

MOI! Pilot report

28.4.2022

Peppi Kääpä
Finnish Heritage Agency



Co-funded by the
Creative Europe Programme
of the European Union



Table of contents

- Introduction..... 2
- Pilots 3
 - Pilots in practice 4
- Data gathering and analysis 5
- Results and feedback..... 6
 - Overall impressions 6
 - Discussion and evaluation processes 7
- Modules..... 9
- NEMO’s training on impact and self-evaluation 11
 - Feedback of the Impact modules’ introductory texts 12
- Improvement suggestions..... 13
 - Structure and content 13
 - Process, guidelines, and scale 14
- Glossary 15
- Framework output and tools..... 15
- Conclusions..... 16
- Appendix A: Feedback questionnaire for observers 17
- Appendix B: Feedback questionnaire for participants 18

Introduction

MOI! Museums of Impact is a European cooperation project co-funded by the Creative Europe Programme dedicated to developing a self-evaluation framework for European museums. The framework is meant to be used by museums as a tool to evaluate their practices and organization and finding their developmental areas, focusing on the impact on society. Using the tool will help museums take a transversal look at their activities, engage in internal discussions about the aims and goals of their work, share views and set development targets.

The project partnership includes development organisations (The Finnish Heritage Agency [FHA], Finnish Museums Association [FMA], BAM! Strategie Culturali [BAM], Hellenic Ministry of Culture and Sports, MUSIS Steirischer Museumsverband [MUSIS], The Museum Council of Iceland [MCI], Institut für Museumsforschung [IFM]), museums (Estonian National Museum [ENM], Museum of Cycladic Art [MCA],) and museum networks (NEMO Network of European Museum Organisations, The European Museum Academy [EMA]).

The partnership is supported by a network of five Associate Partners (Association of Independent Museums AIM UK, Danish Agency for Culture and Palaces DK, National Institute for Museums and Public Collections PL, Netherlands Museumregister Foundation NL, and ICOM Austria AT). Integrating stakeholder knowledge and views into the project has been achieved through open Stakeholder forums, webinars as well as through the piloting of the framework in six pilot museums from five different countries.

The piloting phase took place in six European museums from January to March 2022. The objective of the piloting was to open the evaluation framework draft for testing in real-life conditions. Piloting the framework took part with the support of selected partners present in each pilot event. Participating partners included FHA, BAM, EMA, ENM, MCA, FMA, MUSIS, NEMO and IFM. The museums included in the piloting were Vabamu Museum of Occupations and Freedom (Tallinn, EE), Georg Kolbe Museum and Kunsthaus Dahlem (Berlin, DE), Istituzione Bologna Musei (Bologna, IT), Museum am Dom (St. Pölten, AT) and Turku Museum Centre (Turku, FI). The pilot museums included were selected from partner countries, to ensure strong national support in this phase.

The museums were contacted and selected through the partners' networks. Their selection criteria included willingness to commit suitable resources for the pilot; willingness to perform a participatory evaluation, including staff; interest in developmental evaluation; and willingness to engage with the model beyond the project-life time. In addition, the selection was made in a way as to ensure sufficient variability in the types and sizes of museums selected, as well as geographical balance.

During the piloting phase the museums performed a self-evaluation of its operations according to the framework draft. As comprehensive as possible a composition of the museum staff went through the chosen modules, discussed the themes, and gave a numeral grade to each statement, in sets of facilitated discussion groups. The pilot evaluation was facilitated by individuals from the participating institutions and monitored by those partners present at the evaluation, to observe the use of the model and gather feedback from the users. The same experts were involved in multiple evaluations to guarantee continuity and coherence in the evaluations. Experiences, both from partner observers, and pilot museums, were recorded by observing the process as well as using a feedback form in all evaluations. Two to three days were reserved for the chosen framework module(s) with discussions and feedback.

The purpose of this report is to highlight the main results of the pilots by gathering the observations and feedback from all six pilot events. These results include observers' notes and feedback survey responses from individual participants. These experiences feed into to the further development of the model in the next phase.

Pilots

The chosen pilot museums represent a variety of museum organisations from small to large institutions and from public institutions to private foundations, from all museum sectors. The museums were selected through the MOI project partners' networks. The pilot museums were contacted by the project partners in their countries. With the museums contacted the project sought to have a comprehensive representation of museums of different types, sizes and structures.

Georg Kolbe Museum, Kunsthaus Dahlem and Museum am Dom represented smaller institutions. Georg Kolbe Museum promotes a lively dialogue that links historical issues with the present, with exhibitions and a varied art education program on classical modernism and contemporary art. Kunsthaus Dahlem is an exhibition venue for post-war German modernism (East and West), mainly concentrating on the presentation of sculpture with the exhibition venue. Both museums' directors had a strong interest and curiosity in participating in the project, as the idea of self-evaluation fit well with their idea of strategic development of their institutions. Museum am Dom is a museum of the diocese of St. Pölten with a small number of staff. Vabamu Museum of Occupations and Freedom is a museum with 20 employees, focusing on themes of freedom and civil rights in Estonia. Vabamu's recent change of director motivated the museum to participate in the pilot to have discussions about self-development in the organisation.

Istituzione Bologna Musei and Turku Museum Centre represented larger museum networks. Istituzione Bologna Musei is a publicly funded museum network with 12 museums in the municipality of Bologna. Istituzione Bologna Musei was a partner of the Stakeholder forum in February 2021 and expressed interest in the pilot immediately. Turku Museum Centre consists of 6 museums owned by the City of Turku, and it has previous piloting and evaluating experience with an existing evaluation model for Finnish museums. Turku Museum Centre expressed its interest as they are currently striving for innovation in different areas and saw that the pilot would be of benefit in these processes.

