



This project is funded
by the European Union



Brussels, 29.06.2023

Energy Efficiency Compliant Product 2018 - EEPLIANT3

GA N° 832558

Procurement of service for the development of a WebCrawler prototype/MVP

under Work Package 2 on the Development of IT tools

Date of publication of Q&A version 2: **29 JUNE 2023**

Ref.	Submission Date	Question	Answer
1	19/6/2023	Since exporting into excel and integration to Power BI was mentioned, can we assume that you already have Microsoft 365 licenses and not include that in our offer?	That is correct. Financial offers should be prepared according to the requirements of the two scenarios set out in Section 9 of the tender specifications. Costs relating to potential cloud hosting services, as well as costs for the purchase of Microsoft 365 licenses should be excluded.
2	19/6/2023	In your webpage it says that answering in mail is sufficient but in procurement material also hardcopy is requested. Which information is correct?	Tenders should be submitted via email AND by post according to the deadlines laid out in Section 12 of the tender specification document.
3	22/6/2023	Is it necessary for the MVP to be able to read QR codes?	This feature is one of the minimum functional requirements. Consequently, it is expected that the selected developer will be able to implement it in this prototype/MVP version. In any event, the final scope and specifications of the solution will be determined together with the selected supplier.
4	22/6/2023	Is it necessary for the MVP to be able to identify if the energy labels are valid ones? Not just if an energy label, the small and/or the big, is present?	It is expected that the MVP would be at least able to recognise the presence or absence of the EU Energy Label, read the QR code, and take a snapshot of label for export. However, it is up to the tenderers to propose additional features in their proposals that could be implemented by November 2023. As mentioned, the final scope and specifications of the solution will be determined together with the selected supplier.
5	22/6/2023	Is it safe to assume that the calculations engine only needs to work if the necessary data is present in the product page?	That is correct.

		While reviewing the security and compliance requirements we've met the following (Tender specifications, p 3.1., page 7): <u><i>Requested minimum requirements:</i></u> <ul style="list-style-type: none"> - <i>Compliance with GDPR (General Data Protection Regulation)</i> - <i>Servers used are hosted in the EU</i> - <i>ETC</i> Could you please comment on the last point? ETC - what does it stand for, please?	
6	22/6/2023	<ul style="list-style-type: none"> - <i>Compliance with GDPR (General Data Protection Regulation)</i> - <i>Servers used are hosted in the EU</i> - <i>ETC</i> Could you please comment on the last point? ETC - what does it stand for, please?	This should read “et cetera” (etc.)
7	23/6/2023	Do you mean to perform a crawl in the following?: <i>“5. Ability to launch a search on demand and also automatically/on schedule”</i>	That is correct. The meaning is to run a crawl.
8	23/6/2023	There seems to be a discrepancy between Requirement 14 and User Story #20. While requirement 14 emphasizes the detection of a URL link to the product information sheet and downloading it as an image, User Story #20 suggests the need for collecting, calculating benchmarks, and processing consistency checks on the data from these sheets. These two tasks - detecting a link and extracting & processing data from a PDF file - are significantly different. Could you clarify which one is the actual requirement?	Section 3.1 - Overview of key functional and technical specifications and system requirements and Section 3.2 - Description of the minimum desired features as user stories are complementary and should be read together. Accordingly, it is expected that the selected developer will be able to implement both features as described in Requirement 14 and User Story #20.
9	23/6/2023	You have mentioned that you are seeking a comprehensive analysis of cloud-based and on-site (localized) solutions. Is this analysis expected as part of the project scope, or do you prefer it to be addressed in the tender itself?	According to Criterion 2 under Section 8 - Evaluation criteria of the tender specifications, tenderers should propose a solution for the IT infrastructure and database to be used (cloud hosting service, local infrastructure, or combination). This proposal should be justified based on an analysis/assessment of the associated costs, benefits and risks, also considering future development and scalability. There are no specific requirements on how tenderers should present and substantiate their proposal in this respect.

