts of public participation: Concepts, evidence and policy implications*. Canadian Policy Research Networks Ottawa, 2006.
- [20] F. Merckx, I. van der Weijden, A.-M. Oostveen, P. van den Besselaar, and J. Spaapen, "Evaluation of Research in Context A Quick Scan of an Emerging Field," *Den Haag Rathenau InstituutERIC*, 2007.
- [21] P. Mickwitz, "A framework for evaluating environmental policy instruments context and key concepts," *Evaluation*, vol. 9, no. 4, pp. 415–436, 2003.
- [22] J. Fitzpatrick, J. R. Sanders, and B. R. Worthen, *Program Evaluation: Alternative Approaches and Practical Guidelines* ., 4th ed. San Francisco: Allyn and Bacon, 2010.

Annex 3: WP5 Interview Questions on implementation of the Toolkit

- i) Can you tell us a bit about your role as regards to implementation
- ii) What implementation activities have there been?
 - a. Is there any particular reason for these sort of implementation activities?
- iii) Can you tell us about how, when and where the toolkit activities have been used by each participating institution?
 - a. How did this go (iii), what were particular successes, what were particular challenges?
- iv) Was there any input from teenagers with regards to the toolkit activities?

- a. If there was, what sort of input did they give
- b. Was the teenagers input taken into consideration?
- v) How did the implementers go about engaging the teenagers in toolkit activities
 - a. If they have not, are there any plans to change this (viii)?
- vi) Implementation report (Expectations)
 - a. first report by the end of September 2017
 - b. second report by the end of January 2018
 - c. and third and final report by end June 2018
- vii) Has there been any impact so far?
 - a. What sort of impact do you expect?

Annex 4: Third party interviews - December, 2017

- i. How successful was the implementation of the toolkit activities?
- ii. What challenges did you encounter during the implementation of toolkit activities?
- iii. What was the level of third parties involvement in choosing the toolkit activities?
- iv. Was there any input from teenagers in the toolkit activities?
- v. How was such input taken into consideration?
- vi. Were you happy with the process of implementing the toolkit activities?

Annex 5: Observation and interview instructions



HYPATIA PROJECT OBSERVATION GUIDELINES

Contents

<u>Observation and interview instructions</u>	81
<u>1 Introduction</u>	85
<u>2 Objectives of the Module</u>	85
<u>3 Conducting Observations</u>	86
<u>3.1 Intended outcome of the module</u>	87
<u>3.2 Toolkit guidelines</u>	87
<u>3.3 Indicator of success</u>	87
<u>4 Conducting Interviews</u>	91
<u>4.1 Interview Questions</u>	91

This document is intended to be used by third party members of the project. The total number of third parties conducting the observations will be 9 (Ireland, Austria, UK, Poland, Greece, Estonia, Serbia, Spain and Sweden). Each third party will need to conduct at least one observation at one of the three settings that will implement the Hypatia modules. As the third parties will be implementing the modules in three contexts namely Museums, Schools and Industry, and in order to have a fair overview of how the modules implementation is taking place, each third party is expected to observe one context as indicated below in Table 1: ‘Allocation of observation contexts’ and complete the Table 2: ‘Module objectives and indicators of success’.

Third party’s country	Context
1. Ireland	Museum
2. Austria	School
3. UK	Industry
4. Poland	Museum
5. Greece	School
6. Estonia	Industry
7. Serbia	Museum
8. Spain	School
9. Sweden	Industry

Table 1: Allocation of observation contexts

In addition, each third party will need to choose someone who can conduct the observation to avoid bias. An additional set of interviews will be carried out by 5 third parties (Ireland, Austria, UK, Poland and Greece). This means that third parties from Estonia, Serbia, Spain and Sweden can ignore the interview part of the guidelines and only concentrate on answering the first part of the document related to the observation.

The interviews need to be carried out with one of the following stakeholders: head teachers, teachers, museum staff and industry representatives. The third parties can choose who they prefer to interview.

If you have queries or require further information regarding the observations and interviews, please contact the evaluators using the following emails: kutoma@dmu.ac.uk or tilimbe.jiya@dmu.ac.uk

Introduction

Evaluation involves learning new knowledge through gathering information, making credible conclusions or judgements that can be used in decision making and communicating the findings to an audience. It includes looking at the quality of the content, the delivery process and the impact of the activity on the participants. Evaluation provides an opportunity to determine whether the aim of the activity was achieved or not. For instance, considering the Hypatia project, the evaluation process will assess aspects related to the usability of the toolkit, toolkit user numbers, the effectiveness of the toolkit and the engagement with the toolkit activities, thus the need for the evaluation guidelines. The guidelines include several sections as covered below. The expectation is that these sections will be covered by the parties undertaking the evaluation. A description of what ought to be included is highlighted in the respective sections.