The pilot museums were invited to an online kick-off meeting on November 19th and December 10th, 2021, to introduce the MOI project, Developmental Evaluation and Self-evaluation as a tool, the MOI framework and to give instructions for carrying out the pilot evaluation. A meeting for observers followed on December 17th to give the partner observers more information about the practicalities of observing in the pilots.

Pilots in practice

Pilot	Museum	Modules	Participants in total	Facilitators	Observers in total
Tallinn 11.- 12.1.2022	Vabamu	<ul style="list-style-type: none"> • What we do – Impact goals and strategy • How our organisation functions – Communication, Service Development and Resources • Societal relevance • Relevant and reliable knowledge 	16	1	6
Berlin, 10.–11.2.	Georg Kolbe Museum	<ul style="list-style-type: none"> • What we do – Impact goals and strategy • How we embed the digital – Digital engagement • Societal relevance • Sustainable and resilient societies 	4	1	7
	Kunsthau Dahlem	<ul style="list-style-type: none"> • How we work – Organisational culture and competences • How our organisation functions – Communication, Service Development and Resources • Societal relevance • Relevant and reliable knowledge 	2	1	7
Bologna, 14.–15.2.	Istituzione Bologna Musei	<ul style="list-style-type: none"> • How we work – Organisational culture and competences • How we embed the digital – Digital engagement • Societal relevance • Sustainable and resilient societies • Communities and shared heritage • Relevant and reliable knowledge 	25	2	8
St. Pölten, 21.–22.2.	Museum am Dom	<ul style="list-style-type: none"> • What we do – Impact goals and strategy • How we work – Organisational culture and competences • How our organisation functions – Communication, Service Development and Resources • Societal relevance 	4	4 (1 per module)	4
Turku, 7.–9.3.	Turku Museum Centre	<ul style="list-style-type: none"> • How we work – Organisational culture and competences • Societal relevance • Relevant and reliable knowledge 	26	1	6

Figure 1. Overview of the pilot events and chosen modules.

The pilots began from Vabamu museum in Tallinn over two days on January 11th and 12th, 2022. The pilot host was Estonian National Museum. The museum piloted two modules per day. The number of participants on the first day was 6, on the second day 10. The observers represented ENM, FHA, MCA, and EMA. Most of the discussions were in Estonian, so most of the partner observers focused on observing the group dynamics because of the language barrier.

The second and third pilot were hosted by IFM took place in Berlin on February 10th to 11th, simultaneously in Georg Kolbe museum and Kunsthau Dahlem. Georg Kolbe museum piloted four

modules with 4 participants, and Kunsthaus Dahlem two modules with 2 participants. A German translation of the framework, which was made in collaboration between the German and Austrian partners, and printed card scales were used when needed in both pilots. The observers represented BAM, FHA, NEMO and IFM. Both pilot events had one facilitator with a double role facilitating and partly participating in the discussions, as the overall number of participants was very small.

The fourth pilot was hosted by BAM, and it took place on February 14th to 15th in Bologna at the MAMbo - Museum of Modern Art of Istituzione Bologna Musei. Six modules were piloted in three sessions: the first two modules piloted on the first day and the last four modules on the final day. The number of participants in the sessions were 7, 9 and 9. The discussions were partly in Italian and the groups used an Italian translation of the workbook and printed card scales. Two people facilitated the discussion process, while one concentrated on the discussion process and the other on taking notes. Partner observers represented BAM, FHA, MUSIS and SPK.

The fifth pilot was hosted by MUSIS and took place on February 21st and 22nd in Museum am Dom in St. Pölten. The museum piloted four modules in total. The number of participants in the group was 4, and each member of the group acted as a facilitator in turn. The group used paper prints of the German translations of the workbooks and one computer, and the discussions were in German except for the end discussion between the partners. The observers represented FHA, IFM and MUSIS.

The sixth pilot was hosted by FHA and took place in Turku Museum Centre in Turku on March 7th to 9th. The museum piloted three modules on separate days with different groups. The number of participants on the three days was 5, 12 and 9. The observers represented FHA, FMA and ENM. The discussions were in Finnish.

Observers made notes on a standardized feedback form during the evaluation process, observing the discussion processes and usability of the framework.

Data gathering and analysis

Notes were taken by observers from partner organizations during each pilot event using a standardized form. The questions focused on discussion processes, quality of understanding, challenges in the evaluation process, time management, specific questions, terms, or expressions that posed difficulties and other observations and improvement suggestions. Observers' reports were shared via the project's internal communication channels after each pilot event. The partners compiled a coherent report of all individual observer reports, which helped to outline the key observations in the pilots.

A Webropol survey was opened in the beginning of the pilot phase to gather experiences and feedback of the framework and piloting anonymously. The Webropol link was distributed to each individual participant after the pilot event. The questionnaire consists of 15 questions that focus on the overall pilot experience, benefits of self-evaluation and the framework, functionality of the evaluation statements, enabler modules, impact modules and general improvement suggestions.

The quantitative part of the material has been analysed using Webropol's own features. The qualitative part of both the Webropol responses and the observers' reports have been loosely analysed thematically, thus emerging themes and topics have been discovered and named from the material.