10	23/6/2023	<p>Concerning the feature parity between local (Desktop application) and web application (server-based), should the desktop version have all the same features, including the ability to perform web crawls? Or are you envisioning a subset of the functionality for the desktop application, such as working on existing data?</p>	<p>The features to be implemented should be the same in both cases.</p>
11	23/6/2023	<p>With regard to the ownership of intellectual property rights, you state:</p> <p><i>"PROSAFE and the EEPLIANT3 beneficiaries shall retain full and exclusive joint rights, title or interest in and to the target application including but not limited to codes, patents, inventions, or other intellectual property developed or created in the course of this project, together with the right to market, sell, and distribute the target application."</i></p> <p>Would it be problematic if we use ideas that are already implemented in existing tools, like the internal web crawler tool? Would you claim exclusive rights to these existing ideas, potentially blocking future use of them in internal tools at my company?</p>	<p>The aim is to develop our own ‘in-house’ WebCrawler. As indicated in the tender specifications, PROSAFE shall retain exclusive intellectual property rights on the end product as a whole, and on the newly created material and results produced under this service contract which will be incorporated in the end product including but not limited to codes, patents, inventions.</p> <p>If developers intend to use pre-existing material, such as scripts, or parts of existing technological solutions, or other “existing ideas implemented in existing own tools”, in creating the target application (=WebCrawler MVP), they must indicate this in their proposal and list them.</p> <p>Furthermore, developers must accept from the outset that the use of any pre-existing material in creating the target application will not prohibit or limit by any means PROSAFE from using or exploiting the end-product in any form and for any purpose, including future development. Tenderers must provide in their offer a clear statement confirming that they accept these terms and conditions.</p> <p>PROSAFE will not claim exclusive rights on such pre-existing material incorporated in the end product. This means that PROSAFE will not prevent developers from using this pre-existing material in any other business including use in proprietary tools owned by the developers.</p> <p>Furthermore, the WebCrawler application to be developed should have no dependency on third-party closed-source tools/technological solutions/software that could prevent or limit its use, exploitation, or future development. This excludes cloud infrastructure/cloud-hosting service.</p>
12	26/6/2023	<p>Is it enough to use the EPREL website following the QR code link to extract information? Is processing PDF version of the product information sheet required?</p>	<p>Please see above Question 8.</p>

13	26/6/2023	Is it required to only scrape information from 1 product category for one inspection, e.g., televisions? OR is it required to scrape multiple product categories for one inspection, e.g., televisions and washers?	As specified in Section 3.1 of the tender specifications, the product type in scope for this iteration is televisions.
14	26/6/2023	Is there a preference for using external API or building in-house code for advanced functions such as scraping engine & AI processing?	Please see above Question 11.
15	26/6/2023	Is it allowed to use cloud-based APIs? e.g., AI?	Please see above Question 11. Use of existing AI to be connected with each other is not our target.
16	26/6/2023	On User Story #3: What kind of product information has to be verified? (what is essential)	For the purpose of this procurement, the target essential information/data for extraction is presented in Section 3.1 of the tender specifications.
17	26/6/2023	On User Story #6: What kind of statistical data should the system provide?	In User Story #6, statistical analysis for the purpose of market surveillance planning and priority-setting involves descriptive statistics, data manipulation and aggregations of the collected/extracted target information, as listed in Section 3.1 of the tender specifications. The exact statistics to be implemented in this iteration will be determined together with selected supplier.
18	26/6/2023	On User Story #13: What kind of data should be exported? Ideally, could you provide a template of how the file should look with example data?	Please see above Question 16. The format of the extracted data will be specified at a later stage together with the selected supplier.
19	26/6/2023	On User Story #17: What is the purpose of exporting & importing? The data can be kept there for the lifetime of the system, also backup actions can be automated.	In case of a cloud-based database and with the assumption of a subscription model, if we decide to suspend our subscription (e.g., for 6 months), we would like to be able to back up the database and have the possibility to restore it once we decide to reactivate our subscription. Tenderers should consider this aspect when addressing Criterion 2 of the Evaluation Criteria (Section 8 in tender specifications) regarding the proposed solution for the IT infrastructure and database to be used.

