Enrichment policy: stakeholder survey results

Revision	*Version 1.0*
Last updated	*28/06/2023*
Author(s)	*Maša Škrinjar, Hugo Manguinhas, Ariadna Matas, Europeana Foundation*
Dissemination Level	*Public*

From 8 to 27 March 2023, we invited our audiences to answer the enrichment survey that helped the Europeana Foundation collect information about opinions, views and practices from various stakeholders in the area of enriching digital cultural heritage.

The answers to the survey will support the definition of a policy for enrichments that should set a vision and bring consistency across enrichment efforts in the common European data space for cultural heritage.

This document provides an overview of the questions and answers to the survey. We received a total number of complete 166 responses, 32 of which were incomplete and not taken into consideration.

Table of content

Profiling questions

Q1 What is your name? You can answer the survey anonymously if you prefer not to share your identity, but transparency in who is answering the survey will help us better understand the stakes you represent

Answered: 110. Open text question. Answers to this question are not provided in this document.

Q2 If you fill in this survey on behalf of your organisation, please provide the name. If you answer on behalf of your organisation, please be mindful of answering in your professional capacity and not expressing your personal opinion. You are welcome to fill in the survey again in your personal capacity

Answered: 64. Open text question. Answers to this question are not provided in this document.

Q3 Which of the following roles apply to you and your work? Select all that apply (multiple answers are possible)

Answered: 134		
Choices (multiple possible)	Response count	Response percent
Developer	24	17.91%
Educator (primary or secondary education)	16	11.94%
Educator (higher or further education)	21	15.67%
Member of an aggregation service	29	21.64%
Student or in higher education	6	4.48%
Researcher (including PhD study)	43	32.08%
Policymaker	6	4.48%
Culture enthusiast	28	20.89%
Heritage professional or other	79	58.96%
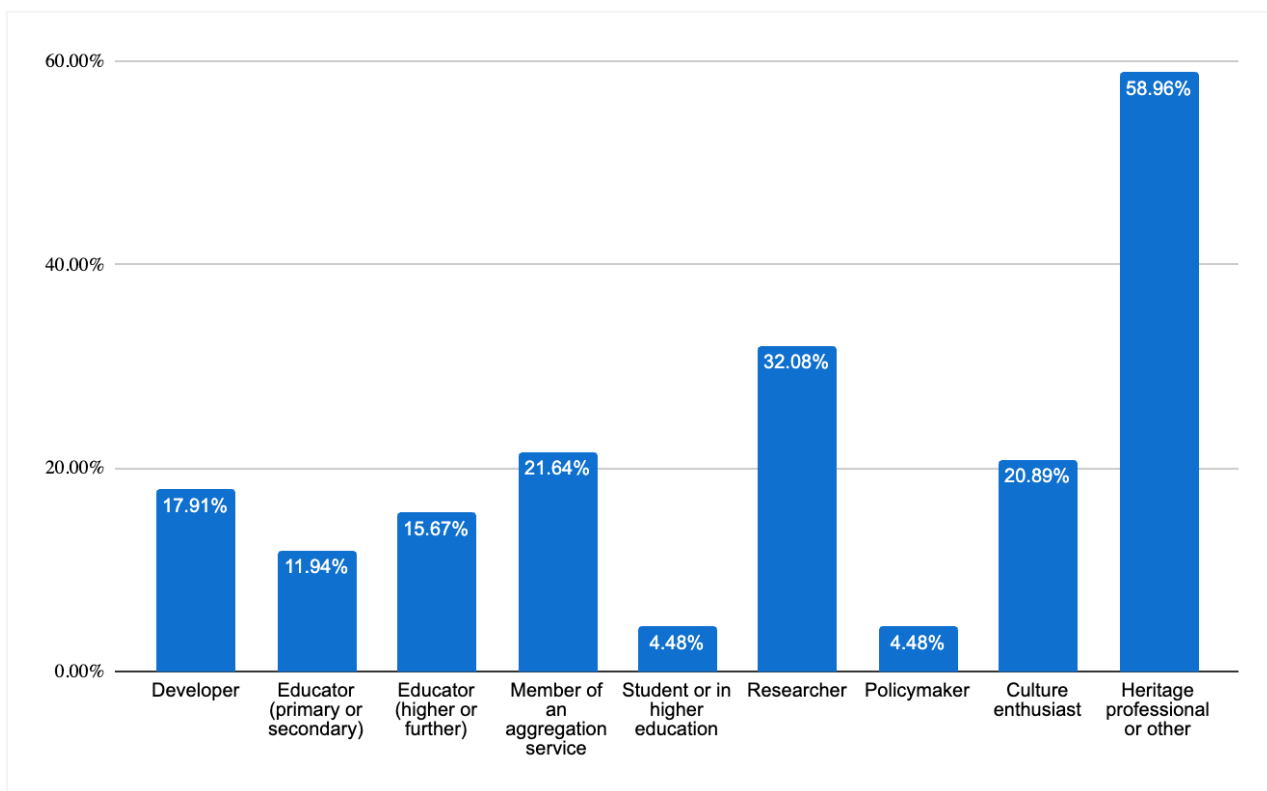
Q4 How do you engage with enrichments? (multiple answers are possible)

Answered: 134		
Choices (multiple possible)	**Response count**	**Response percent**
I produce enrichments	57	42.54%
I (re)use enrichments	49	36.57%
I develop tools to facilitate the production of enrichments	43	32.09%
I organise enrichment activities	44	32.84%
I am generally interested in the topic of enrichments	80	59.7%
Other	13	9.7%
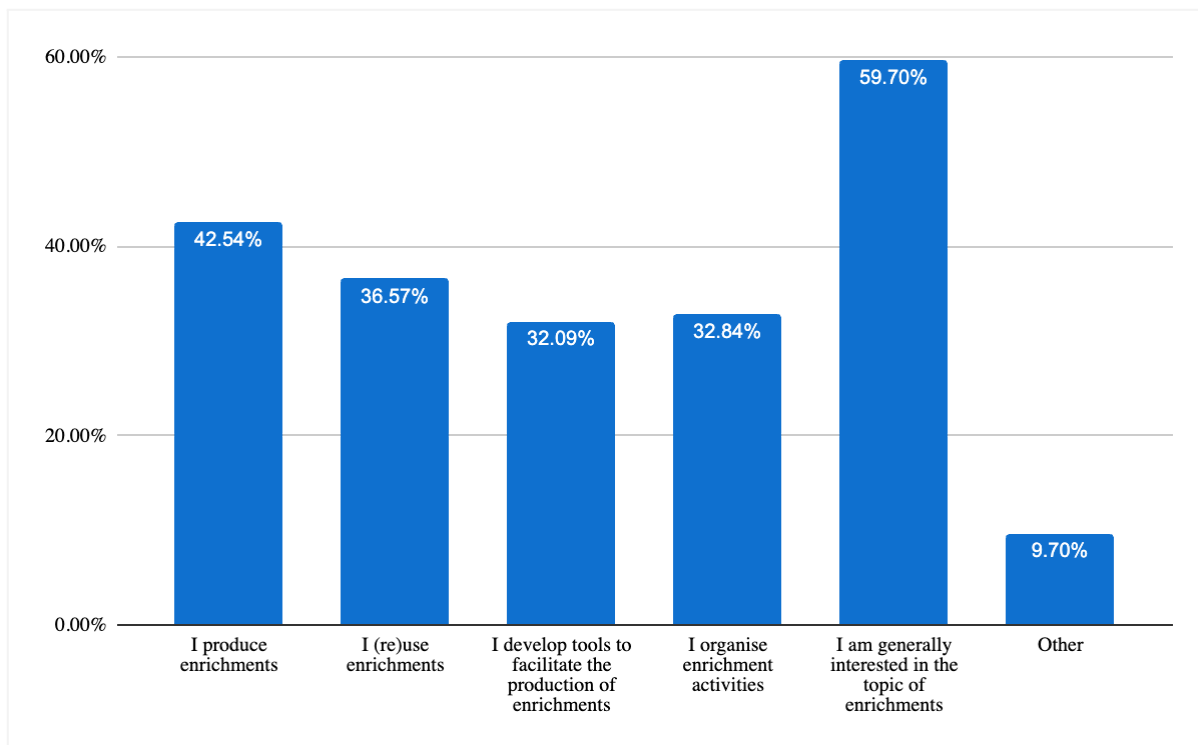
Q5 Which of the following types of enrichments are you most interested in? (multiple answers are possible)

