# Data product survey (human-led discovery)

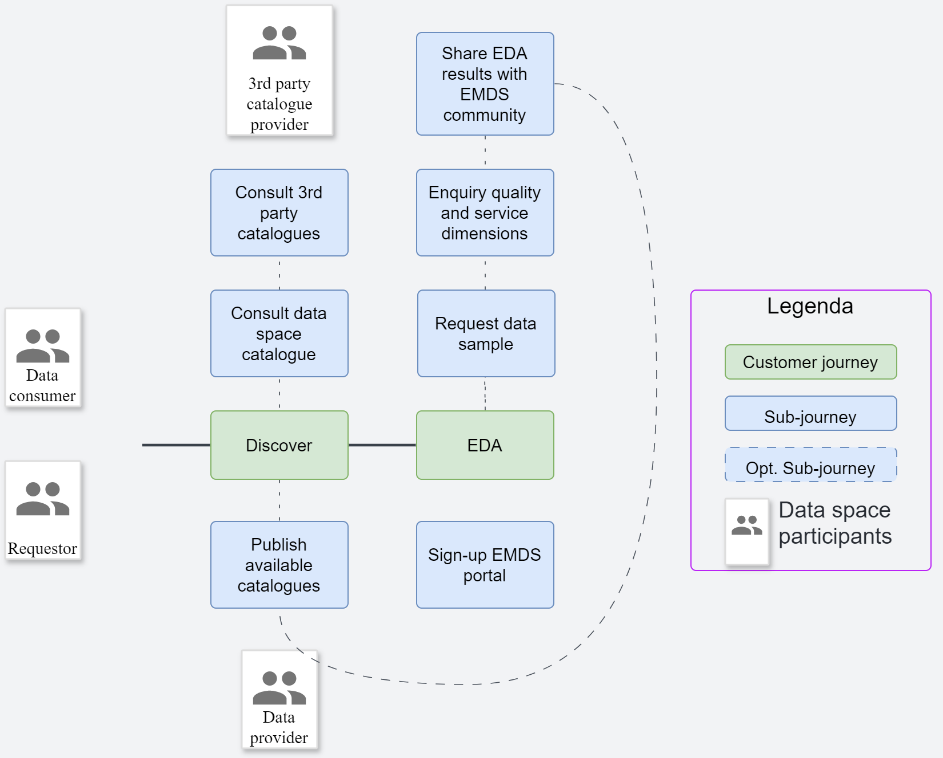
## What is it.

A potential data consumer (e.g., a requestor, or participant) performs an explorative survey of available data in the mobility sector. She peruses EMDS and 3rd party catalogues (like <https://data.europa.eu>) where data providers published their data products.

The data consumer can request more information on the data product exposed on the EMDS, she will, however, need an EMDS portal account to establish a communication channel with the data product provider. Without being an active participant of the EMDS data space, the data consumer might obtain a sample of the data she interested in, to perform further evaluation. The data consumer can share impressions and experiences on the EMDS portal with the EMDS mobility interest group.

## Detailed overview

### Graphical representation



### Actors

|  |  |  |
| --- | --- | --- |
| Actor | Description | Notes |
| Requestor | The legal entity that wants to join the EMDS and qualify as Participant | We assume that the entity involved in the survey will join EMDS if the result of the survey is positive. |
| Data consumer | A participant engaged in data sharing in the EMDS requesting and using data provided by a Data Provider. | Participants can assume different roles (e.g., consumer, producer, intermediary, etc.) |
| Data provider | Participant exposing Data Sources via a Connector; a Data Provider may be an enterprise or other organisation, a data marketplace, an individual, or a “smart thing”. | The DS Authority might release credentials or delegate the release to a third party. |
| 3rd-party catalogue provider | External providers of mobility data catalogues that wish to integrate EMDS data products in their catalogue listings. | EMDS does not have control over these 3rd-party catalogues, but EMDS participants can choose to publish their data products there through the EMDS catalogue |