20	26/6/2023	On User Story #20: What are the details of the consistency checks and benchmarks that are to be run here?	Consistency checks refer to checks between information displayed directly on a product page and information contained in the product information sheet (and in the EPREL database, see User Story #5) in order to verify their consistency—i.e., that information is the same across sources.
21	28/6/2023	Could you please provide list of shops which should be covered in MVP scope?	The list of shops (root URLs) to be targeted will be specified by the project after the selection of the preferred supplier.
22	28/6/2023	What is the desired frequency of data updates or refreshes for the scraped information?	As indicated in the tender specifications, speed and frequency are non-essential requirements for this first iteration. However, the system must be as optimised as possible taking account of the number of pages to be crawled, the volume of the data to be scraped, and the benchmarks of other similar solutions.
23	28/6/2023	Are there any restrictions on the use of third-party libraries or open-source components in the development of the web crawler application?	Please see above Question 11 and Question 14.
24	28/6/2023	Do you have any preferences regarding the reporting or notification mechanisms for monitoring the status and progress of the web crawler?	There are no preferences at this stage. Tenderers may propose best-suited solution(s) considering client needs, and fitness to all functional and non-functional requirements, as set out in the tender specifications. Specifications in this respect will be defined together with the selected supplier.
25	28/6/2023	Do you have any preferences regarding the logging and auditing of scraping activities for compliance or troubleshooting purposes?	There are no preferences at this stage. Tenderers may propose best-suited solution(s) considering the client needs, user experience, and fitness to all functional and non-functional requirements, as set out in the tender specifications. Specifications in this respect will be defined together with the selected supplier.
26	28/6/2023	Are there any specific data integration or data synchronization requirements with other systems or databases?	Apart from the interoperability requirement with the open space of the EPREL database and MS Power BI, there are no other specific data integration or data synchronization requirements at this stage.
27	28/6/2023	Do you have any expectations regarding configurability of the web crawler application?	Configuration preferences will be discussed together with the selected supplier following the award of the contract.
28	28/6/2023	Do you have any expectations regarding the handling of changes in the website structure or layout that may affect the scraping process?	We are aware that changes in the website structure or layout can affect the scraping process. Tenderers may propose potential solutions/mitigations that could be implemented to address this issue.

29	28/6/2023	Should we highlight differences between retailer's product description and data from EPREL database?	Please see above Question 20.
30	28/6/2023	Does the project have a specific budget and/or monthly investment? If yes, please specify.	We do have a specific budget for developing this MVP, however, we cannot disclose any related information. Our preparatory research indicates that the available budget is sufficient to cover the development of this prototype/MVP solution, considering its size and scope, as described in the tender specifications.
31	28/6/2023	What are the most important factors to qualify?	The evaluation criteria and weighting are specified in Section 8 and Section 11 of the tender specifications, alongside the assessment procedure.
32	28/6/2023	What are the qualifications of subject matter experts? Which contribution can we expect from them?	The selected developer will work closely with a seasoned IT consultant in the role of a Product Owner, the PROSAFE team, and a small group of subject matter experts. These experts are senior market surveillance officials with extensive experience in market controls in the energy labelling and ecodesign fields.
33	28/6/2023	Which ticket & story-tracking tool is used by product owner, and development teams? Will you be providing these tools or could we you own?	It is expected that developers will use their own collaboration/PM tools and provide access to 1-3 members of the project team, including the product owner. We ask tenderers to indicate these tools in their offer.
34	28/6/2023	Should we make separate screenshots on Mobile devices?	There are no specific requirements for mobile devices.
35	28/6/2023	Who is responsible for infrastructure procurement and setup on-premises or cloud?	Procurement for any required IT infrastructure will be managed by the project.
36	28/6/2023	Do we need any automation to check ED/EL regulations? Or will this be done manually by the inspector?	Please see User Stories #15 and #20, and also Question 20 above.
37	28/6/2023	Is there a fixed list of targeted websites for crawling at this moment?	Please see Question 21 above.