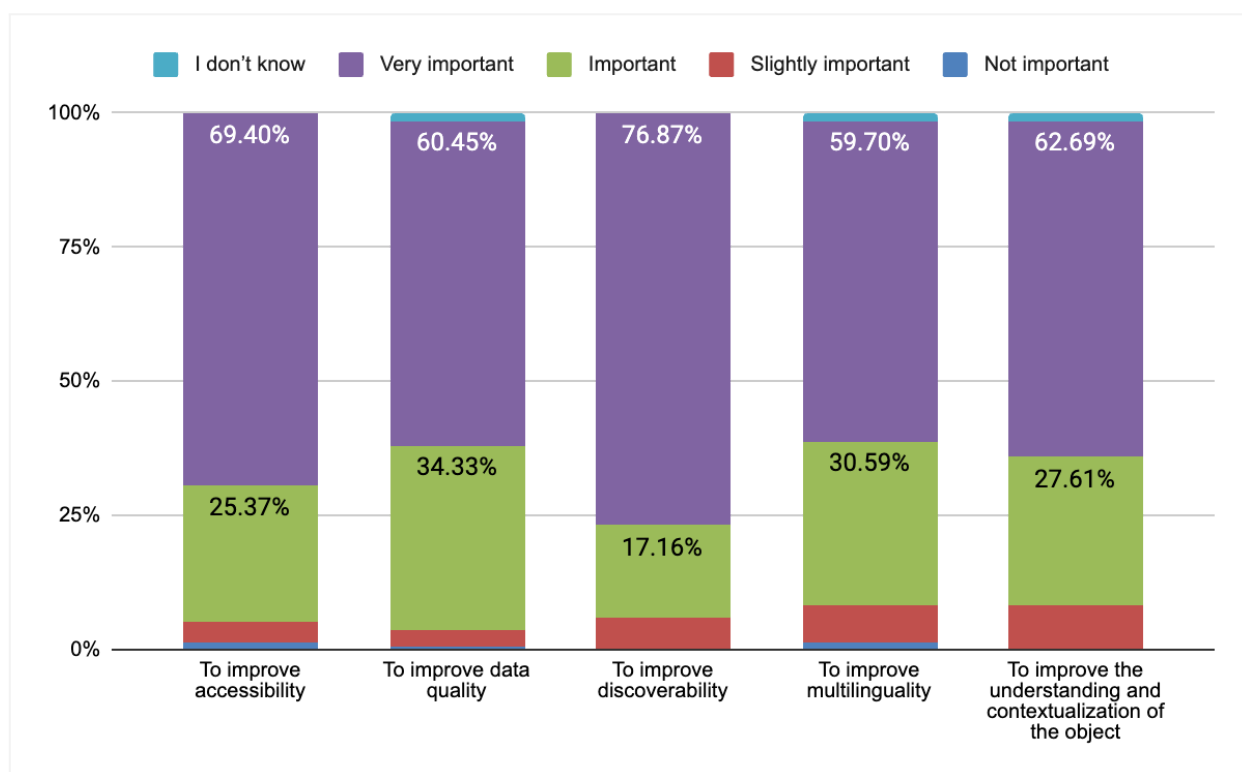| Answered: 134		
Choices (multiple possible)	Response count	Response percent
Metadata enrichments	120	89.55%
Transcriptions	71	52.99%
Translations of transcription	51	38.06%
Captions	35	26.12%
Subtitles	34	25.37%
Other (Please specify)	23	17.16%
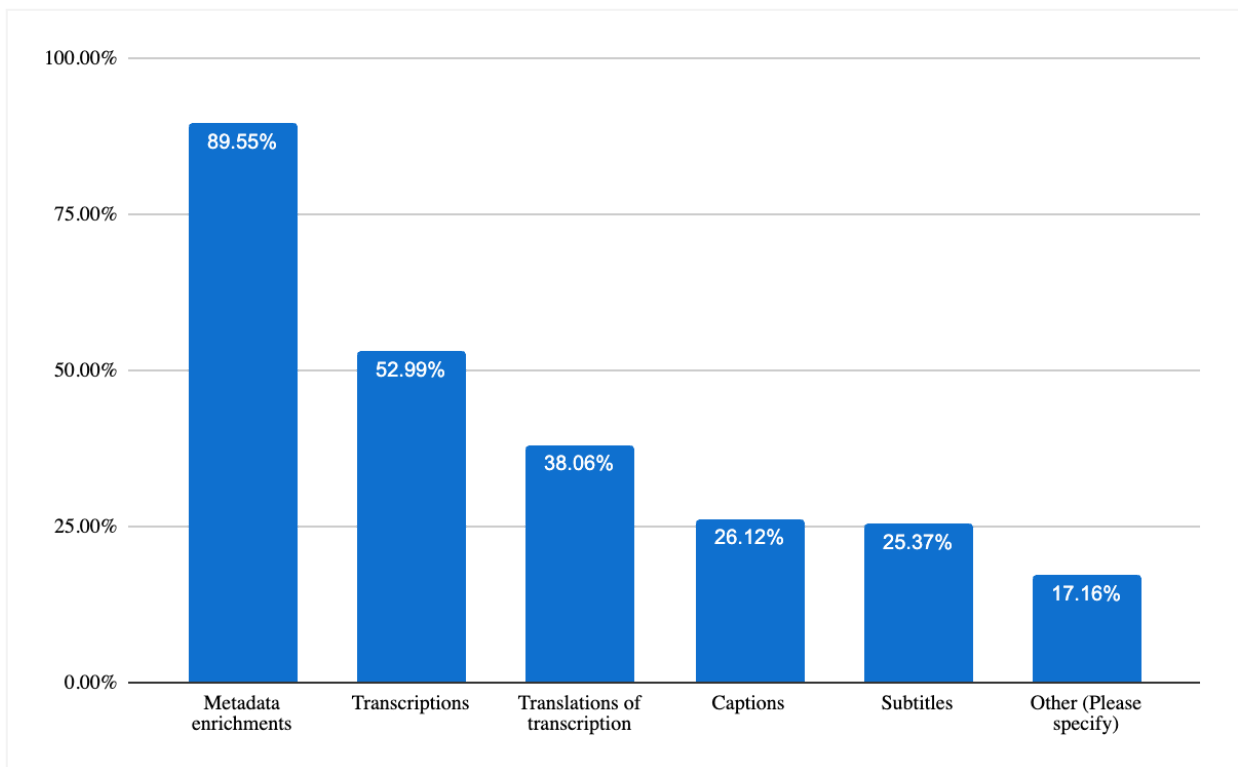
Participants provided additional information about the types of enrichments they are interested in: image recognition; user-generated responses to objects; AI enrichments; manual labelling using NLP Annotation tools; digital storytelling; option to source and share stories from the public; metadata and content enrichments for photographs by using machine vision and automatic keyword generation; comments, links to other digital resources related to the object; collection [of] new CH materials; image recognition on photo / video / film collections; representation of the qualities of the image, for example colours, recognized forms etc; adding linguistic annotation to text transcripts in order to enhance search; content data enrichment; artificial intelligence including image and facial recognition, automated translations; contemporary and artistic approaches, techniques etc.; geodata (geotagging), rephotography; enrichment of 2D images and 3D models; handwritten text recognition; annotations.

Data quality

Q6 Where do you see the value of enrichments for cultural heritage data?

Answered: 134					
Statements	Not important	Slightly important	Important	Very important	I don't know
To improve accessibility	2 (1.49%)	5 (3.73%)	34 (25.37%)	93 (69.4%)	0 (0%)
To improve data quality	1 (0.75%)	4 (2.99%)	46 (34.33%)	81 (60.45%)	2 (1.49%)
To improve discoverability	0 (0%)	8 (5.97%)	23 (17.16%)	103 (76.87%)	0 (0%)
To improve multilinguality	2 (1.49%)	9 (6.72%)	41 (30.59%)	80 (59.7%)	2 (1.49%)
To improve understanding and contextualization the object	0 (0%)	11 (8.21%)	37 (27.61%)	84 (62.69%)	2 (1.49%)

22 participant provided information about areas where they see the value of enrichments, which can be summarised as follows: digital storytelling; developing new services and functions; enabling visualisation of relations in the data; improving relevancy of metadata; connecting items between collections by e.g. emphasising previously hidden relations between cultural heritage objects; improving reuse potential of data in education and research; making metadata more interesting for audiences.

