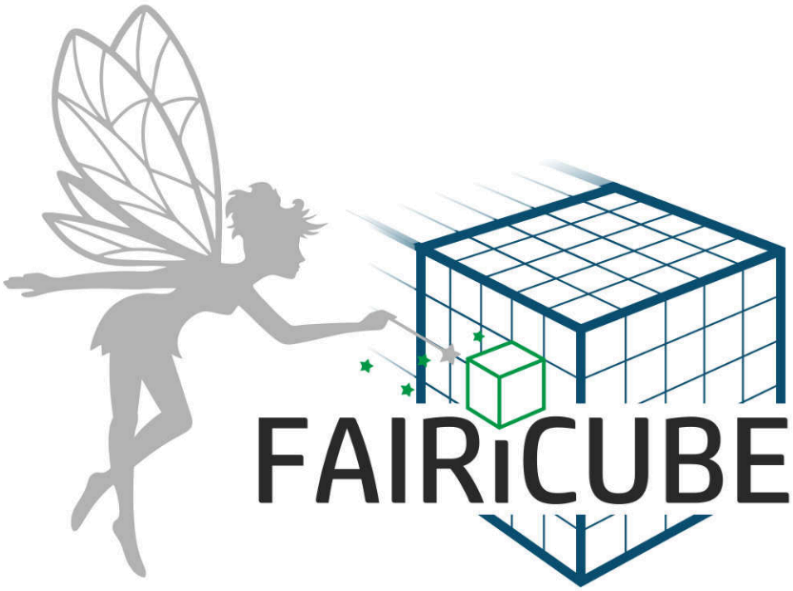

FAIRICUBE – F.A.I.R. INFORMATION CUBES

WP1 Manage

D1.2 Validation report

Deliverable Lead: 4SF
Deliverable due date: 30/06/2024

Version: 1.1
28/06/2024

Document Control Page

Document Control Page	
Title	D1.2 Validation report
Creator	NILU
Description	D1.2 Validation report
Publisher	"FAIRICUBE – F.A.I.R. information cubes" Consortium
Contributors	4SF, NIL, EPS
Date of delivery	30/06/2024
Type	Text
Language	EN-GB
Rights	Copyright "FAIRICUBE – F.A.I.R. information cubes"
Audience	☒ Public ☐ Confidential ☐ Classified
Status	☐ In Progress ☐ For Review ☒ For Approval ☐ Approved

Revision History			
Version	Date	Modified by	Comments
0.1	29/05/2024	Stefan Jetschny	Drafts structure and basic content
0.2.	18/06/2024	Jaume Targa	
0.3.	19/06/2024	María Colina	Further explanation of all sections
0.4	20/06/2024	María Colina and Jaume Targa	Further improvements following initial feedback
0.5	24/06/2024	Jaume Targa	Final document for review
1.0	26/06/2024	Wies Vullings	Review and format checking
1.1	28/06/2024	Maria Ricci, Stefan Jetschny	Improvements due to review, creation of section 2.7, improvements in WP4 checklist

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

Disclaimer

This document is issued within the frame and for the purpose of the FAIRICUBE project. This project has received funding from the European Union's Horizon research and innovation programme under grant agreement No. 101059238. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

Table of Contents

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

List of Figures

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

List of Tables

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

1 Introduction

In today's data-driven world, data science has emerged as a critical field that leverages statistical techniques, machine learning algorithms, and computational tools to extract meaningful insights from vast amounts of data. Organisations across various sectors, including healthcare, finance, retail, and technology, rely heavily on data science to inform decision-making, optimize operations, and drive innovation. However, the reliability and effectiveness of data science projects hinge on robust validation processes that ensure the accuracy, consistency, and applicability of the developed models and analytical methods.

Validation is a crucial step in the data science lifecycle, encompassing a series of checks and assessments designed to confirm that the outputs of a data science project meet the desired objectives and perform well in real-world scenarios. Without rigorous validation, data science models and insights may lead to erroneous conclusions, suboptimal decisions, and potential financial or reputational damage. Validation helps to ensure the model's predictive performance and generalizability to new, unseen data. It detects and mitigates issues such as overfitting, bias, and variance, confirms the robustness and reliability of data preprocessing steps and feature engineering, and evaluates the impact of assumptions and methodological choices on the final outcomes.

