# Data Product publication

## What is it.

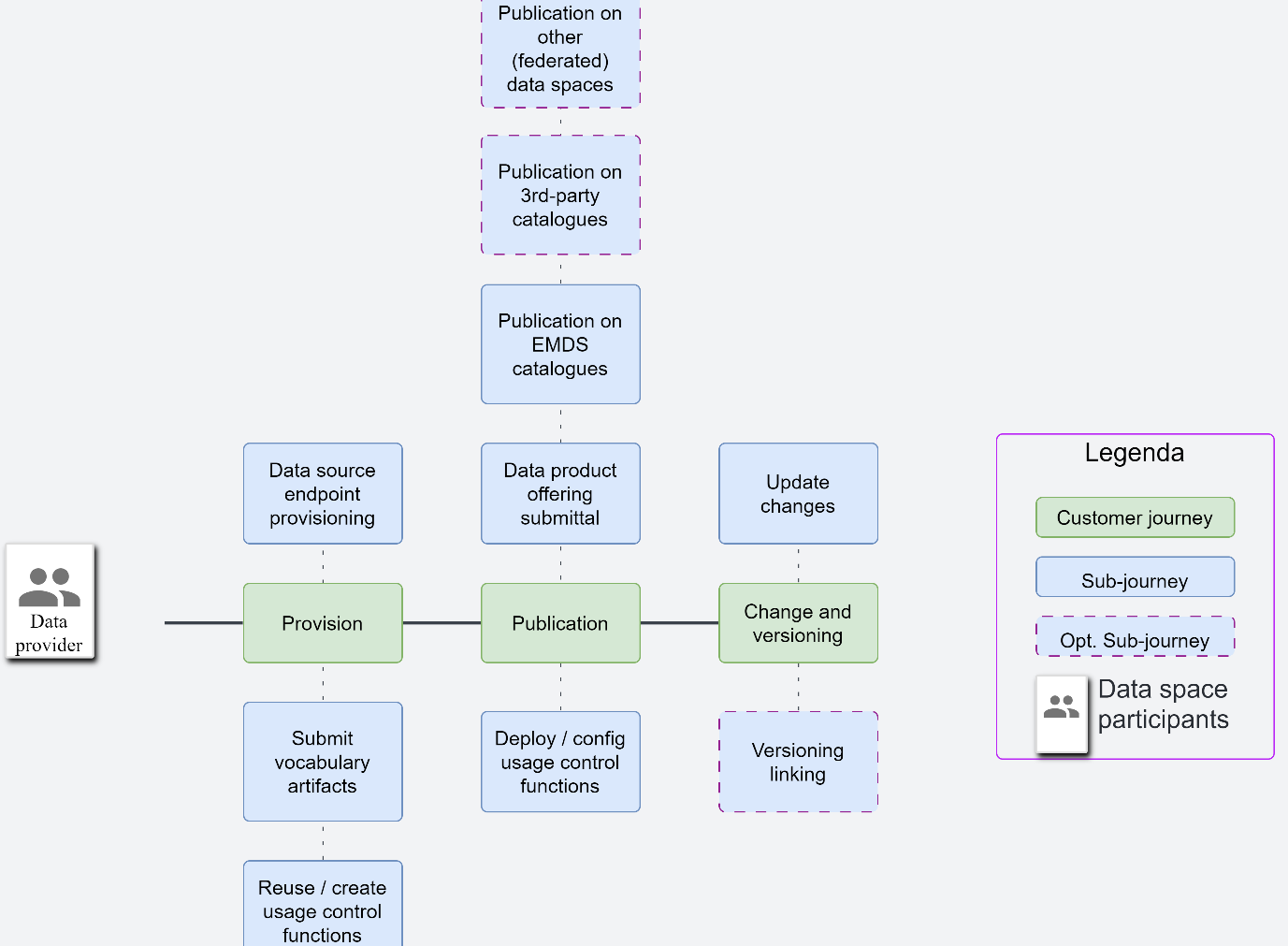
A data provider participant wants to publish a consumption-ready data asset in the form of a data product offering on the EMDS data space.