Q7 Please elaborate briefly on your previous answer

Answered: 52. Open text question. A few answers to this question are provided below with an indication of the profile of the respondent.

[*researcher*] Enriching data for collections is essential for ensuring that our institutions are inclusive, accessible and collaborative.

[*educator (primary or secondary education)*] Cultural heritage is very important in every child's life so all of the enrichments given should be updated, accessible, improved, easy to discover, and in many languages also in context easier to learn and remember.

[*culture enthusiast, researcher, developer, educator (higher or further education) and heritage professional*] Pressure on programme producers and educators is higher than ever before. Whatever (1) enhances our ability to search datasets and (2) makes it more accessible to a greater number of people, is something I welcome.

[heritage professional] The improvement of accessibility and discoverability would contribute to our users. They could find what they are looking for much easier with a lot less effort on their side. Improvement of the data quality would be beneficial for the users as well as for the institution because of the long-term preservation.

[developer] Data quality is important but we see it as a means to improve other objectives, for example usability.

[member of an aggregation service and heritage professional] Without knowing the source of the enrichment there's a need for caution, it can add value but also misdirect (for example if place names or peoples' names aren't disambiguated, enrichments can point to the wrong thing).

[heritage professional] Having access to transcribed text for example is important for researchers, for people with visual impairment, and for discoverability by search engines. The same goes for descriptions of visual pieces. Multilinguality is relevant especially when we think about text written in languages other than English.

[Heritage professional] Enrichments in cultural heritage are beneficial and work in a multitude of ways. They improve the context and description of items, and aid in research by presenting several new terms for search and recontextualising, offering the opportunity to new audiences to discover and access heritage in multiple available formats.

[member of an aggregation service and heritage professional] Enrichment helps to find data, as does multilinguality. The overall aim would be to improve understanding and contextualization. Here multilinguality also helps a lot.

Q8 Enrichments may not always be 100% correct. For example, they may be imprecise or inaccurate. How important is the correctness of an enrichment when you (re)use enriched data?

Answered: 103. Open text question. A few answers to this question are provided below with an indication of the profile of the respondent.

[*heritage professional*] This is very important - we've already had an example where a record was incorrectly enriched (interviewee shared name with someone famous) and the famous person was upset to be misidentified as interviewee as it looked like they had been talking about a deceased family member in a derogatory manner. So I think we should tread extremely carefully with automated enrichments.

[*cultural enthusiast and policy maker*] Apparently, data cleaning is a critical aspect of trustworthy data management, as it ensures that data is free from errors, inconsistencies, and inaccuracies that may result from the use of faulty enrichment techniques. While data enrichment can provide significant benefits in terms of enhancing the value of data, it is important to recognize that the correctness of enriched data is pivotal for its successful reuse.

[*member of an aggregation service and cultural heritage professional*] 100% important in metadata. 75% importanta in audio captions, translations and subtitles

[*researcher and student or in higher education*] As correct as possible and including sources. In cultural heritage there's often not a single interpretation, so the viewpoint should be clear.

[*member of an aggregation service*] Very important, otherwise there is the risk of adding wrong information to content; by doing this we run the risk of losing trust (trust of the content providers and the audience)

[*member of an aggregation service and developer*] It is very important. If you're searching for certain information and you found some records you should be able to rely on their accuracy. Additionally unreliable results introduce a kind of noise in the search results which is hard to detect and eliminate. Also when an inaccurate enrichment is exposed it may be shared further by other services and spreaded over the Internet. This may be later hard or impossible to clear.

[*researcher and heritage professional*] It is very important, therefore it is essential that every enrichment is checked. Especially automatic enrichments should be double checked, as they are sometimes not correct, i.e. appropriate for a certain context.

[*member of an aggregation service*] Better somewhat imprecise enrichments than poor or no enrichment. Proper disclaimers should be included and the audience asked to contribute in the correction of eventual errors.

[*member of an aggregation service and developer*] It very much depends on the goal, what you want to do with the data in the first place: If its use is a scholarly purpose, the correctness is of highest importance, if the goal is mere findability the correctness is less important, you might find what you are looking for anyway. (Given a search option that offers fuzziness.)

[*researcher and heritage professional*] Absolute correctness is not a requirement, especially for transcriptions. But it is nice if, as a user, you are informed about the quality of the enrichments.

[*culture enthusiast, researcher and developer*] It is important to know for automatic enrichment the confidence score that the system has assigned to those. In this way the user of enrichments can better work with them. Depending on the task not 100% correct is better than no enrichment at all (e.g. for improving the discoverability of collection items). In other cases though, e.g. when searching for something specific, the correctness of the enrichment is key to avoid false positives in the results.

[*researcher*] I think there should be some room for both imprecision and inaccuracy especially in metadata enrichment to leave an opening for multiple interpretations

[*researcher and heritage professional*] I think that the correctness of enrichment data will be a moving target. The important thing is to create frameworks that provide transparent information about the quality of the data and signal possible issues with correctness. I find it interesting that this issue is in many ways similar to that of content moderation and dealing with misinformation.

Transparency

Q9 Is the distinction of enrichments from authoritative data provided by cultural heritage institutions important to you?

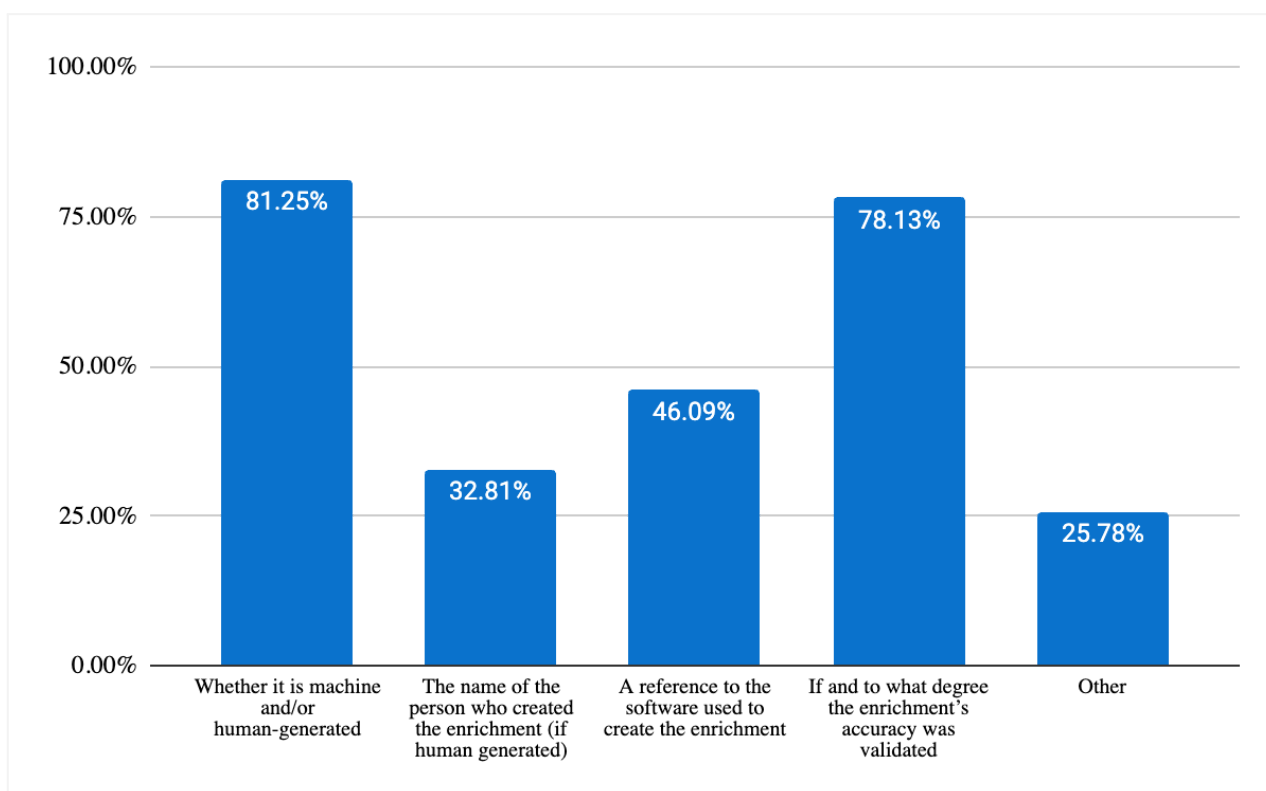| Answered: 131		
Choices	Response count	Response percent
Not at all	3	2.29%
To a small extent	10	7.63%
To a medium extent	39	29.77%
To a great extent	68	51.9%
I don't know	11	8.39%
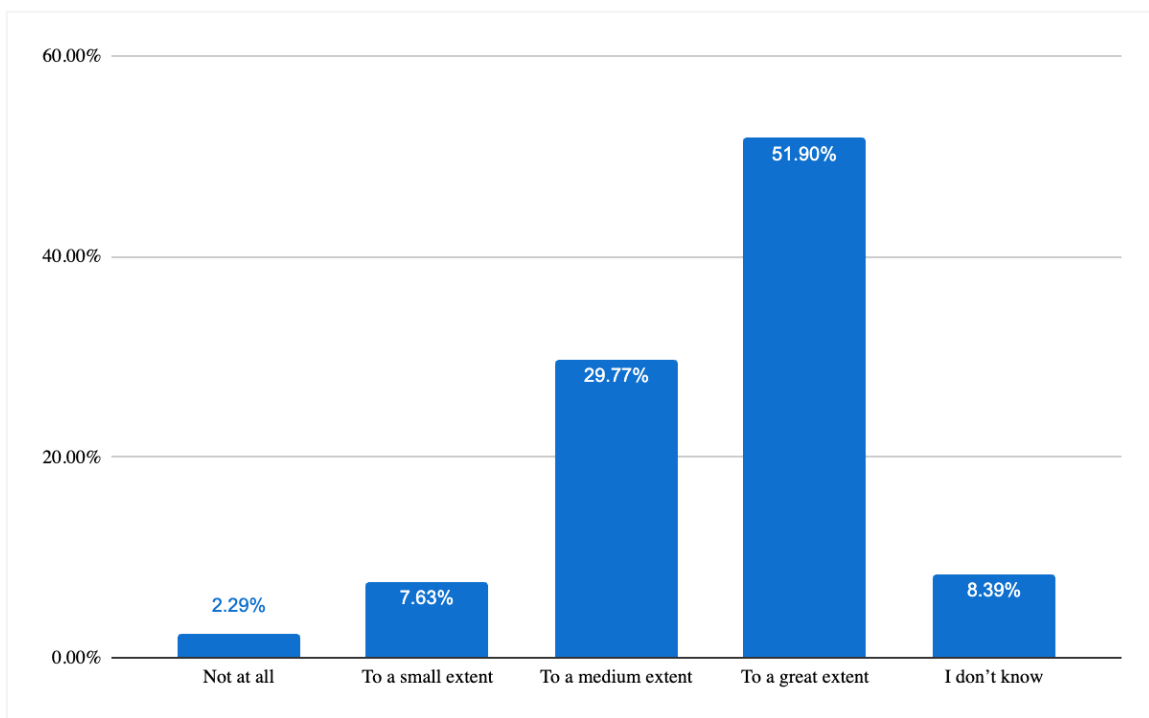
Q10 Which information about the production or source of the enrichment would you like to see? (multiple answers are possible)