Results and feedback

Overall impressions

The framework was seen as a useful tool for evaluating the organisation's concrete actions and finding potential development targets. Majority of the feedback respondents felt that the framework initiated reflection about one's job and the organization and enabled transversal polyphonic discussions, which was regarded a positive experience. There was an experience that the framework challenges the museum to rethink its goals and practices, asks questions and supports the discussion by bringing new perspectives. The evaluation process led to important value discussions and focusing on the concrete things.

When asked about the functionality of the framework's contents as a whole, it was said that while it was clear and easy to use, some terms and expressions were too abstract, wide, or unclear, which led to multiple interpretations. Some respondents pointed out some overlapping in the modules and themes, and that some statements were very broad covering too many topics to give one grade to. In some cases, there were challenges with the translation of some terms. Some respondents stated that using the framework requires pre-planning, such as already initiated impact discussions in the museum to make the self-evaluation as beneficial as possible.

How well do you agree with the following statements?

5 = strongly agree, 4 = moderately agree, 3 = neither agree nor disagree, 2 = moderately disagree, 1 = strongly disagree

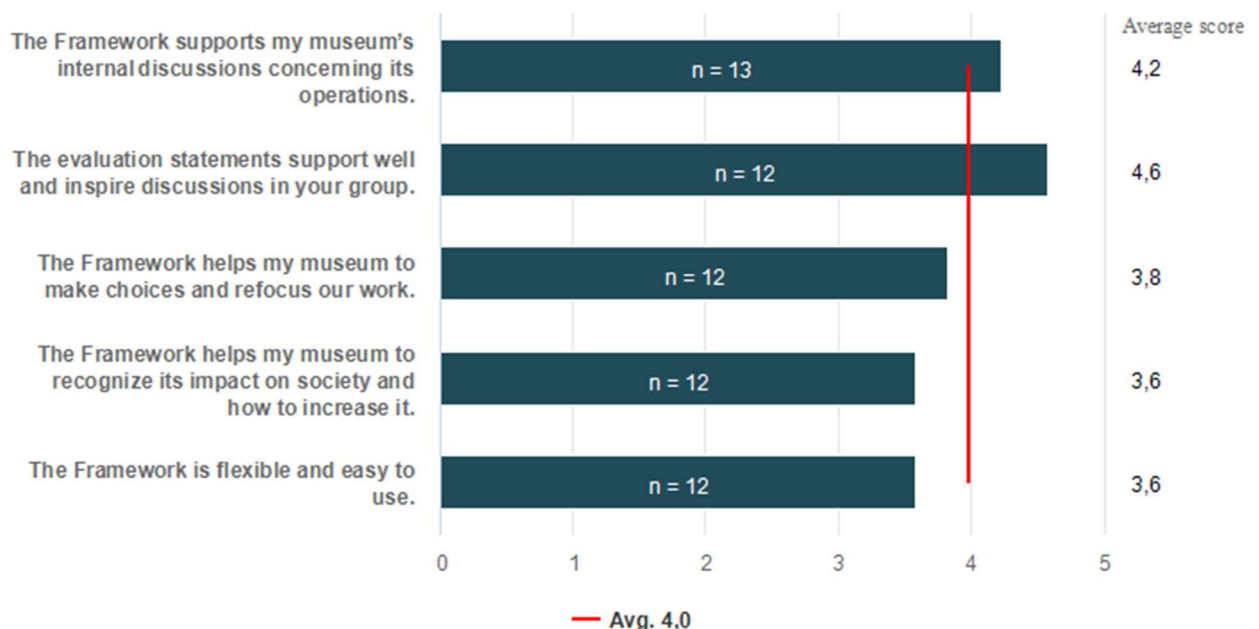


Figure 2. Participants' distribution of opinions on the framework. (Webropol)

Discussion and evaluation processes

In most pilots the atmosphere was good, and the participants were able to discuss freely and openly. It seemed that natural leaders and sometimes the directors voiced their views more, and quieter participants participated more actively when voting on the grade or if the topics concerned specifically their field of expertise. In some pilots, staff hierarchies set some boundaries for open transversal discussion. Anonymous participation was suggested with some sensitive-regarded topics, for example.

Common understanding of the terms and topics was often created in the discussions, which enhanced a dialogue. Sometimes the discussion turned into evaluating success rather than finding the potential developmental areas, regarding modules with the numeral scale in particular. In some of these cases, the facilitator led the discussion in the needed direction. The Enabler module ***What we do – Impact goals and strategy*** encouraged more open and deeper discussions as it posed open questions or statements without a numeral scale, whereas in general the scale seemed to somewhat limit and not always encourage a deeper discussion.

The printed versions of the workbook were mostly used in the pilots. Three pilot museums had printed physical cards showcasing the scale value with definitions, and these were occasionally used in addition to the workbook. In most cases the workbook or the evaluation scale was projected on a screen for support. The visibility of the workbook to all participants seemed to result in more engagement in the discussion. The participants didn't use the printed workbook very actively individually, and a few participants thought that it was useful that each have a copy. The scale cards were occasionally seen as useful, because they seemed to soften possible hierarchies or imbalances in the discussion and there were more contributions from previously quieter participants. In small groups the cards were mainly used as a tool to break up the discussion when it was stuck, or no one wanted to start the conversation.

The first pilot set the precedent for the time use with the framework. On average, a little over one to two hours was spent going through each module regardless of the group's size. It took roughly 10 minutes to discuss one evaluation statement, not longer than 20 minutes. Two modules per day seemed to be the total maximum the participants were able to engage in relatively actively. A two-hour slot for one module per day in the morning seemed ideal in efficiency and discussion dynamics, because the impression from full-day pilots was that the event causes participants to get fatigued and lose motivation.