38	28/6/2023	Is there a proposed way to integrate with the EPREL database?	According to User Story #5, the web crawler should be able to read the QR code on the energy label on each product page, and then download the label and all the available product data from the open/public space of the EPREL database.
39	28/6/2023	As per user story 8 “As an inspector, I want the web crawler to have a desktop user interface”. What does ‘desktop interface’ mean in this context?	We mean a frontend application.
40	28/6/2023	As per user story 9 “As an inspector, I want to be able to program the web crawler, so that it can automatically crawl websites periodically.” What does ‘program’ mean? Is it about changing crawler scheduling parameters? Or is it about changing crawling parameters? Or is it about adding new web sites?	User Story #9 describes our desire to have scheduling options which can enable users to scrape data from the targeted websites automatically, whenever they choose, e.g., “the first Monday of every second month, at 12:00 am” (indicative example).
41	28/6/2023	As per requirement “GUI-based operation ready to scale up across variable cross-platform frameworks (desktop, mobile)” Please provide a list of OS/browsers/devices to support	The developed solution should be able to run across all main/most popular OS, browsers, and types of devices: <ul style="list-style-type: none"> - OS: Windows 10/11, macOS 10.x/11.x/12.x/ 13.x, Android 8+, iOS 11+ - Browsers: Google Chrome, Microsoft Edge, Safari - Devices: PCs/Macs, tablets, and smartphones
42	28/6/2023	As per requirement “Multi-user operation through individual password protected user accounts; Implementation of two roles: administrator and user” Is it necessary to support external identity providers’ sign in (google, Microsoft, other)?	There are no specific requirements in this regard.
43	28/6/2023	When would PROSAFE consider the MVP of the EEPLIANT3 project to be a success?	The MVP would be a success when the data crawled would be found useful by the market surveillance authorities, the officers would find it easy to use and, overall, the tool developed would improve the efficiency and effectiveness of their work related to online inspections. The tool would also be a success if delivered on time and it could be easily transferred to the next project so that it can be further developed. Some indicative success factors are indicated in Section 3.1 (p. 80) of the tender specifications. KPIs to measure success will be discussed and

			agreed upon together with the selected supplier. Tenderers are free to suggest such KPIs in their offer.
44	28/6/2023	What is the long-term vision for the EEPLIANT3 project?	With this prototype/MVP, our ambition is to start building a sustainable solution that serves our needs, as these are described in the tender specifications, and which will be built in such a way that is scalable and adaptable, as the intention is to upgrade/develop it further in the near future by covering more EU markets, more retailer websites and product types, and by adding more functionalities, deeper automation and machine learning.
45	28/6/2023	What are the key challenges and pain points faced by market surveillance authorities (MSAs) in enforcing EU product energy efficiency legislation?	Please refer to Section 1 of the tender specifications.
46	28/6/2023	Will the root URLs of the to-be-crawled websites be provided?	Root URLs will be provided by the project after the selection of the preferred supplier. Please also see Question 21 above.
47	28/6/2023	Should the crawler run through the websites all together or would one by one be sufficient for the MVP?	Developers are free to propose the best possible implementation in this respect, considering constraints. Developers will be provided with pre-selected websites to start testing the application during its development; however, the concept shall be based on web crawling (not scraping!) and must be suitable to cover multiple EU markets in the future.
48	28/6/2023	What does EEPLIANT3 mean with user story #10? Does the user type 'programmer/software developer' refer to the supplier's developer or an additional external developer?	As mentioned, our intention is to continue the development of this WebCrawler MVP in the future, under a potential follow-up action to EEPLIANT3. Thus, the term refers to the developer or developers who may be contracted in the context of a potential future procurement for this purpose.
49	28/6/2023	Are suppliers free to suggest KPIs to measure the success of the crawler for the MVP?	Please see above Question 43.

DISCLAIMER

This document is part of the EEPLIANT3 concerted action that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 832558. The content of document represents the views of the author, and it is his sole responsibility; it can in no way be taken to reflect the views of the European Climate, Infrastructure and Environment Executive Agency (CINEA), the European Commission or any other body of the European Union, who are not responsible for any use that may be made of the information it contains.