Answered: 128		
Choices (multiple possible)	Response count	Response percent
Whether it is machine and/or human-generated	104	81.25%
The name of the person who created the enrichment (if human generated)	42	32.81%
A reference to the software used to create the enrichment	59	46.09%
If and to what degree the enrichment's accuracy was validated	100	78.13%
Other (Please specify)	33	25.78%

33 participants listed other types of information they would like to see about the production or source of the enrichment, which can be summarised as follows: date of enrichment creation; context in which enrichment was created; why it was created; if human generated: role (function) of the person and their (educational) background, name of the organisation they work for; if machine generated: software settings.

Q11 Please elaborate on why you would like to see the information you indicated in the previous question made available, when (re)using enrichments.

Answered: 73. Open text question. A few answers to this question are provided below with an indication of the profile of the respondent and divided by the answer provided to the previous question.

<u>Answered 'if and to what degree the enrichment's accuracy was validated' to question 10</u>

[*researcher, developer, educator (higher or further education) and heritage professional*] This would help me predict how much time must be given to checking of work, and correction of transcriptions.

<u>Answered 'whether it is machine and/or human-generated', 'a reference to the software used to create the enrichment' and 'other' to question 10</u>

[*heritage professional or other - commercial photographic archive*] Accuracy and authenticity and a reasoning, mostly for results that appear which are so clearly wrong. CHIs should be the authority and when basic mistakes are visible it reflects poorly on them but it is more understandable if it is a machine made reference.

<u>Answered 'whether it is machine and/or human-generated', 'if and to what degree the enrichment's accuracy was validated' to question 10</u>

[*developer*] Information or data. Knowing how to distinguish between the work done by a human or a machine on initial data allows one to better understand the direction that has been given to this interpretation and enrichment and to contextualise it in order to measure its limits or interests.

[*developer*] Whether it is machine or human I would find it relevant to a small extent, because I would assume enrichment is generated by human and machine. But it might be interesting to know from a system evaluation perspective, if too many mistakes are being generated by the machine, it might be possible to let the human also verify automatic annotation. If data enrichment is community sourced I find it better not to disclose personal data. If it is made by staff then it is relevant to know who created the enrichment. If I would be coordinating a specific team of annotators I would definitely want to know who created the enrichment. So it depends on the situation. I like to know the date created, also to understand the process and trace a history of annotations.

[developer] Provenance information is key for users, e.g., researchers that build their research on data they get from a specific source. I think at least the responsible institution should be indicated, but maybe also the person (attribution of a nano publication). This depends a bit on the context (Is it a library or a research institute in an academic environment? etc.).

<u>Answered 'whether it is machine and/or human-generated', 'a reference to the software used to create the enrichment' and 'if and to what degree the enrichment's accuracy was validated' to question 10</u>

[*researcher*] The above all relate to trust so you can make a judgement on the likely accuracy of the data.

[heritage professional] Especially when the enrichment is machine-generated, it is very important to know exactly how the metadata was generated and if the accuracy of the enrichment was valid. In that case, when mistakes show up, we have a better understanding why especially this mistake showed up and in which other places we can expect those errors. This is of course different when the enrichment is human generated, because then it is more likely that the mistake stands alone.

[researcher and heritage professional] […] as a user, it is nice to have a good sense of the quality of enrichments. For this, information about the production or source of the enrichments is essential. The name of the person who created the enrichment is of less importance to me, but the person's background might be of interest (e.g. professional / volunteer; social background, etc.).

Answered 'whether it is machine and/or human-generated', 'the name of the person who created the enrichment (if human generated)' and 'if and to what degree the enrichment's accuracy was validated' to question 10

[culture enthusiast and policymaker] When it comes to enrichment data, it is crucial to ensure that the publisher and accuracy of the data are guaranteed. Therefore, I would like to see information about the institution that produced or provided the enrichment data, as well as any accreditation or certifications they may have. It is also important to know how the data was collected, processed, and validated, as well as any potential biases or limitations.

[heritage professional or other - cultural diplomacy] When using such information as a basis for scientific reflections or as a for qualitative statements to third parties, it is important to be able to assess how it was generated (similar to a scientific text or study).

Answered 'whether it is machine and/or human-generated', 'the name of the person who created the enrichment (if human generated)', 'a reference to the software used to create the enrichment' and 'if and to what degree the enrichment's accuracy was validated' to question 10

[member of an aggregation service and developer] Only if I have these bits of information available can I estimate its value/correctness/reliability.

[researcher] I think the more information we have the better. Then when mistakes are found they can more easily be examined and corrected.

[member of an aggregation service and policymaker] Because you need all this information to decide whether the enrichments are usable in your case.

[researcher, educator (higher or further education) and heritage professional] The production context of enrichments helps understand and interpret information about objects. Also, it helps check the accuracy and correctness of enrichments.

[heritage professional] To avoid circular reporting/citogenesis. To allow data re-users to choose whether (and which) enrichments to reuse. We work with pseudonyms as we place less value on the individuals formal credentials, and rather on what their work on the platforms show. However, it is important to be able to track what enrichment has been done by what user to be able to handle e.g. recurring mistakes or vandalism. It's important that

the information on the source of the enrichments is machine readable and included for any data re-users, so that they can choose which data to reuse.

Q12 If you already contribute enrichments, or plan to do so (e.g. by participating in a crowdsourcing campaign), would you like your name acknowledged?