She starts by provisioning the technical infrastructure and artifacts (i.e., the data source endpoint, metadata and linked vocabularies, and the definition of usage controls that she can find on an EMDS repository (reuse) or build herself.  
She then proceeds to publish the data source as a data product on the EMDS data space, the process is not limited to a mere asset publication on the participant’s (or data intermediary) connector, it also involves providing information to internal and external search tools, and activating the asset’s sharing agreement (usage control policies might require a custom control functions with additional configurations).

During the lifecycle of the data product offering, the data provider might need to update its assets or the way it is consumed, as well as supporting previous versions in a machine-automatable way (preferably using linked information)

## Detailed overview

### Graphical representation



### Actors

|  |  |  |
| --- | --- | --- |
| Actor | Description | Notes |
| Participant | Stakeholder in the EMDS data space; every Participant is assigned a unique identity by the EMDS authority. | 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”. | As defined in D2.1 |

### Customer journey: Provision

|  |  |  |  |
| --- | --- | --- | --- |
| Sub-customer journeys | Description | Examples of ISO 25010 Qualities | Review / tests |
| Data source endpoint provisioning | Data provider prepares and secures a data source for data space consumption. | 1. Security: Can the data provider integrate the source using choosing the right API authentication method? 2. Security: Can the endpoint of the data source be masked/obfuscated when the data is consumed? 3. Functional Suitability: Can the provider integrate different data source methods (APIs, Data bases, file systems, etc)? 4. Fault Tolerance: Can the data provider adjust the need for a recovery system that ensures high service continuity and performance? 5. Performance: Can the data provider provide a limit on the volume and speed of data? 6. Security: Does the data provider need to certify the ownership of the data source platform? 7. Security: Is the data provider required to indicate the owner / sovereign of the data? | **1: Test that the component/connector complies with NIST 800-53 AAA controls. More in detail: Test coverage: Pass test on secure authentication procedures (the result ranks higher as more authentication procedures are supported).**  **Assess: the account that is needed to access the data source is managed and stored in a secure way.**  2: Assess that a connector provides a proxy http(s) data plane, that masks the data source  **3: Assess the availability of multiple data planes that support multiple protocols. Refer to D2.4 for an overview of the most used protocols. The higher the coverage, the higher the ranking.**  4: Setup a test where an HTTP(s) data source provides Active/Passive endpoints in a load balanced fashion. Test the data source stability by failing it over during critical sessions like: endpoint provisioning, data sharing execution.  5: Assess if QoS and rate limiters are available options on the connector’s data plane.  6: This cannot be tested without a data provenance framework.  7: Assess that the Data Product publication contains information on the ownership of the data. |
| Submit vocabulary artifacts | The data source must be supported by metadata and standardized vocabularies. Such artifacts must be validated by an EMDS interoperability facility. | 1. • Flexibility: Can the data provider submit a new vocabulary? 2. Flexibility: Can the data provider choose from different standardized vocabulary that meet the requirements of the data source? 3. Functional Sustainability: Can the data provider submit custom metadata fields and link their semantic definition? 4. Flexibility: Can the data provider access standardize vocabularies to then use them in the data source before provisioning the data? 5. Integrity: Does the EMDS interoperability facility validate the vocabulary used at the data source? 6. Functional Suitability: Does the system provide feedback or error messages if a vocabulary submission fails validation? 7. Functional Suitability: Can the data provider view the validation status of used vocabularies at the data source? 8. Functional Suitability: Can the data provider request to update/modify a vocabulary standard? 9. Interoperability: Does the metadata associated to the data product follows a given standard or model from the data space? | **2-4: Test that the vocabulary hub’s assets are available to every data space participant.**  **Rank higher if the reference to a vocabulary asset is integrated in the data sharing process.**  **1-6: Test that a data space participant can publish new vocabularies (test vocabulary), and the vocabulary hub returns convenient feedback of the operation.**  8: Assess if and how a change in a vocabulary asset is executed. Rank higher if integrity controls are in place.  **3-9: Assess that the vocabulary hub covers semantic or domain standards for the metadata. The more the coverage of standards, the higher the ranking.**  5-7: Assess that the vocabulary hub validates a new entry or a modification of an existing vocabulary. Test the validation process by submitting a test vocabulary, and assess if the service provides a feedback on the result of the validation. |
| Reuse or create usage control policies / functions | Data providers can tap at the EMDS repository of usage control policies and constraint functions. Or they could write new ones and test them in a test sandbox. | 1. Functional Suitability: The data producer uses an API or to deploy and config.ure usage control policies. 2. Functional Suitability: The data producer uses a GUI or to deploy and configure usage control policies. 3. Flexibility: Can data providers integrate new usage control policies into existing data products? 4. Flexibility: Can data providers submit new usage policies to the EMDS repository? 5. Functional Suitability: Can the data provider apply multiple usage control policies simultaneously to a single data source? 6. Maintainability: Do any new development of usage control policies need to follow a set of requirements or standards from the dataspace governance framework? | **1: Assess how Usage Control Policies are deployed. Rank the result by API coverage and ease of use (i.e., avoiding multiple calls with parameter passing).**  **2: Assess how Usage Control Policies are deployed. Rank the result by GUI coverage and ease of use (i.e., auto-completion, validation, interface to repository if available).**  3: At the moment no data space allows that.  4: If the EMDS usage control policy repository is available, verify that a new policy can be pushed on it. Rank the result by scoring the publish process, and if security and integrity controls are available (e.g., a review process).  5: Depending on the formalism used to express UCPs, either test a policy that covers two use cases (OR, and AND), or concatenate two test use policies. For instance: consumer is within the EU AND is a company.  6: Part of test 2 that covers integrity and security controls. |
|  |  |  |  |