The decisions on the grades were made jointly and the scale was mostly used as it is intended. According to feedback, the grade definitions were well formulated and structured, but sometimes they didn't match with the evaluation statements and the grades were rather used as rating success. N/A option was sometimes used if it was the consensus that the group was not able to take position on a statement. There was a lot of discussion interaction, and the final decision on the grade value came either after proposing ideas or the value was first proposed by each participant after which they made short statements on their decision. In some groups, especially in the case of managerial staff, coming to a unanimous decision seemed effortless and smooth. In the case of differing opinions within the group, the final grade seemed to be compromised even though mean values were suggested to be avoided. As the definitions of the scale values were quite specific and not the standard 1 to 5, the groups returned to recall the definitions from time to time, which suggests that the definitions were difficult to remember.

The presence of the facilitator was different in every group. In some pilots, the facilitator's presence was not as apparent, whereas in some pilot groups the facilitator took a more active role in leading the discussions. In some cases, it was noted that facilitators were not able to exercise their potential to steer the discussion, that is to intervene when the discussion was dominated by leading participants or when the perspective narrowed, partly because discussions can be influenced by the familiar hierarchy of the team. One facilitator seemed to be enough for pilot groups of all sizes. However, some observers suggested that the process could benefit from two facilitators, where one is concentrating on the discussion and other on taking notes, as it was sometimes challenging for one person to manage both. Two facilitators with different roles were used in one of the larger pilots.

Some observers pointed out that the numeral scale and presumptions in the statements/formulations of the statements can give the framework a normative character, which might result in participants trying to over-evaluate or answer the "right thing" instead of finding the potential development areas. In some cases, the scale was seen as encouraging participants to tick a box and go through the statements at the expense of having a deeper discussion. Moreover, sometimes the discussions seemed to focus more on deciding on the grade rather than the actual topics. On the other hand, the numeral scale was seen as a benefit for the process as a concrete step to end a certain topic and move on, as otherwise discussions can start to meander and lead nowhere.

The number of participants in the pilot events varied from 2 to 12 persons. Two participants and one facilitator marked the minimum for the use of the framework, whereas a group of 10-15 participants should be the maximum group size. A group of over 10 participants can be challenging to facilitate and to get a balanced discussion interaction.

The intermediate stages of the discussion were usually recorded in the open text fields after each theme, at least by the facilitator. Sometimes the notes were made together projecting the workbook on a screen for the whole group to see, which seemed to increase participation. A development plan was either made at the end of each module or after the whole evaluation process or planned to be made after the pilot event. The outcome of the discussions and next steps were unclear for some pilot museums, and more instructions seemed to be expected. In some pilots the development plan was completely out of attention, which can be due to time management, not engaging enough with the instructions or because the development plan is a separate document at the end of all modules (not included at the end of each module and can be left with less attention).

Modules

Module	Times piloted	Pilot museum
Societal relevance	6	Vabamu Museum of Occupations and Freedom, Kunsthaus Dahlem, Georg Kolbe Museum, Museum am Dom, Bologna Musei, Turku Museum Centre
Relevant and reliable knowledge	4	Vabamu, Kunsthaus Dahlem, Bologna Musei, Turku Museum Centre
What we do – Impact goal and strategy	3	Vabamu, Georg Kolbe, Museum am Dom
How we work – Organisational culture and competences	3	Museum am Dom, Bologna Musei, Turku Museum Centre
How our organisation functions – Communication, Service Development and Resources	2	Vabamu, Museum am Dom
How we embed the digital – Digital Engagement	2	Georg Kolbe, Bologna Musei
Sustainable and resilient societies	2	Georg Kolbe, Bologna Musei
Communities and shared heritage	1	Bologna Musei

Enabler module

Impact module

Figure 3. Modules in order from most often piloted to least piloted.

The modules were chosen by the pilot museums according to their priorities and interests. Enabler modules were not obligatory, but at least one relevant Impact module was suggested for piloting. Enabler modules were piloted in 5 out of 6 museums, and at least one impact module was piloted in each museum. ***What we do - Impact goals and strategy*** and ***How we work - Organisational culture and competences*** were the most popular enabler modules, as both were evaluated in half of the pilot events. ***Societal relevance*** and ***Relevant and reliable knowledge*** were the most popular impact modules: Societal relevance was evaluated in all pilot events and Relevant and reliable knowledge in 4 out of 6.

In general, the framework was regarded as a functional tool that enables discussion and creates a frame for it as such discussions could take more time resources without a specific tool. In some pilots, the experience was that the framework sparked discussions about new and previously untreated topics and perspectives that were found to be interesting and inspiring. However, there was parallel feedback on similar challenges repeating in the framework in all pilots. General challenges with both Enabler and Impact modules were related to definitions, formulations and translations of certain terminology and ambiguous expressions, which left room for multiple possible interpretations. Some overlapping in the modules and themes was mentioned, and that some

statements were very broad covering too many topics to give one grade to. In addition, repeating challenges had to do with the perspective the statements were meant to be approached from – the individual employee, organisation or audience perspective, and the situation the statements are evaluating – the current performance level or the present discussions and ideas within the organisation.

The impact aspect seemed to be forgotten in some of the pilot discussions especially regarding the individual statements. The term “enabler” caused some confusion in few pilot events. In some pilots it was pointed out that it could be interesting to have access to other museums’ evaluation results and compare situations, which could help the institution in mapping out their position and relevance in society in comparison with others. The following section discusses in more detail the feedback that arose within the different modules.