Answered: 131		
Choices	Response counts	Response percent
Yes	47	35.88%
No	26	19.85%
I don't know	43	32.82%
I'm not planning to create enrichments	15	11.45%
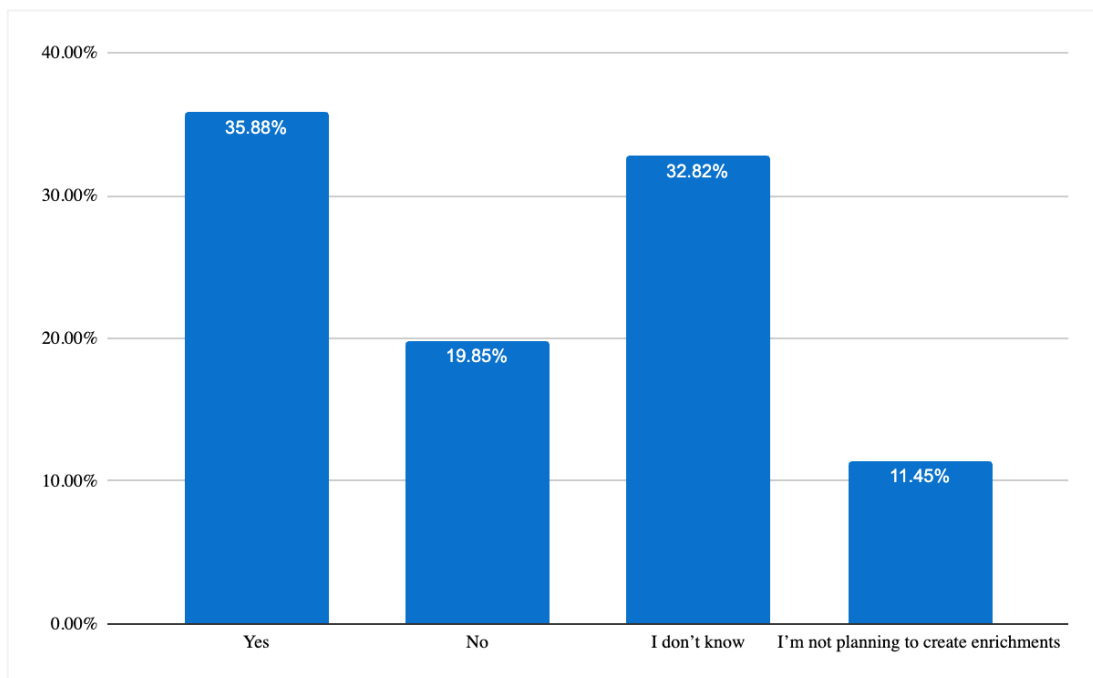
Reusability and interoperability

Q13 For what purpose(s) do you or would you (re)use enriched data?

Answered: 79. Open text question. Answers to this question are not provided in this document.

[culture enthusiast and policymaker] Enriched data can be (re)used for various purposes, including gaining new insights, improving decision-making, and identifying trends and patterns [...] Enriched data can also be used to identify common challenges and issues in the preservation of cultural heritage, leading to more effective solutions and collaborations. I think, overall, (re)using enriched data can lead to better decision-making, increased knowledge, and improved preservation of cultural heritage.

[member of an aggregation service] To increase the usability of the contents we aggregate for the target population. This may help to increase the access for such resources and this way the readiness of the content providers to continue and provide additional resources.

[heritage professional] To have more relevant information in one place + to create more social and cultural value for researchers and the général public + to update our metadata and pictures from other sources (we have many b/w pictures), to stay relevant.

[researcher, policymaker and educator (higher or further education)] Education, research.

[developer] As a developer, I am interested in the uses made of these enriched data and the way to exploit them by tools. I am collaborating on a project to collect and valorise orphan collections in rural areas and I am also required to create and manage metadata to build collections. Thus, my reflections are double, as a curator and as a technician. The enrichment of data challenges me in these two dimensions, in particular for the enhancement of media by Web tools, and the addition of intelligence from metadata for the benefit of end users.

[researcher] Research, publication, public outreach, teaching,...

[*educator (primary or secondary education)*] Teaching in classroom

[culture enthusiast, researcher and developer] Research purposes, new services

[*educator (primary or secondary education)*] I would use it in my English language classes.

[researcher and developer] Improve the findability of collection items; enhance the appearance of data-intensive websites by adding images on the topics, etc.

[member of an aggregation service and heritage professional] By returning enriched data and translations, and incorporating automated curated new terms into your database, you are able to create new finding tools for your own collection, as well as cross-referencing between collections outside of your museum. This can create new research opportunities and increase the discoverability of your collection. The incorporation of enriched data and

translations into your database can help improve the accessibility and discoverability of your collection. It can also help to uncover new connections and relationships between different objects or collections within your museum. By using automated curation and translation tools, you can efficiently add new terms and metadata to your database, making it easier to search and explore your collection. Furthermore, by cross-referencing your collection with other collections outside of your museum, you can create new research opportunities and uncover new insights. This can help to expand the reach and impact of your collection, and facilitate collaboration with other institutions and researchers. In summary, incorporating enriched data and translations into your database can help to enhance the discoverability and accessibility of your collection, while cross-referencing with other collections can lead to new research opportunities and insights.

[researcher and educator (higher or further education)] For finding primary sources for research purposes and for educational materials. For illustrating research publications, educational content and project websites.

Q14 Do you agree with the following statement? *Data resulting from enrichment efforts should have no new copyright protection. Barriers to (re)using data resulting from enrichment efforts should be removed to the extent legally and ethically possible. That is, no new rights should be applied to enrichments, whether they are machine or human generated.*

Answered: 128		
Choices	Response count	Response percent
Strongly disagree	6	4.69%
Disagree	16	12.5%
Agree	37	28.9%
Strongly agree	59	46.09%
I don't know	10	7.81%
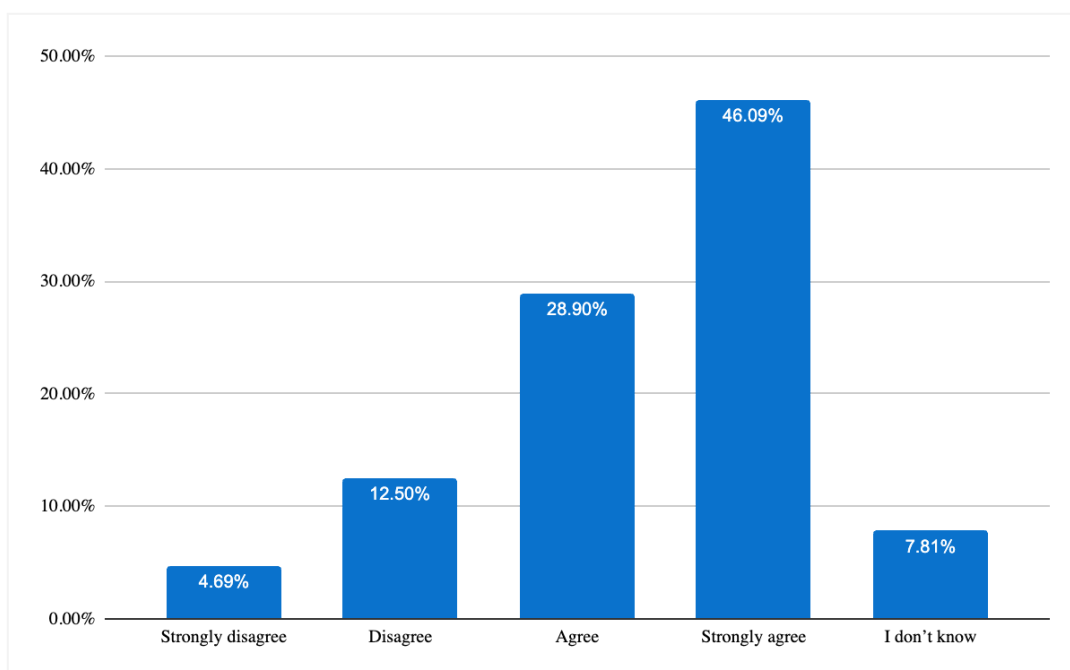
Q15 Please elaborate briefly on your previous answer

Answered: 68. A few answers to this question are provided below with an indication of the profile of the respondent.

[*cultural heritage professional or other*] [*strongly agree*] The above statement safeguards an important underpinning of our IP system: that facts and data should and cannot be protected because they would stop being useful if expressing them would be the exclusive rights of those who have expressed them. Copyright protection arises when there is a sufficient level of creativity which is an important aspect of ensuring the balance of copyright. This principle should be strongly defended by cultural heritage institutions who have a mission to serve public access, which is already hampered by overly broad copyright laws. Extending copyright protection to cover facts, data and other non original forms of expression (computer or human generated) would further limit the ability for cultural heritage institutions to fulfil their missions.