### Customer journey: Publication

|  |  |  |  |
| --- | --- | --- | --- |
| Sub-customer journeys | Description | Examples of ISO 25010 Qualities | Review/Tests |
| Data product offering submittal | The data producer securely publishes a data product offering via an administrative interface (e.g., API or U/X). The connector where the data product is published updates data space registries and catalogue with the data product offering entry. | 1. Security: The data producer administrative interface is protected with authentication. 2. Security: The administrative interface is secure and preserves confidentiality of session. 3. Functional Suitability: The data producer has access to overviews of data products offerings, sharing agreements, available data planes. 4. Functional Suitability: Clear feedback is provided on validation errors and warnings during the publishing process. 5. Interaction Capability: The system provides tooltips/help to ensure the data provider correctly understands and fills the associated metadata fields of the data product offering. 6. Reliability: The publishing process is resilient to interruptions, and it can recover from network, system or timeout failures. 7. Performance: The connector’s latency to update the data space registries and catalogue after a data product offering is published. 8. Functional Suitability: Is it possible for a data producer to apply a monetization scheme to the data product offering? 9. Functional Suitability: DIS1-3, DOF1,2 | **1: Test that the administrative interface of the connector complies with ISO 27001 Annex 9.x (Access Control procedures) or, alternatively, NIST 800-53 AAA controls. More in detail: Test: secure authentication procedures (Rank by security of authentication).**  **Assess: administrative accounts are managed and stored in a secure way.**  **2: Test that the administrative interface of the connector complies with ISO 27001 Annex 9.x (Access Control procedures) or, alternatively, NIST 800-53 AAA. More in detail:**  **Assess: administrative accounts are segregated (multi-tenancy of connectors’ operators).**  3: Assess if the administrative interface of the connector provides overviews of data products, data sharing agreements, available data planes and the status of these assets.  4: Simulate submission inconsistencies to test error-handling. Rank by error clarity, notification channel, and timeliness of error report.  5: If U/X is available, refer to ISO 9241 (<https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en> ) to extract a test suite that evaluate the friendliness of the interface.  6: Refer to the proposal above.  7: Measure the average time of asset publication propagation to the data space catalogue and the vocabulary hub.  8: Assess that the system supports metered consumption, as: usage count, rating, billing… |
| Deploy / config usage control functions (usage control functions are not usage control policies!). | The data producer links usage control policies to the data product offering. The data producer is also responsible to deploy and configure the constraint functions, or external lookup functions (for instance: a geo-IP service to verify that a consumer is located in the EU) that implement the policies. | 1. Functional Suitability: The administrative interface allows the producer to deploy and configure usage control policies. 2. Functional suitability: The administrative interface for usage control policies has a GUI. 3. Analysability: The system provides an overview of the enforcement and application of the usage control policies. 4. Functional Suitability: The data producer deploys custom usage enforcement functions. 5. Functional Suitability: Usage control policies can support enforcement by external or manual input. 6. Functional Suitability: The deployment of new usage control policies and enforcement functions can be tested before they are applied to a data product offering. 7. Accountability: Security measures are in place to ensure that only authorized data producers deploy and configure usage control policies and custom functions. 8. Performance: The system minimizes the impact of usage control policies to the performance of the data product offerings. 9. Integrity: The system guarantees that non-compliance with usage policies prevents access to the data product offering. 10. The policies language allows development of new operators to support new models of consumption and constraints. 11. Interoperability: The policy language is complete; it allows the translation of usage policies from another policy language.   