Enabler module ***What we do – Impact goals and strategy*** was piloted in three museums and regarded as the most interesting and fruitful module in terms of the discussion quality. Observers pointed out that the absence of the numeral scale and including only open-ended questions resulted in freer and deeper discussions of the theme. It was stated that this module fits well as the starting module in the evaluation process. However, in some pilots it seemed to come as a surprise that the module presupposes that the institution has set some impact goals in advance, which posed some challenges for the evaluation process.

How we work – Organisational culture and competences module was piloted in three museums, and it was the most manageable module in the case of a larger museum network and its pilot group due to the cross-sectoral nature of the themes, which resulted in a vivid discussion. In another case example, however, it was felt that some statements, regarding the theme of recruiting for example, were irrelevant or difficult to answer from the everyday staff’s point of view. This supports the idea that the evaluation process should include staff members from all possible levels, as some questions can be more specifically targeted towards the management, for example.

How our organisation functions - communication, service development and resources module raised some questions, in particular in the *Customers and audiences* section, on statements that cover many different topics and include many words with slight meaning differences. The functionality of these statements was questioned because the case can be that the museum strives on some topic and not on the other(s) that are under evaluation in the same statement.

How we embed the digital – Digital engagement module was piloted in two different sized museums. The term digitality was understood almost completely as digital communications in one case, and thus going through this topic of discussion was mainly on the communications team that was present in the evaluation. In another case, digital engagement as a term posed some difficulties with understanding, and the statements posed some difficulties with the modern and business-oriented terminology. It was suggested that the digitality themes be woven in the whole framework instead of packaging it as one module – this could also help with understanding the broad term digitality/digital engagement and that it concerns many areas in the museums’ operations.

Impact module ***Societal relevance*** posed some difficulties from the perspective of strategic planning and impact thinking inside the museum in addition to challenges with terminology and perspective. It can be difficult to evaluate the topics if the organisation has not thought about its relevance in society or had discussions and plans before the evaluation. In some pilots the module was felt to be too long with many overlapping, and it was felt that some statements were too complex. It was apparent in some pilots that staff from conservation/collections units didn’t see this module very relevant for their point of view, whereas other units’ the managerial staff seemed to engage with the

topics more smoothly. In the case of a larger museum network pilot, this module was considered too specific to discuss at a network/organisation level as opposed to an individual museum level.

Sustainable and resilient societies was piloted in two museums, and it posed some difficulties in both cases due to its wide interpretations and sustainability being unfamiliar for their operations. There were challenges in understanding what sustainability encompasses because it is a very broad term. For example, in the case of an art museum the subject was felt to be unfamiliar and somewhat irrelevant. It was stated that some of the statements should be reformulated as not all European countries/institutions are on the same page in practice on environmental issues and sustainability. In the context of this module, it can be felt by the evaluating institution that it is behind or going a different direction if they are not familiar with the theme and provided with additional information or know-how about it. Specifically, the *Sustainable Development Goals* were hoped to be defined in the framework, because they are explicitly addressed in the statements.

Communities and shared heritage module was piloted in one museum and its main challenges were related to unclear terminology, again, particularly regarding the central term *community*. The unclarity of the term resulted in the module to be difficult to evaluate engagingly. **Relevant and reliable knowledge** module was piloted in four museums and the main challenges had also to do with key terminology, such as *research* and *content*. Some museums rely on academic research more than others, so the research question should be clarified, because it can be interpreted as non-academic research done for/in exhibitions, for example. In the context of the decolonisation theme, a question raised whether the statement could include/additional statement be added addressing provenance research questions, as decolonisation is not a relevant notion in all museums.

NEMO's training on impact and self-evaluation

NEMO's training "Enhancing your Museum's Impact: Self-evaluation & development" on March 17th and 23rd, 2022 was a collaborative effort between the MOI project and partner NEMO. 20 museum professionals from Europe participated in the training. The training included an introduction to impact thinking and the MOI project, and discussion on the framework's topics and sharing experiences on the participants' museum work. The MOI project also gathered feedback of the introductory texts of the four Impact modules of the framework draft, further development in mind.

The first two modules, **Societal relevance** and **Communities and shared heritage**, were discussed in the first webinar on March 17th, and **Sustainable and resilient societies** and **Relevant and reliable knowledge** were covered in the final webinar on March 23rd. The modules' introductory texts were sent to the participants as pre-reading material. Based on the experience of the first webinar, the whole modules with evaluation statements were sent to the participants before the final webinar to get the big picture of the framework.

The discussion sessions in NEMO Online training webinars were recorded by one recorder or facilitator. The following questions were discussed: 1. *Do you agree with the contents of the introduction?* 2. *Is there something to add to it?* 3. *Are there any changes to be made?* The main discussion points and development ideas regarding the introductory texts or the theme in general were recorded.

Feedback of the Impact modules' introductory texts

Societal relevance module's introductory text got feedback on its orientation/perspective to societal relevance and the museum's internal activities striving for societal relevance. It was suggested that the introductory text's perspective should be more future-oriented as society is constantly evolving and changing: Are we still relevant in the future? Society is changing quickly – do museums change the same way? It was also suggested that it be thought about whose voice can be heard in the text: voice of employees, directors, museum boards? It was suggested to map out in the introduction who can shape societal relevance activities in museums; it should be emphasised that all departments, including collections, are in key roles. The questions raised were: Who can do it - is it enough when the director is doing it? With whom should the museum contribute to do it?