[*member of an aggregation service*] [*strongly agree*] Enrichments are usually "small coin" work and do - from my perspective - not justify a new copyright.

[*researcher*] [*strongly agree*] I think it difficult to make a blanket statement like this - so is context dependent. However as a general principle adding further layers of IP to the original is not helpful and will add further barriers to the reuse of the enrichment.

[*researcher*] [*strongly agree*] Especially if using crowdsourcing, adding copyright protection would be an abuse of the unpaid labour of volunteers.

[*cultural heritage professional or other*] [*strongly agree*] Data Enrichment is an extension of knowledge and lies under the access to digital heritage. Digital heritage as all heritage ought to be publicly accessible and related information should remain free to use to further research, expand the knowledge of the general audience; and reduce, to the extent possible, the tendency of "hoarding" information and knowledge behind paywalls. This in turn can also generate different ways for remuneration within the copyright policy system as a whole as for example the rising participative financing.

[*researcher*] [*strongly agree*] Open data is the way forward. We need to be inclusive of different uses (within the bounds of ethics of course) though there should of course be proper citation.

[*cultural heritage professional or other*] [*strongly agree*] Data on cultural heritage should generally be as widely/openly available as possible in my opinion. I see no upside in copyright protecting data resulting from enrichment efforts.

[*educator (primary or secondary education)*] [*strongly agree*] Whatever about current rights remaining in force - no new rights should be applied to an object just because the metadata to it is enriched - if this is the case then it should only be applied to the metadata and that seems pointless to me and the purpose of what metadata is for.

[*member of an aggregation service*] [*strongly agree*] I think all CH data should be freely available to everyone. This would apply especially to new data added.

[*cultural heritage professional or other*] [*strongly agree*] If the enrichment is about adding, connecting, representing, etc. mere facts, they should not be copyrightable (and are not copyrightable in most jurisdictions). If the collection items give way to new artistic works in their own right (poetry, remixes, visualisations with a particularly creative touch, etc.) they are obviously under copyright, and it is up to the creators to decide to what extent they would like to share them.

[*member of an aggregation service*] [*agree*] Agreed, but in the case of human-made translations and/or subtitles, they might just fall under copyright protection. But I would advocate for them to be offered under an open licence.

[*student or in higher education (higher or further education)*] [*agree*] The use and re-use of CH should be easy and facilitated for every kind of end-user, context or entity (organisation).

[*researcher*] [*disagree*] I would like to change the phrase: We need authors and sources of information, but as in scientific practice copyright or licensing could enable re-use.

[*cultural heritage professional or other*] [*disagree*] Surely a lot of work will go into producing some enrichments and possibly the people putting in the work will need to monetize the enrichment to continue being able to work and produce more enrichments.

[*researcher*] [*disagree*] I think it depends on the quality of the enrichments. There might be hybrid business models where enrichments can be free or available for a fee, depending on quality or the general process through which these were created (public money vs own investment of organisation).

[*member of an aggregation service*] [*disagree*] I think that metadata enrichment should be labelled as CC0 in the context of the Europeana landscape.

[*cultural heritage professional or other*] [*strongly disagree*] There is an overwhelming amount of intellectual effort (property) that is used to create enrichment tools. And there are also masses of intellectual property used to train these tools. And finally the resulting enrichments can be accurate copies - text or image - of in copyright works.

[*cultural heritage professional or other*] [*strongly disagree*] There should be new rights associated with data enrichment. At the same time, efforts should be made to ensure that such enrichments are openly licensed, ideally CC0.

Collaboration and support

Q16 What types of partnerships would be most helpful to you as you collaborate with other people in the area of enrichments?

Answered: 68. A few answers to this question are provided below with an indication of the profile of the respondent.

[*heritage professional*] Projects with cultural heritage institutions in the same domain, collaboration with universities etc.

[*heritage professional*] Exchange of best practice, especially in relation to critical cataloguing and machine-generated enrichments.

[*researcher, policymaker, educator (higher or further education*] Technical support.

[*member of an aggregation service and student or in higher education*] Cultural collaborations with museums and cultural institutions, local communities, students, legal entities, NGOs

[*developer*] Undoubtedly, with museums and collection holders.

[*researcher*] Partnerships with IT developers, diversity and inclusion groups (to advise on accessibility needs)

[*member of an aggregation service*] Partnerships with other institutions for sharing thesauri and controlled vocabularies, as well as tools for enrichment.

[*member of an aggregation service and policymaker*] Communities, communities and communities.

[*heritage professional*] Partnerships with experts in natural language processing and corpus linguistics are vital for adding the best possible linguistic annotation to texts.

[*member of an aggregation service*] Partnerships with other aggregating services to share our enrichments (if wanted) and to be able to pick up on enrichments others have already added. In our context, this would especially apply to our connection to national aggregators, but also some regional and thematic aggregators (dealing with archival collections). Partnerships with other aggregators and similar service providers when it comes to running crowdsourcing campaigns with the aim to enrich data e.g. on a shared topic of interest. Partnerships with collection management system providers to explore the opportunities of giving enrichments back to the cultural heritage institutions. At the moment, it seems that enrichments are often happening on the aggregation level only, while it could be useful to have those enrichments happening at the source level in the first place or at least having an opportunity to integrate enriched data with the source data somehow (if wanted by the data owners).

[*heritage professional*] Share methodology, business cases, technology, workflows, case studies. Purpose would need clear integration to organisational priorities (collections care and compliance, align with public programme activity).

[*researcher, educator (higher or further education) and heritage professional*] We collaborate with artists, experts (specialists culture knowledge holder) from communities of origin.

[*heritage professional*] Sharing knowledge and experiences.

[*culture enthusiast, developer and heritage professional*] I believe shared policies and technical standardisation are very important. Also the legal aspects.

[*wikimedian*] We act in an area that lies between the community (crowdsourcing on Wikimedia's platforms) and institutions. We are a project organisation and like to work actively with minorities that are underrepresented on Wikimedia's platforms because that knowledge is often lacking and in need of expansion. The majority of our heritage projects are in collaboration with other institutions. We believe that it is crucial to form partnerships with the groups/people from where the material comes, and include them as participants in the enrichment work according to the motto "Nothing about us without us".

Q17 What kind of support would you need to conduct enrichment efforts?

Answered: 65. A few answers to this question are provided below with an indication of the profile of the respondent.

[*culture enthusiast and policymaker*] Conducting enrichment efforts can require a significant amount of resources, including funding, access to data sources, and expertise in data management and analysis. In addition to these resources, having a supportive community can be crucial in facilitating effective enrichment efforts. One type of support that would be helpful is access to a professional community where experts in data management, analysis, and related fields can provide guidance and advice on enrichment initiatives. Such a community could provide a platform for sharing best practices, exchanging ideas, and collaborating on projects. Ideally, participation in this community should be conditional on revealing institutional affiliation and other relevant details to ensure that all participants have the necessary expertise and experience to provide meaningful contributions. This would help to ensure that the community provides high-quality and reliable support to enrichment efforts. In addition to community support, other forms of support that could be helpful include access to high-quality data sources, funding for research and development, and access to training (for example, we had [this webinar](#) last month) and resources on data management and analysis. By providing these types of support, organisations can help to facilitate effective enrichment efforts and promote progress in various fields.

[*researcher*] Organisational and administrative and legal frameworks.

[*member of an aggregation service   and heritage professional*] Partners that help to ensure the quality of enrichments. Many small institutions would be willing to work on enrichments but don't have the capacity to check the quality but the knowledge to contribute.

[*heritage professional*] money to pay specialists.

[heritage professional] Awareness of ongoing initiatives and opportunities to participate. Possibility of remuneration. Capacity building activities.

[heritage professional] We are at the beginning of implementing machine-generated enrichments, so we will need support for this first, exploratory phase. This can be in the form of advice on which applications to use or avoid, which are the best match for our specific situation and collection. Etc.

[researcher] Human and technical support, upgraded training on latest technologies.

[heritage professional] Financial to foster these relationships, or to afford the user friendly enrichment tools. and/or tangible online tools which facilitate enrichments.

[researcher] Free handwriting transcription tools.

[heritage professional] Mostly technical support to set up an enrichment process. And enough resources (time and staff) for validation.