Objectives of the Module

Looking at the table below, please provide an indication of whether the objectives we have selected to concentrate on of this module have been met:

Third Parties	Context	Objective	Indicator(s) of success	Outcome (Y/N)
<ul style="list-style-type: none"> • Ireland • Poland • Serbia 	Museum	Understand what sex and gender stereotypes are and identify them	Evidence of an understanding of gender stereotypes e.g. Engineering jobs are mainly for boys	
<ul style="list-style-type: none"> • Austria • Greece • Spain 	School	Increase awareness about the negative impact they can have on their own representations of science, the world of science and technology, and their study/career choice	Evidence of reflection on more positive ways of representing STEM in Schools e.g. Gender neutral STEM materials such	

			language, role models in STEM	
<ul style="list-style-type: none"> • UK • Estonia • Sweden 	Industry	Learn about careers in STEM and develop an interest in them, regardless of their sex.	Evidence of how industry has/is helping with the promotion of STEM careers for both boys and girls e.g giving talks and presentations about STEM careers for both girls and boys	

Table 2: Module objectives and indicators of success

If the chosen answer is ‘Yes’, please give examples of how they have been met and if the answer is ‘No’ please give an insight into why they have not been met.

Conducting Observations

The collection of data based on observations is an example of qualitative evaluation tool which involves watching, recording and analysis of behaviour as it occurs in a ‘natural’ setting. Observation enables the evaluators to understand the participant (s) engagement with specific tasks and define the key issues that may be followed up in the interviews. As the development and implementation of the toolkit is central to this project, particularly in terms of engaging teenagers in STEM, it becomes imperative to observe how teenagers are engaging with the activities in addition to how institutions are supporting the toolkit implementation and subsequent adoption. As such, a further guideline would be the completion of the following sections specifically related to observed elements within the institution.

Intended outcome of the module

Describe in your own opinion, what is the intended outcome(s) of your assigned module in Table 1:

Toolkit guidelines

Please give an indication of the effectiveness (degree of success) of the toolkit guidelines provided as far as facilitation and gender inclusion is concerned.

- c. Were there any challenges in following the guidelines in order to implement or facilitate the modules? If they were, please highlight what these challenges were.

- d. In your own opinion and understanding of gender inclusiveness, were there any challenges in ensuring gender inclusiveness? If they were, please highlight what these challenges were

Indicator of success

Please use the following table to measure the success of the module and indicate the results in the outcome column:

	<p>balance or did the game activities reflect a gender balanced protagonists?</p> <p>Institutional level</p> <p>xxviii. To what extent has the aim or mission of your institution shaped the activity?</p> <p>xxix. Was the physical learning environment appropriate for the planned activities i.e. was it appealing to both girls and boys in terms of design, materials, and availability of role models?</p> <p>xxx. If yes, please give examples of how this was achieved?</p> <p>Societal/Cultural level</p> <p>xxxi. Did the activity cause participants to question gender norms that may be established at the societal/cultural level, i.e. that girls are naturally more nurturing or that boys are more logical</p>	<p>viii.</p> <p>ix. Yes No</p> <p>x.</p> <p>xi. Yes No</p> <p>Please elaborate on the given answer</p>
--	--	--

	<p>xxxii. Did the activity 'clash' with societal or cultural norms, and if so did this give rise to difficulties with respect to achieving the intended outcomes.</p>	<p>.....</p> <p>xii. Yes No</p> <p>Please elaborate on the given answer</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>
--	---	---

Conducting Interviews

Interviews are another type of evaluation tool which is used to understand and obtain useful information from the participant's experiences. To get an in-depth understanding of the implementation and adoption of the toolkit, interviewing relevant stakeholders in addition to observations becomes a necessity. The expectation is that aside from conducting observations with teenagers, the third parties conduct an interview with a stakeholder of their choice from the institution implementing the module such as head teacher, museum coordinator and industry representative. While all third parties are conducting the observations, only 5 of the third parties (Ireland, Austria, UK, Poland and Greece) conduct the interviews.

Interview Questions

- xxiii. How useful were the guidelines on facilitation and gender inclusiveness in ensuring that the modules were effectively implemented?

- xxiv. What risks/complications/challenges did you encounter while conducting these modules e.g. availability of resources, availability and understanding of the staff members, overall management of the activity?

- xxv. If there were any risks involved, how were they tackled? What measures were put in place to minimise those risks?

xxvi. Were there any challenges faced when these activities were conducted or during implementation or adoption?

xxvii. Describe the strategies which were used to implement the activities?

a. Did they work well?

i. Yes

ii. No

b. If there were any challenges, kindly highlight them while describing how they were overcome?

xxviii. What improvements can be made to the modules for future use?

xxix. Any other comments