Regarding the module **Communities and shared heritage**, the assumption that all museums have impact was questioned. Some terms caused confusion, such as *actors* and *communities*. It was said that it should be addressed that it is very museum-specific how museums can define and work with its communities. However, the term communities should be defined clearly.

Regarding the module **Sustainable and resilient societies**, the museum's mediator role in preserving cultural and natural environments through giving visitors tools and ideas was seen an important aspect in the introduction, but it was questioned whether it is visible in the statements. Sustainability as a concept was discussed and it was regarded as too ambiguous. The environmental aspect of sustainability was largely emphasized in the module, and the social aspect could be highlighted more. The sustainability aspect was felt to be lacking under the *Collections and content* theme in comparison with other themes. It was suggested that a section be added to the module addressing if there is a person in charge of or willing to advocate for sustainability or environmental issues in the museum, who could act as a facilitator or advocate for the theme.

The sentence "Museums need to recognize their responsibilities and strengths as producers, custodians and intermediaries of information and content" was agreed on and regarded as important in the introductory text to the **Relevant and reliable knowledge** module. The policies of open data sharing were discussed and including GDPR and intellectual rights topics into the module was suggested. In addition, interdisciplinarity was suggested to be included in the statements, because it is a common theme in today's museums. People's different ideas of what a museum is, was discussed and whether it should be addressed in the statements. For example, it could be asked how much the museum is working on to explain what the museum is to overcome the symbolic distance between the museum and citizens. Questions raised were: How does history affect the collections and information we share? Should the burden of the museums' history be included in the module?

In the wrap-up session the participants shared their overall views on the framework and the information they got during the event. The feedback was mainly positive. The framework was seen as a versatile and usable tool that enables self-reflection and discussion within the museum to define impact goals and to strive towards development. The structure and modularity of the model were seen as allowing the museum to choose the sections that are relevant to them and to use the framework in conjunction with other projects and existing evaluation tools. The evaluation scale and grade definitions were regarded as well-formulated and concrete.

Improvement suggestions

Structure and content

The framework's wider structure and content didn't pose major challenges. The content seems to be relevant for museums, and no important themes or topics are missing. There seemed to be some overlapping between modules/themes/statements, and some statements were felt to be too broad covering too many topics. Some statements are evaluating several and differently performing operations in the museum, which may not be eligible for giving a common grade. The content of some individual statements should be updated, and, for example, too long statements could be shortened or split into separate statements. This would also shorten the framework and make the process less time-consuming and tiring for participants. The impact aspect seemed to be forgotten in some of the pilot discussions especially regarding the individual statements, but this can be due to not engaging with the instructions enough.

As the development plan seemed to be out of attention in a few pilots, it could be beneficial to rethink the placement/process leading up to it in the framework. The development plan should be emphasised as an important concrete step after the evaluation. In one case the discussion quality increased when encouraged to conclude the main points after two of the modules. Concluding questions are currently included only at the end of the first Enabler module ***What we do - Strategy and impact goals*** and in the development plan, but it should be considered to add few concluding questions after each module in addition to the discussion notes boxes. One pilot museum suggested that the scale and 3 open questions after each module would be a good combination. This could both increase the discussion quality in between the modules and make the development plan phase easier. In addition, there were some concrete improvement suggestions to the content of the development plan: it could include a "one year after" and a "funding period" section.

The approach in the framework was regarded by some as putting too much emphasis on improvement. If the framework is regarded as a normative tool, it can result in overestimation and trying to answer the "right thing" or putting institutions on the defensive. In some cases, it was suggested to emphasize the evaluation as a learning process and that the language could somewhat be adjusted to focus on the learning opportunities. However, it must be highlighted that all impact areas and themes are not relevant for all museums and that the evaluating museum decides on what areas are important in their evaluation. It should be clearly stated that the evaluation is not a test, but a tool to help museums see their current performance level to map out the relevant development areas.

In one pilot digitality themes were suggested to be woven in the whole framework instead of packaging it as one module. This could also help with understanding the broad term digitality/digital engagement and that it concerns many areas in museums' operations, not only communications, for example. It was suggested that modules covering topics such as digitisation could require different levels or stages for different museums, for example "beginner" to "advanced".

Process, guidelines, and scale

It is necessary to include some instructions on setting impact goals and how the museum can start the discussions on the subject before the evaluation, because some of the modules assume that the impact goals have already been set. Also, it would benefit the museum if the evaluation process started by defining important concepts. It was suggested that the museum would review the workbook, decide on the context and aims of the evaluation, and select the relevant topics or statements for the museum before the evaluation. In addition, interpretations on whether the statements evaluate internal or external activities could be done beforehand. This phase would save time and keep the group more focused. Also, these issues could be addressed in the guidelines or clarified with the glossary, because terminology is central for understanding the perspective.

According to the feedback, kind of “ambassador” organisation’s contact information and resource links could be included in the framework’s guidelines to provide the user with more information. It was suggested that the institution get a reminder via e-mail, for example, to do the evaluation again after 1-2 years to remind that it is an ongoing process. It could be stated in the instructions and that the evaluating institution can address this topic in their own development plan. According to the pilot results, this type of evaluation process could take place every 6 months to 3 years.