*[developer]* Sometimes the ways of evaluating enrichments.

[member of an aggregation service] technical support and funding for crowdsourcing campaigns.

[*member of an aggregation service and heritage professional or other*] A software tool for easily adding enrichments by humans; software to automatically propose enrichments, expertise to (automatically) assess quality.

Participation, diversity and inclusion

Q18 How valuable do you believe are contributions of individuals (i.e. crowdsourcing) for enrichment efforts?

Answered: 123		
Choices	Response count	Response percent
Not valuable at all	0	0%
With limited value	11	8.94%
Valuable	52	42.28%
Very valuable	56	45.53%
I don't know	4	3.25%
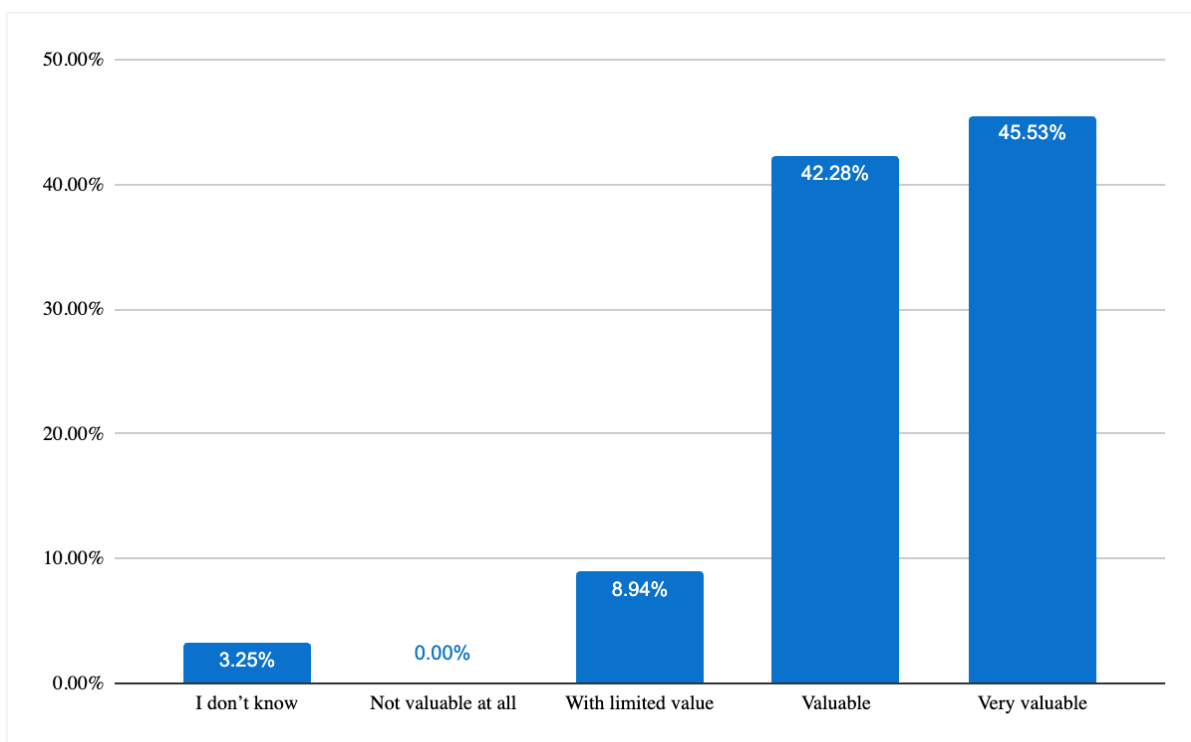
Q19 What, if any, concerns do you have about crowdsourcing in cultural heritage?

Total responses: 77. A few answers to this question are provided below with an indication of the profile of the respondent.

[*culture enthusiast and policymaker*] Crowdsourcing in cultural heritage can offer many benefits, such as increased community engagement, enhanced data quality, and the potential for new discoveries and insights. However, there are also some concerns that need to be addressed to ensure that crowdsourcing initiatives are effective and sustainable. One significant concern is privacy and security. As the question suggests, developers working on cultural heritage projects need to be mindful of the privacy issues related to publishing data about the owners of cultural properties and building plans. It is important to ensure that

appropriate safeguards are in place to protect sensitive information and prevent unauthorised access or use. Another concern is the potential for bias or inaccuracies in the data collected through crowdsourcing. To address this, it is essential to establish clear guidelines and quality control measures to ensure that the data collected is accurate, reliable, and representative. Finally, there may be concerns about the sustainability and long-term viability of crowdsourcing initiatives. It is important to ensure that there is ongoing support and resources available to maintain and update the data collected through these initiatives, and that there is a clear plan for how the data will be used and shared over time. By addressing these concerns and implementing appropriate safeguards and quality control measures, crowdsourcing can be a valuable tool for cultural heritage research and management. However, it is important to approach these initiatives with care and consideration to ensure that they are effective, ethical, and sustainable over the long term.

[*member of an aggregation service and developer*] A certain level of experience with the data and the topic the data covers needs to be there in order to make the enrichment useful for all purposes, especially any scholarly use. That level of expertise should be proven somehow.

[*researcher and heritage professional*] Crowdsourcing produces a lot of bad and/or meaningless results, but it's also extremely valuable, so my concern is to do with how to effectively sort this different data. The system that I'm designing will include a filter, meaning that crowdsourced enrichments will be checked before going live.

[*researcher and Student or in higher education]* Intentional manipulation and bias.

[*researcher]* The contributions need to be verified.

[*member of an aggregation service and heritage professional*] Verifying opinions - is the enrichment based on one person's opinion or several? How do you cope with differences?

[*member of an aggregation service and heritage professional*] It needs to be clear that crowdsourcing should not or can not replace the important work of cultural heritage professionals and their important domain knowledge around the objects they handle and contextualise. Crowdsourcing should be targeted at either an educational focus (i.e. the subtitle-a-thon for Europeana XX that was an opportunity for students to contribute to a subtitling activity in a limited and defined subset) or support the existing work of CHI professionals for the public good and in an enjoyable way for the participants. In any way, I think that crowdsourcing is most suited for enrichments on a smaller and well-defined scale.

[*member of an aggregation service*] Speaking on behalf of our content providers, I think the concerns that I've heard the most often are (1) potential misuse, e.g. someone torpedo-ing a crowdsourcing campaign with a biassed view of things and (2) the fear that crowdsourced input requires lots of personnel time to review and verify. On the other hand, I also know of projects where crowdsourcing has been really successful as it taps into people's own histories or local knowledge (e.g. projects about identifying people from a certain village or town or region on old photographs or projects about identifying places, buildings, streets on old photographs).

[*researcher and heritage professional*] I am not sure that all people participating are reliable and knowledgeable for a certain subject.

Q20 How can enrichment efforts be designed so that minority communities can best contribute?

Total responses: 65. A few answers to this question are provided below with an indication of the profile of the respondent.

[*culture enthusiast and policymaker*] Training programs can be developed that are specifically tailored to the needs and experiences of minority communities. This can involve working with community organisations and leaders to identify the skills and knowledge needed to effectively contribute to the enrichment efforts, and developing training materials and resources that are accessible and culturally appropriate. It is also important to ensure that the data management tools and software used in the enrichment efforts are user-friendly and accessible to all members of the community, regardless of their technical expertise. This can involve providing training and support to help community members navigate and use the software effectively.

[*culture enthusiast and heritage professional*] Co-creation workshops to engage communities to provide their input.

[*heritage management consultant*] Direct outreach to each community, while at large it may seem impossible but there are several NGOs that work with minorities that can facilitate outreach. Furthermore, minorities may not be engaged with cultural heritage, or understand why their contributions are essential, and how and why digital enrichment works, these best work with participatory practices such as mapping.

[*culture enthusiast, educator (primary or secondary education) and heritage professional*] If a crowdsourcing project has a wide reach, it should represent the entire spectrum of opinion and not prioritise individual opinions.

[*heritage professional*] The interface of the systems in which the contributions can be added must be as clear and user friendly as possible.

[*researcher*] Make them part of long term partnerships where people are properly compensated for their time and given their due credit.

[*heritage professional*] Let the minority communities give input on that (instead of "us" designing for "them"). So define your user group(s) and make sure they are represented in the planning/preparation phase.

[*culture enthusiast, educator (higher or further education) and educator (primary or secondary education)*] Multilinguality, active promotion and involvement of minorities.

[*culture enthusiast, researcher and developer*] Approach such communities explicitly.