Task 1.4 aims to track validation activities across work packages. This deliverable report (D1.2 Validation report) aims to provide a comprehensive and holistic framework for the validation of FAIRiCUBE. As there are several FAIRiCUBE deliverables addressing validation on various levels, D1.2 aims to provide an overarching hierarchy of all these activities and put them relation. Finally, an outlook is given how FAIRiCUBE is providing online guidelines and checklist as web services for documenting and communicating validation results effectively.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

2 Overview of validation activities in FAIRiCUBE

FAIRiCUBE validation framework includes validation parts across FAIRiCUBE's work packages. Figure 1 provides a diagram on how the different validation deliveries link to each other. This report (D1.2 - Validation Report) provides a comprehensive overview of all validation activities within the FAIRiCUBE project, serving as a meta-analysis to ensure each component meets required standards. This report is crucial for demonstrating the robustness and reliability of the project's validation efforts, thereby supporting its overall integrity and success. Other key deliverables include:

- D2.5, which validates each Use Case at a thematic level, ensuring they align with project goals and user needs;
- D3.6 details validation for data processing and machine learning applications, emphasizing algorithm robustness and ethical considerations;
- D4.6 focuses on ensuring that the sharing components of the FAIRiCUBE Hub function correctly both individually and together;
- D5.3 outlines the validation checks for datasets ingested into the system, ensuring data integrity and quality;
- Lastly, D6.11 assesses the ethical use of AI, ensuring that AI applications are fair, transparent, and free from biases.
-

Together, these deliverables ensure comprehensive validation and ethical compliance across the FAIRiCUBE project. While the deliverables listed above, merely encompass the validation tasks, they establish dependencies and relationships across the project. The ethics assessment primarily affects data handling, processing, and use case (UC) work, rather than the Hub itself at this stage. The development of the Hub and the UCs serves as the foundation for outreach, communication, and dissemination activities. Furthermore, there is a hierarchical structure within the project, progressing from data collection and processing to addressing the research questions of the UCs, all performed on the FAIRiCUBE Hub.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

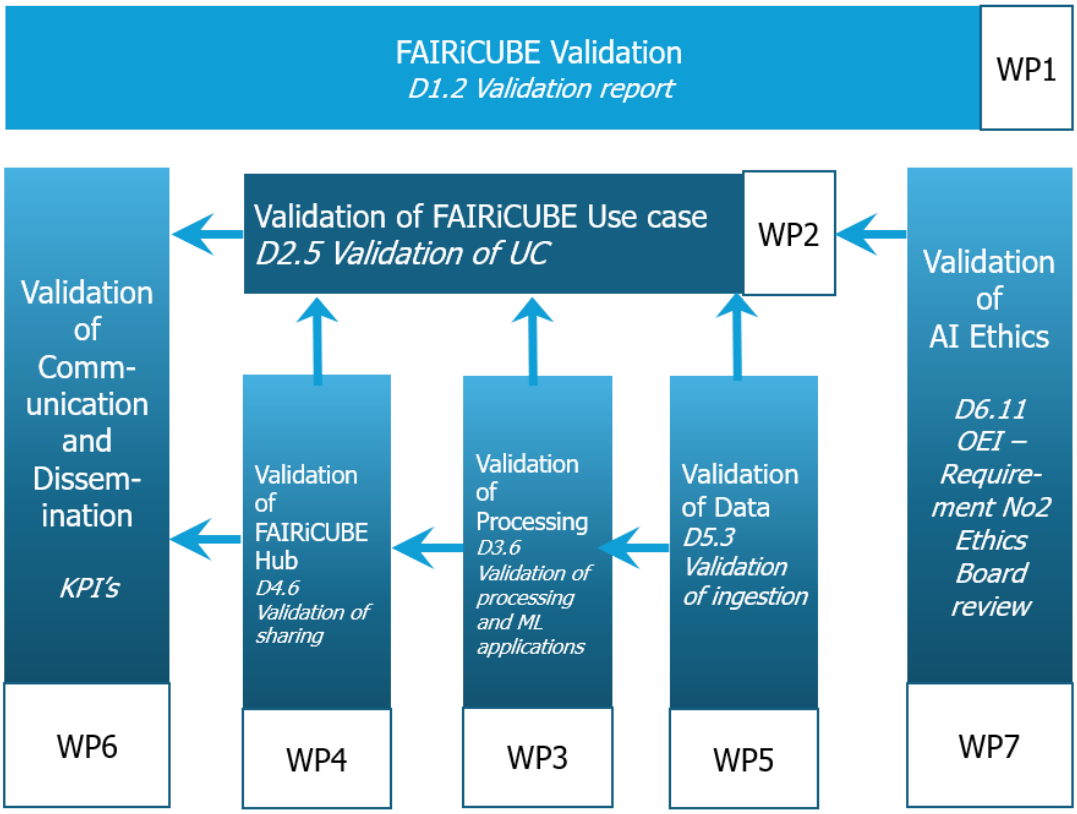

Figure 1 : FAIRiCUBE validation framework

Table 1 summarises the validation parts that each deliverable across different Work Packages include.