It could be helpful to encourage facilitators to be more active and offer a toolkit of moderating techniques to ensure a balanced and constructive discussion. It was noted in some pilots that facilitators did or could not use their full potential in steering the discussions. To keep the conversations manageable, the groups could be divided into smaller groups to discuss different modules with as much departmental diversity as possible, which would require participation of more than one facilitator. In addition, tips could be given on how to handle sensitive content and how to carry out anonymous participation in the discussion. During the evaluation process the discussion notes and remarks have more importance than the given grades, so perhaps the process could benefit of one person who uses the workbook; is taking notes, writing concluding remarks at the end of each module, and finally filling the development plan.

Enabling anonymous participation was hoped for the future of the framework. Even though in most cases the topics weren’t considered sensitive, the issue could be addressed in the framework’s guidelines. It was suggested that in such a case, the group would decide the anonymous topics or statements beforehand, and the topics would be addressed anonymously in written form. Sensitive topics could be suggested with a secret ballot, or the whole team could review the statements before the evaluation session and privately submit the statements they would prefer to answer anonymously. It was also hoped to be clarified whether managerial staff should participate in the whole evaluation process or only in selected sections.

It could be beneficial to emphasize the idea and methodology of self-evaluation and impact in the guidelines and introductions of using the framework. It would be important to highlight the impact thinking both in the guidelines and individual statements, as it was stated that if the statements don’t remind that impact is the goal, then participants can get lost in the descriptions of the activities in the statements. The numeral scale seemed to lead to an evaluative and normative tone in the discussions and finding the areas for development seemed to stay in the background. This suggests that the development, processes, goals, and strategy should be highlighted more as opposed to the grading system, because in the end the evaluation is not so much about the ratings but about the thematic areas for development.

It was evident that the evaluation process and discussions should start on the topics with open questions, which is why the first Enabler module could be recommended as the starting point. It was proposed that besides numbers, colours could be used as a grading or marking method. The numeral scale, especially with the voting cards, seemed to limit the discussion at a “marking exercise”, not encouraging a proper discussion. However, the scale should be included because the current grading was seen as a beneficial step for the discussion process in general, to move smoothly from one topic to another.

Glossary

The main challenges in the evaluation process were related to unclear terminology and wordings, and complex expressions and formulations of the statements. Difficult terms in the pilots were mostly broad key terms, such as *community*, *sustainability* and *digital engagement*, or more vague terms such as *operating environment*, *actions*, *services*, *content*, *research*, *resources*, and *knowledge platforms*. A detailed list of the challenging terminology has been compiled and will be used in the further development of the framework.

A glossary is useful for creating a frame for mutual understanding when the terminology is defined and clarified. Concrete examples can suggest what more vague or broad definitions can refer to in different contexts. The key terms for the use of the model should be defined so that the museum staff can start the evaluation process on common terms. However, undefined terminology and being able to discuss different interpretations led to more interactive and fruitful discussions, so not all terms should be strictly defined as this reduces the discussion quality.

Framework output and tools

The printable workbook with the note boxes was regarded a good and functional format. The workbook functioned especially well when it was projected on a shared screen and making notes jointly lead by the facilitator. It was noted in the pilots that participants rarely used or engaged with the workbook very actively individually. Few felt that the workbook was useful for each individually. The scale cards that were used unprompted by some pilot museums, or a similar card format, could be more engaging. It was suggested that different formats could be combined, and perhaps the workbook and cards could alternate more throughout the modules/process. The scale cards can encourage a more balanced discussion, and they could work especially in larger pilots.

A common observation was that participants can lose motivation towards the end especially if two modules are evaluated successively. It was suggested that the evaluation process could include more engaging ways of participating to keep the attention high throughout the process. Small exercises or mini games were suggested as kind of ice breakers. There was a suggestion that possible additional tools could be provided in a kind of “facilitator’s toolbox” together with other materials and instructions for the facilitator. To guarantee good instructions on self-evaluation and the MOI! framework without a personal briefing by the project, it was suggested that some information could be made available in short instruction videos. Regarding design, it was suggested that each module could have one more round of design with colours, or that the cards with statements and scale would be designed with different colours for different modules and emphasizing the impact context.

Conclusions

The piloting phase put the framework draft to test in real-life conditions. It provided the project with useful knowledge about how the tool functions in different museums and organisations, environments, and groups, and how the framework and the evaluation process were experienced by the users. The framework was piloted in six European museums from small to large and public to private institutions and foundations in Estonia, Germany, Italy, Austria, and Finland. The pilot group sizes varied from 2 to 12 participants with at least one facilitator facilitating each group, and 4 to 8 partner observers observing the event.

According to the pilots, a two-hour slot for one module per day in the morning was ideal in efficiency and good discussion dynamics. One facilitator was sufficient for groups of all sizes, but the evaluation process could benefit from at least two facilitators with different roles. Most pilot groups used the printed workbooks, and three pilot museums used in addition printed physical cards showcasing the scale value with definitions. The numeral scale was used as it is intended and the decisions made jointly. The interpretations of some terminology and statements were created in the discussions. Even though the scale was sometimes approached as a normative tool to rate success rather than find the areas for development, it was seen as a benefit for the process to mark the discussion on a certain topic and move on. A development plan was either made after each module, the whole evaluation process or planned to be made after the pilot event.

All modules were piloted at least once; ***Societal relevance*** and ***Relevant and reliable knowledge*** being the most piloted modules, whereas ***Communities and shared heritage*** was piloted only once. All Enabler modules were piloted at least twice, and the first Enabler module ***What we do – Impact goals and strategy*** was regarded as the most functional in the beginning of the process in terms of good discussion quality and interaction.