### Customer journey: Discover

|  |  |  |  |
| --- | --- | --- | --- |
| Sub-customer journeys | Description | Examples of ISO 25010 Qualities | Review/tests |
| Consult data space catalogue | A requestor or data consumer browses and search the data space catalogue (augmented with qualitative and human-readable information) through the EMDS search tool. | 1. Functional suitability: The catalogue system provides human or interactive interfaces to perform complex and free-text searches across data products. 2. Functional suitability: The content of the catalogue can be published to Internet search engines. 3. Functional suitability: The catalogue is accessible to standard data space connectors (programmatic interface). 4. Functional suitability: The catalogue provides usage statistics of the different data products. 5. Functional suitability: The catalogue maintains updated quality metrics about the different data products. 6. Security: The catalogue provides AAA, if needed, to consult the catalogue? 7. Performance: The catalogue provides results in a timely manner, so that either interactive or programmatic interfaces work as generally expected. | **1: Assessment: If an Online U/X is natively available, evaluate individual search features. If the Data space catalogue exposes an API, assess the technical debt to integrate it with a data search tool that is representative for EU projects, criteria: Open Source, hosted. Or EU-driven project.**  2: Assessment and test: If the data space catalogue can be indexed by a web spider, test a mock indexing on major search engines. If not, assess the technical debt to implement. Rank higher if conditional indexing is available.  3: These tests are already performed in data product publication.  4: Assessment: the feature is present, assess its completeness. The feature is not present: assess the effort to capture the telemetry of data product usage and integrate it in a catalogue. If the stack provides observability, ranks higher.  **5: Assessment: either the data product specification provides the necessary metadata to report quality, or the catalogue must be extended with an “-AP” profile. Ranks higher in the first case.**  **6: Test: if the catalogue is part of a data space, test that authentication works, and the access is logged. Ranks higher if the audit is part of an observability framework.**  7: Synthetic timeliness tests. |
| Consult 3rd party catalogues | An entity, involved in a mobility project, browses and searches a 3rd-party data catalogue where EMDS data products are listed, augmented with qualitative and human-readable information. | 1. Functional suitability: 3rd-party catalogues contain an extract or copy of the EMDS catalogue. 2. Functional suitability: The 3rd party catalogue provides an indication that the entry is up to date, and the degree of completeness of the information that has been imported from the EMDS catalogue/producer. 3. Functional suitability: The 3rd party catalogue provides information of how to join the EMDS data space, if the data product is restricted to EMDS participants. 4. Interoperability: See criteria 6 on the “Request Data Sample” Sub-CJ to provide a data sample on 3rd party catalogues. | **1: Assess the effort required to integrate an external catalogue (choice to be done). Ranks higher if the publishing is part of the data space catalogue publishing protocol**.  2: Assessment: evaluate such feature if available (most probably not). Otherwise, evaluate or the cost to implement it in a way that doesn’t disclose the content of the EMDS catalogue to a 3rd party if this is not desired.  3: Assessment: prove that the information that the data space catalogue can share with a 3rd party includes non structured information like: “How to subscribe to EMDS”.  4: Already provided by another CJ. |
| Publish on available catalogues | The data product provider /owner updates multiple catalogues so that the discoverability of EMDS data products is maximised. | This CJ has been skipped at present.  To be clear: this CJ is about the requestor that, after assessing/discovering a data product, updates a 3rd party catalogue with the relevant information. It is foreseen as a from of crowdsourced data documentation, delegated to data consumers. |  |

### Customer journey: EDA (Explorative Data Analysis)

|  |  |  |  |
| --- | --- | --- | --- |
| Sub-customer journeys | Description | Examples of ISO 25010 Qualities | Review/tests |
| Request data sample | The data provider can share part or a synthetic data set for test, integration, and quality assurance purposes. The data sample does not require EMDS participation. | 1. Functional suitability: The catalogue can link to a data sample of the data product that has been provided by the data producer. 2. Functional suitability: If a data sample is not provided, the system (e.g., catalogue or connector) implements a function to send a “Data sample request” to the data producer. 3. Security: The system allows the data provider to retain a record of access to a data sample that has been specifically requested. 4. Security: The system allows the data provider to define usage policies / restrictions on the data sample. 5. Interoperability: The above characteristics are available to 3rd-party data catalogues. The system allows a data sample to be accessed either through an EMDS connector (without the need to undergo a full subscription to the data space), or to upload the data sample to a 3rd party staging zone. | **1: Assess and test: for instance, test that DCAT distribution property allows the publication of a data product for test data. Test that the information is published as part or as a separated data product that is referenced by its production counterpart.**  **2: Assess how this request can be performed in practice. For instance, via a data sharing agreement request on the data product, but with parameters that allow redirection to a test data data product, or data plane.**  3: As test 4 of:“Consult data space catalogue”  4: As tests defined in data product publication.  5: Assessment: this feature is likely non-existent. The assessment should provide a two-legged evaluation of: 1: which is the best fitting solution; 2: the solution implementation cost, with an eye on data security and non-disclosure. |
| Enquiry service and data quality | The requestor can demand the data provider more information or observability data documenting the quality of service and quality of data of the latter’s data product. | 1. Functional suitability: The system allows the requestor to request quality information from the data provider (assessment) if the information is not available in the EMDS catalogue (or 3rd party catalogue). The data provider can refuse or oblige with the request. 2. Security: If required by data space governance, the system authenticates a data quality assessment request (via EMDS credentials, could include attestations). 3. Interoperability: The system encodes and transmits the data quality assessment in a standard model for data quality representation (e.g., Open Data Product Specification: DQ). 4. Interoperability: Upon a data quality assessment request, the system allows the data provider to publish the report directly on the EMDS catalogue (and update the data product record). The requestor must be notified or informed (at data quality request) where and when the information will be available. | **1: Assessment: a data quality request should be implemented as a data sharing request. Assess if this is possible without custom implementations, and what is the feedback channel for the result. (Similar test to “Request data sample (2) sub-journey.)**  **2: Test: if the above is true, test a policy that verifies a mock attestation (e.g., “based in EU”) of the requestor. The result ranks higher if attestations are fully supported.**  3: Choose a data quality report format, test that the data quality report can be validated against the report format definition.  4: Likely not implemented at all. Skip test but keep this in mind. |
| Share EDA results /experience with community | The information pertaining the discovery and its results can be sent back to the community. A data provider might use the information to update the data product entries in internal and external catalogues. | 1. Functional suitability: The system allows a requestor to share the results of their EDA on the EMDS catalogue or to submit it to the data producer for further publication. 2. Integrity: The data provider has the sole right to accept the information of the requestor’s EDA or data quality assessment. 3. Functional suitability: The system allows EDA results to be automatically added to 3rd-party catalogues in sync with EMDS catalogue, if the data product policy allows that. 4. Accountability: The requestor provides credentials / digital certificates prior to sharing the results of their EDA. 5. Non Repudiation: The system informs the requestor that the result of her EDA have been submitted for consideration. | 1,2,3,4,5: These tests are skipped until this part of the customer journey is validated by the EMDS governance team. |