Functional Suitability: UC1-3, UC8, UC11 | **1: Assess the completeness of the administrative interface (either API or GUI) so that it covers the most needed use cases for the deployment of usage policies: upload a new policy, (optional)bind a policy with a custom enforcement function, assign a policy to a sharing agreement, delete a policy, re-use an uploaded policy, persist uploaded policies.**  2: Assess if a GUI is available to define and deploy usage policies and bind them to custom enforcement functions.  3: Feature assessment: assess if the system provides an observability record of enforced policies. Rank higher if the information is available in an organised report or GUI.  **4: Expandability assessment: assess if the system provides an API or libraries to embed custom enforcement functions that can be invoked by usage policies.**  5: Feature assessment: assess that the usage control framework supports external enforcement.  6: Feature assessment: assess if the system provides a built-in facility to simulate policies (max score), or if a separate test environment (min score) is required.  7: Assess that only the data producer can associate a policy to a DP during its publication. If multitenancy is allowed, published assets should not be manageable cross-tenancy by default.  8: Load test: create an enforcement function that artificially allocates memory and CPU resources in a loop and monitor performance management of the connector.  9: Smoke test on policy enforcement: bugs, user-space safeguards, access rights override (conflicting rights), etc.  **10: Assess that the policy language is extensible and score the results by flexibility and availability of development and testing facilities for new operators.**  11: If the stack uses a policy formalism that is different from ODRL, provide an analysis or examples of 1:1 policy translation in ODRL. Refer to the set of policy patterns defined in IDSA’s Position Paper on Usage Control [https://internationaldataspaces.org/wp-content/uploads/dlm\_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3..pdf )](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3..pdf) |
| Publication on EMDS catalogues | The data producer published their data product offerings on data space EMDS catalogues, facilitating their discoverability by other data space participants and requestors. | 1. Functional suitability: The data producer can, at any time, make a data product visible in the data space catalogue(s). 2. Functional Suitability: Can the data producer track the performance and visibility of their data product offerings on the search and discovery tools? 3. Functional Suitability: can the data producer use a GUI to publish the data product offerings in the data spaces catalogue and discovery tools? 4. Functional Suitability: can the data producer remove a data product offering from the EMDS catalogue? 5. Functional Suitablity: DOF6-9, DIS2-8 | **1: Test the process of catalogue publication for a data product under the following conditions: a new data product is published in the catalogue; an existing data product is published on the catalogue; a new data product cannot be published on the catalogue; a data product is de-published.**  2: Assess that the data space catalogue provides an overview of searches and hits for data products. Ranks higher if an option exists to make these statistics public or accessible only by the owner of the data product.  3: Assess that a GUI is available.  4: Feature assessment: catalogue de-publishing or make the catalogue entry private. |
| Publication on 3rd-party catalogues | The data producer advertises a data product offering on external data portals, often including promotional and descriptive information that is not present in the data space catalogue. | 1. Interoperability: A data product offering can be published on discovery portals other than the data space own’s. 2. Functional completeness: the same publishing KPIs indicated for the data space catalogue (e.g., publication policies) are valid for external catalogues or search tools. 3. Interoperability: The data space catalogue can be included in an European directory of data portals, like Prepdspace4mobility (<https://mobilitydataspace-csa.eu> ) 4. Interoperability: The system provides a pre-populated list of search and discovery tools where to publish a data product offering. (Also for Federated data spaces) 5. Interoperability: The system syncs updates to the data product offering (data, metadata, usage policies) across discovery tools where it is published. 6. Interoperability: The native metadata model can be translated to a European Search Tool of choice. | **1: Assess if the stack provides an interface to 3rd party catalogues.**  **If not, what is the effort to build one? Rank in terms of technical debt (relative to the rest of selected stacks being analysed).