[*member of an aggregation service and policy maker*] Open, accessible in multiple languages, analog, guidance, simple.

[*heritage professional*] Don't put up any barriers to minority community participation.

[*culture enthusiast*] Engage local communities. Provide in person activities. Not only online

[*heritage professional*] As many methods and formats of feedback as possible. Actively seek engagement. Clearly describe purpose and benefit in terms of outcome to those communities and overall.

[*heritage professional*] By using materials that resonate with the minority communities.

[*member of an aggregation service and heritage professional*] Involve minorities from the start.

[*member of an aggregation service and heritage professional*] Targeting them? But how do you balance differences in world view and different triggers (i.e. what is offensive to one minority may not be so to another).

Environmental sustainability

Q21 How can the environmental impact of enriching data be mitigated through any of the following activities? (multiple answers are possible)

Answered: 120		
Choices (multiple possible)	Response count	Response percent
Selection of enrichment method(s) and technique(s)	22	18.33%
Creation of enrichment software and tool(s)	25	20.83%
Assessment of the quality of enrichment(s)	19	15.83%
I don't feel it is important	16	13.33%
I haven't considered it	64	53.33%
Other (please specify)	13	10.83%
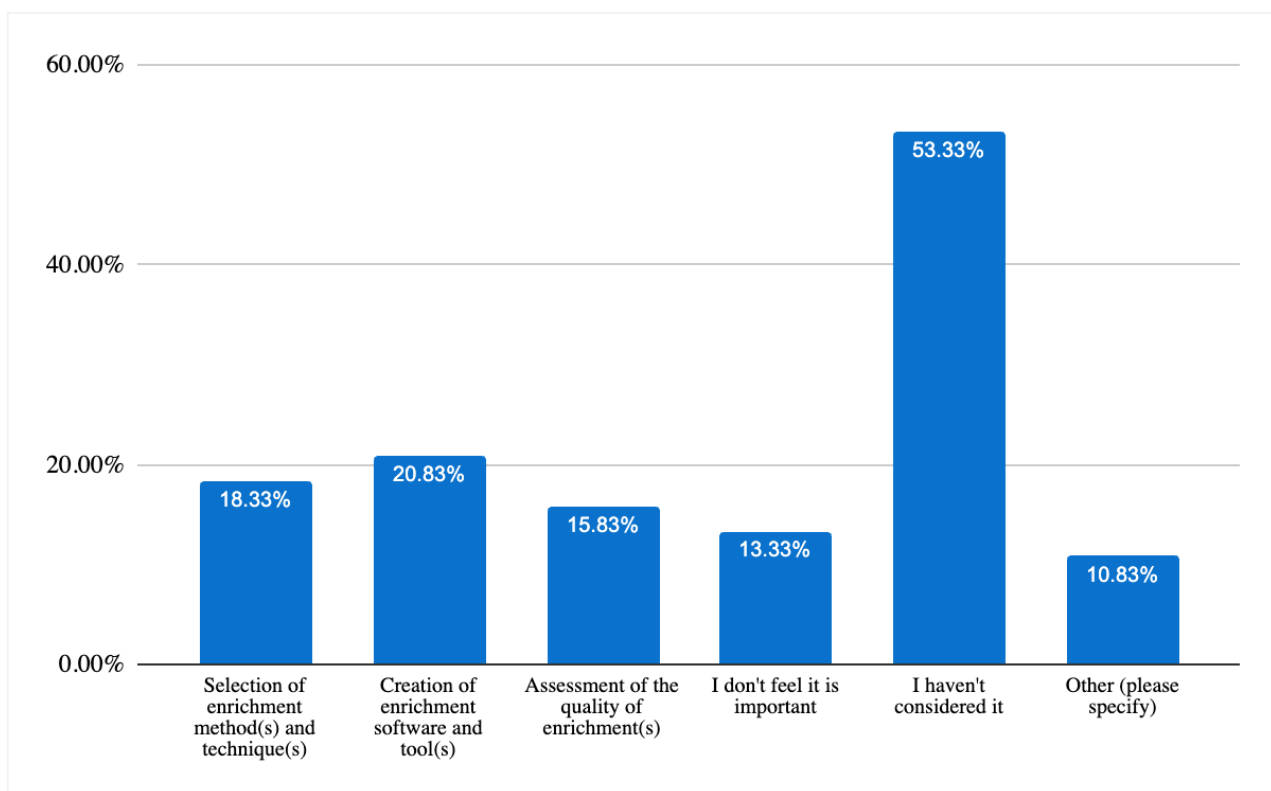
13 participants listed other activities through which the environmental impact of enriching data could be mitigated, which can be summarised as follows: data storage; prioritising human-generated over AI-generated enrichments; assessing the impact of the entire lifecycle; ensuring energy, equipment and labour are sustainable.

A few answers to this question, that were indicated in the comment box, are provided below with an indication of the profile of the respondent.

[heritage professional] As with all current practices in heritage management and digital heritage management the environmental impact assessment ought to play an important role, these could underline the policies on data enrichment by opting for more environmentally friendly methods, and inviting proposals for new less energy-demanding software.

[heritage professional] Bearing carbon footprint in mind when starting enrichment initiatives, and creating software tools.

[heritage professional] Crowd-sourced enrichment is, in this context, preferable to energy-intensive AI models, which produce content as-good-as or slightly worse than human agents, but for orders of magnitude greater power costs.

[heritage professional] User friendly enrichment software could ease the effort of the users and pull them in so they do more work.

[policymaker and educator (primary or secondary education)] I'm actually not sure how it can be mitigated - if it's all online that cuts down on air travel - but even online has some environmental effects that would need to be further mitigated.

[member of an aggregation service and researcher] I think it doesn't make any sense to consider the environmental impact of enriching data isolated from an assessment of the impact of the entire operation of the digital collection and the platform where it is hosted.

[researcher and student or in higher education] Sustainable energy, equipment and labour.

[developer] Assess the impact of the entire lifecycle. Imo, environmental impact should be analysed in relation to the entire cultural heritage repository and not in relation to data enrichment alone.

[member of an aggregation service and heritage professional] Selection of methods, techniques and infrastructure. First you need to be sure that the benefits outweigh the carbon footprint of doing the work, then you need to look at the carbon footprint and evaluate ways of reducing this, and then perhaps consider if there are ways of offsetting this impact.

Additional comments

Q22 What haven't we asked you about enrichment efforts that we should have, and what's the answer?

Answered: 30. A few answers to this question are provided below with an indication of the profile of the respondent.

[*researcher and student or in higher education*] The purposes of enrichment vary. For whom are they for?

[*heritage professional*] Their communication: for crowdsourcing, more awareness in the cultural heritage and education sectors for their use and the public for accessing digital heritage. Ethics, while I personally advocate for crowdsourcing and machine enrichment other stakeholders might have differing opinions that may be beneficial to the policy. Values of cultural heritage, whether these tools are using values-based methodologies and if they should.

[*researcher and heritage professional*] I would like to participate in enrichment activities when I am retired (and that would be in a few months), but I don't know whom to address. I have extensive knowledge in applied arts and languages.

[*heritage professional*] How much does your institution spend or intends to spend on getting new enrichments? At the moment there are no funds planned for the new enrichments.

[*wikimedian*] How can we ensure that enrichment efforts are not wasted or duplicated?: The answer to this has two parts. The first is to ensure enrichments are not burdened by restrictive licences which prevent their re-use. Free licences such as CC0 for data and CC-BY-SA/CC-BY/CC0 for media or longer texts should be the recommended default for new enrichments. The second is the discoverability of enrichments. Here the institution hosting the original content has an important role to either include and expose enrichments that have been done, or to provide a mechanism by which other platforms can clearly mark that they have enrichments to the particular content. Persistent identifiers, linked data etc. can play an important role here but two-way connections are necessary for someone starting at e.g. a Cultural Heritage Institution collections website to discover that the current object they are looking at has enriched data in Europeana, Wikimedia Commons or on some other more niched platform. Use of open and accessible APIs to both contribute and read data should be the recommended standard for enrichment platforms.

[*heritage professional*] The scale of enrichment i.e. what % of your online collection has been enriched, the treatment of legacy data vs new data, use controlled vocabularies and linked data enrichments. AI enhancement and the ethical considerations of modification (which may make data usable) vs. a pure record for historical posterity. Tracking changes and issues surrounding authenticity.

[*member of an aggregation service and heritage professional*] What should you do when an enrichment goes wrong?