In general, the framework was clearly structured and easy to use. It initiated reflection and brought new perspectives about one's job and the organization, enabled transversal discussions beyond unit borders, and supported important value discussions. It was seen as a useful tool for evaluating the organisation's concrete actions, rethink its goals and practices and finding potential development targets. The main challenges with the framework and evaluation process were related to unclear instructions on what the museum needs to do and know before the evaluation, unclear terminology, and some overlapping between the modules, themes, and statements.

The improvement suggestions covered the whole process and the structure, content, guidelines, scale, glossary and output and tools of the framework. There were no major challenges with the wider structure and content, but rather small adjustments here and there, better guidelines and rethinking the final format and tools of the framework were needed. The next step is to finish the fine-tuning and create the final output and tools of the framework.

Appendix A: Feedback questionnaire for observers

Feedback Questionnaire for Observers

Pilot Museum:

Partner:

Facilitator:

Observer:

Number of Participants:

- 1. *What was your overall impressions of this Pilot and how the participants interacted with the Framework?***
- 2. *How did the discussion processes work in general in this Pilot?***
 - *How was the quality of discussion interaction?*
 - *How was the quality of understanding?*
 - *How was the quality of decision making (with regards to scale of the framework)*
 - *How was time management?*
- 3. *Did the pilot museum have enough adequate information and instructions in beforehand of the process? If not, what was missing?***
- 4. *What challenges were there in this Pilot evaluation process?***
 - *Sufficient number of facilitators?*
 - *Time slots? Pauses? Schedule?*
 - *Premises?*
 - *Understanding of the questions and questionnaire?*
 - *Group and discussion dynamics?*
- 5. *What challenges did the facilitator face?***
- 6. *How long did it take to discuss one question/one module? (estimation)?***
- 7. *Which questions did pose very specific difficulties and why would you think this is so? Are there terms, expressions or formulations being discussed or even apparently misunderstood?***
- 8. *Was there a sufficient representation of all staff categories and departments/units present in the evaluation (if this can be estimated)?***
- 9. *Do you recognise any impact from museum hierarchies impacting the result of the evaluation? Or other biases?***
- 10. *Other observations and improvement suggestions for the self-evaluation tool and/or its guidance material?***

Appendix B: Feedback questionnaire for participants

Feedback Questionnaire for Pilot Museums

Dear Participant of MOI Framework Pilot, your feedback is very important for the further development of the model. Thank you in advance for your help in improving European museum self-evaluation framework.

Guide to answering: note that you only need to answer those questions that are relevant for you, e.g. concerning the modules you used in your own evaluation. The questionnaire is meant to be answered individually and privately by all those who participated in the self-evaluation. No names will be recorded, and all answers are handled as confidential.

We suggest to get acquainted with the feedback questions before starting the self-evaluation in your museum. You are asked to complete the questionnaire within a week from the end of the evaluation.

Name of the museum:

Module(s) evaluated:

- Impact goals and strategy
- Organisational Culture and Competences
- Communication, Service Development and Resources
- Digital engagement

- Societal Relevance
- Communities and Shared Heritage
- Relevant and Reliable Knowledge
- Sustainable and Resilient Societies

Open question: Please explain why these modules have been chosen, and if some modules are not used, specify why (lack of time, lack of relevance, other reasons)

Staff categories present in the evaluation:

- Management and leadership
- Museum services (e.g. exhibitions, collection work, education, conservation)
- Other services (e.g. customer services, accounting and HR, maintenance, communications)

1. What kind of overall impressions are you left with after using this Framework?

2. How did the Framework's contents (modules, discussions, questions) work together as an entity?

3. How did you feel about working on the Enabler modules:

Open question: Were the questions comprehensible? Was this topic easy to address and to discuss? Were the evaluation statements relevant overall and as individual statements considering the topic?

Impact goals and strategy:

Organisational Culture and Competences:

Communication, Service Development and Resources:

Digital engagement:

4. Was something missing (a theme or a question) from the Enabler modules in your opinion?

5. How did you feel about working on the Impact modules:

Open question: Were the questions comprehensible? Was this topic easy to address and to discuss? Were the evaluation statements relevant overall and as individual statements considering the topic?

Societal Relevance:

Communities and Shared Heritage:

Relevant and Reliable Knowledge:

Sustainable and Resilient Societies:

6. Was something missing (a theme or a question) from the Impact modules in your opinion?

7. How well did the Framework's evaluation statements work in your opinion? Were they clear and topical? Were there any difficulties with the wording? Were they relevant to impact?

8. How well do you agree with the following statements?

A The Framework supports my museum's internal discussions concerning its operations.

strongly agree, moderately agree, neither agree nor disagree, moderately disagree, strongly disagree

B The evaluation statements support well and inspire discussions in your group

strongly agree, moderately agree, neither agree nor disagree, moderately disagree, strongly disagree

C The Framework helps my museum to make choices and refocus our work.

strongly agree, moderately agree, neither agree nor disagree, moderately disagree, strongly disagree

D The Framework helps my museum to recognize its impact on society and how to increase it.

strongly agree, moderately agree, neither agree nor disagree, moderately disagree, strongly disagree

E The Framework is flexible and easy to use.

strongly agree, moderately agree, neither agree nor disagree, moderately disagree, strongly disagree

9. Is there anything you would add to or change in the Framework to make it more relevant to the needs of your organization?

10. Do you think your museum can benefit from using such kind of framework regularly? What kind of benefits do you see in using this kind of evaluation?

11. Other observations or suggestions to improve the Framework's structure, content, and/or process?