**  **2: Assess that publishing to a 3rd party catalogue or search tool doesn’t add technical debt. Configuration, and policy setting are not technical debt. Metadata translation IS technical debt.**  **3: Assess that the data space catalogue is public, if that’s the case, assess what work would be needed to publish, and sync, its content with PrepDSpace4Mobility.**  4: Assess that the system provides an interface to choose among external search tools.  5: Assess and test that the system performs automatic updates to the chosen 3rd party catalogues. Ranks higher if automation can be excluded or configured.  6: Choose an European project that provides data catalogue and assess / test how to integrate native data space catalogue entries with the chosen tool. |
| Publication on federated data spaces | If the EMDS partners with other data spaces, data producers might want to publish their data product offerings on federated data space catalogues, accessible by participants of the other data spaces. | 1. Interoperability: The system allows to publish a data product offering on a data space other than the EMDS. 2. Interoperability: A federated data space can publish data product offerings on EMDS, 3. The publication across federated data spaces requires the translation of a data product offering. 4. EMDS participants need to obtain a separate identity or are automatically attested in a federated data space. 5. The publication across federated data spaces is performed by data intermediaries, which manage the data product offerings translation and cross-data space attestation. 6. Interoperability: Usage control policies can be translated among federated data spaces. 7. Interoperability: search and visibility are expanded transparently across federated data spaces. | 1, 2: Assessment: Give an existing data space, assess that this CAN be done or not. More qualitative assessment comes in the KPIs that follow.  3: Assessment and compatibility testing. As federated data spaces might use different standards and policies, document the level of translation that is necessary to publish data products on both directions. If the chosen stack provides tools to aid in this translation, it ranks higher.  4: Assessment. Data spaces might use different IAM frameworks. Assess if the stack provides the necessary technology to federate identities, claims, and attestations. If not, assess what cost a data producer or consumer must sustain to maintain different identities (e.g.: ranks lower if one is required to recreate identity and re-publish a data product in the federated data space. Ranks higher the more transparent the process is.)  5: Assessment. If there is no way to federate data space other than having intermediaries at both data space’s boundaries that re-publish data products on behalf of data producers and provide access to data products on behalf of consumers. Asses the cost and technical debt required to implement such strategy.  6: Test. If the federation requires policy translation, test the translation of a representative usage policy; evaluate coverage and feasibility.  7: Assessment: does my product in a federated data space achieve equal visibility as native ones? |

### Customer journey: Change and versioning (**No testing for now**).

|  |  |  |  |
| --- | --- | --- | --- |
| Sub-customer journeys | Description | Examples of ISO 25010 Qualities | Review / Test |
| Update changes | The producer manages the lifecycle of data product offerings, implementing a support policy for existing and past versions thereof. | 1. Functional Suitability: Can the data provider update a data product offering to reflect the latest changes from the data source and automatically update any ongoing agreements that require an update data product? 2. Functional Suitability: Can the data producer notify consumers on updates to the product data offering? 3. Functional Suitability: can the data producer manage the data product lifecycle from a GUI? 4. Functional Suitability: Can the data producer update usage control policies based on the updates of the data product lifecycle on existing agreements that allow policy updates? |  |
| Versioning and linking | The producer provides multiple versions of the data product offering, in case contractual agreements demand so. | 1. Functional Suitability: Does the system support versioning of data product offerings to manage existing and past versions? 2. Functional Suitability: is it possible to manage the different versions of the data product offering from a GUI? 3. Functional Suitability: Is the version of the data product offering detailed on the associated metadata? |  |