Table 1 : Overview of validation deliverables

Nr.	Title	Objective
D1.2	Validation report	Overview of all validation activities across FAIRiCUBE.
D2.5	Validation of UC	Description of the validation process to be performed for each Use Case at a thematic level. Covers both the Use Case specification as well as wider User assessment & fit-for-purpose checks.
D3.6	Validation of processing and ML applications	Description of the validation process to be performed for each Use Case at an analysis and processing level. Covers data processing validation including algorithm implementation validation, comprehensiveness of documentation and benchmarking; machine learning validation; ethical and bias validation.
D4.6	Validation of sharing	Description of the validation process to ensure that the components of FAIRiCUBE Hub on information, data, processing and portrayal each function individually, as well as jointly.
D5.3	Validation of ingestion	Description of the validation process to be performed for each dataset ingested to FAIRiCUBE. Covers data validation checks as well as statistical and spatial comparison of dataset characteristics with source dataset.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

KPIs	Validation of communications	Description of the validation measures for communication and dissemination through key performance indicators (KPIs) proposed in the project.
D6.11	AI ethics assessment	Description of the validation process to be performed for each Use Case to assure relevant considerations pertaining to ethical use of AI are met.

2.1 Validation of FAIRiCUBE

An overview of the validation activities under FAIRiCUBE is provided under D1.2 within WP 1. This report provides a comprehensive overview of the validation activities conducted across the entire FAIRiCUBE project. It documents methodologies and approaches for validating data and models, provides guidelines and checklists for consistent documentation of validation results, and offers an overarching hierarchy of all validation activities to ensure coherence and integration across the project. It is crucial for demonstrating the robustness and reliability of the project's validation efforts, thereby supporting its overall integrity and success.

2.2 Validation of WP2

The validation of WP2 (Use Case) is summarised in delivery D2.5. This document outlines the thematic validation processes applied to each Use Case within FAIRiCUBE, ensuring they align with project goals and assessing their suitability and effectiveness. It defines validation criteria, conducts user assessments, and performs fit-for-purpose checks to ensure practical applicability. Additionally, it documents the validation process and outcomes and ensures effective communication of results to stakeholders. D2.5 is strongly linked to D3.6 (Validation of processing and ML applications) and D4.6 (Validation of sharing). Table 2 shows an overview of the validations to be performed in the deliverable. A wider perspective is provided by D2.5, where more detailed information is given.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

Table 2 : Validation checks WP2

Process	Check type	Characteristic	Description
Use case specifications	Clear Goal defined	Specific	The objective is clearly defined and addresses a precise problem.
		Measurable	The objective's success can be measured.
		Achievable	The objective is attainable given available resources.
		Realistic	The objective is practical and relevant.
		Time-bound	The objective has a clear deadline.
	Required datasets identified	Relevant datasets	Identification of datasets that are most suitable for achieving goals.
		Resource estimation	Estimation of resources needed for acquiring and storing the datasets.
	Required ML/AI approaches identified	Suitable approaches	Identification of the best processing and ML/AI approaches to achieve the goals.
		Processing resources estimation	Estimation of the resources required for processing.
	Workflow designed	Workflow outline	Detailed steps of applying processing and ML/AI resources to the datasets.
		Documentation	Optimal documentation of the workflow using diagrams and regular updates.
	Visualisation of outputs designed		
User assessment/fit-for-purpose	Support the users' work	User assistance	Providing support for users to effectively utilize outcomes within their operational contexts.
	Service orientation	Transparent service chain	Establishing a transparent service chain allowing users to trace all alterations applied to data.
	Reliability	Method documentation	Public availability of production methods and complete metadata, independent QC of final products.
	Applicability	Fit-for-purpose	Ensuring products meet spatial, temporal, and quality requirements for user operational environments.
	Data systems stability, reliability, and interoperability	Stable and interoperable systems	Maintaining stability, reliability, and interoperability of data systems used for product creation and sharing.
		Standardised feedback collection	Formalised collection of user feedback through standardized questionnaires.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

2.3 Validation of WP3

The validation of WP3 (ML processing) is summarised in delivery D3.6. This deliverable focuses on validating data processing and machine learning applications within each Use Case. It ensures algorithms are correctly implemented, documentation is thorough, and benchmarking is conducted. It describes the validation process for data processing workflows, including data cleaning, transformation, and integration, and outlines procedures for algorithm implementation and documentation. Additionally, it addresses the ethical implications and potential biases in machine learning models.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

Table 3 : Validation checks WP3

Process	Check type	Characteristic	Description
Data processing Validation	Algorithm implementation validation	Technical Robustness and safety	Ensuring robustness and safety of the implementation through e.g. unit tests that can verify that unit of codes behaves as expected in isolation (inc. individual components of our data processing pipeline and algorithms).
		Assess the interactions	Conduct integration tests to assess the interactions between different components of the system, verifying that data flows smoothly between processing stages and that the overall pipeline functions correctly.
		End-to-end testing	Assess the entire data processing workflow. This involves testing the complete pipeline with controlled (i.e. synthetic) but representative data to ensure that it produces the expected results.
		Cross-validation	Cross-validation to assess the model's performance across different subsets of the data, ensuring it generalizes well to new, unseen data
	Benchmarking	Monitor compute resources	Monitoring and storing the consumption of computational resources as defined and described in the FAIRiCUBE GitHub repository.
		Re-run and compare	If performance is in question, re-run of an application can be advised. The monitoring results of a compute task can be compared with other similar tasks as listed in the FAIRiCUBE knowledge base (htts://fairicube-kb.dev.epsilon-italia.it/).
	Comprehensive documentation	Documentation and transparency	Document the rationale behind design choices, assumptions, and dependencies to allow transparency of the processing and ML application methods.
		Meta-data	Update, complete, and maintain meta-data records associated to the data set. This applies to the FAIRiCUBE analysis / processing meta-data.
Machine learning validation		Dataset preparation for training	Create subsets of data separate for training, testing (and validation), check the selection method (random, by consecutive index).
		Define appropriate validation metrics	Select and document appropriate metrics such as total accuracy, precision, recall, F1 score, area under the ROC curve to evaluate the performance of your ML method. Define expectations first, establish baseline methods to compare against.
		Prevent/Test overfitting and underfitting	Comparing performance / accuracy metrics from applying an ML model to different datasets which have not been included in the training of the ML model will give insights on over or underfitting.
		Statistical bias validation	Checking the statistical distribution of the input feature both within one feature space and across features will avoid unwanted biases in the training of the ML model. Some ML methods require certain statistical distributions (Gaussian or even distribution, etc.), some other methods require scaling of the data features.
		Human agency and oversight	Implementation of human oversight mechanisms such as human-in-the-loop, human-on-the-loop, and human-in-command approaches. At each step of the ML application, user feedback and interaction should be foreseen.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

2.4 Validation of WP4

The validation of WP4 (Share) is summarised in delivery D4.6. This document describes the validation procedures for the sharing components of the FAIRiCUBE Hub, ensuring that information, data, processing, and portrayal modules work effectively both independently and together. It outlines the process for validating the interoperability and integration of different components, ensuring correct functionality of data sharing mechanisms and interfaces. Additionally, it documents any issues or improvements needed and assesses shared components to ensure they meet user requirements and project standards.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

Table 4 : Validation checks WP4

Component	Process	Check type	Characteristic	Description
Core	Hub	Findability	Findability of components	FAIRiCUBE Hub UI has clear navigation structure and section names; index of all components is up to date; a search function exists.
		Authentication	Authentication across components	Common, secure authentication method available across all components; standard authentication protocol implemented; user roles defined and documented.
		Interplay	Accessibility of datasets	Data resources can be retrieved for use from their metadata records; access modalities clearly described (e.g. local download, cloud APIs…)
			Accessibility of analysis/processing resources	A/p resources can be retrieved for use from their metadata records; access modalities clearly described (e.g. Jupyter notebook, Python library)
			Accessibility of linked datasets and a/p resources	Provenance data or a/p resources or derived data resources are described and accessible
			Storage and accessibility of processing results	Processing results are securely stored and readily accessible for inspection and reuse by authenticated as well as unauthenticated users
			Visualization of processing results	Visualization software for gridded and non-gridded data is an integral component of the Hub and available to users
		User support	User support/helpdesk	Users know how to ask for support in case of questions or technical problems; user support is conducted in a structured manner (e.g. ticketing system); enquiries are timely answered.
Information	Documentation	Findability	Findability of required information	Information is organised and structured; consistent labeling of sections; search function is implemented
		Correctness	Correctness of available information	Periodic reviews of the content are conducted; information is kept up to date; broken hyperlinks are replaced
		Completeness	Completeness of available information	All components/topics mapped to respective documentation/articles; external expert review mechanism is implemented
		Usefulness	Usefulness of available information	Periodic reviews of the content are conducted; structured system for collecting user feedback is in place (e.g. comments, rating)
		Ease of Provision	Ease of provision of documentation	Regular maintenance of the system is scheduled; contribution methods are intuitive and documented
	Knowledge Base	Findability	Findability of required information	Information is organised and structured; consistent labeling of sections; search function is implemented
		Correctness	Correctness of available information	Periodic reviews of the content are conducted; information is kept up to date; broken hyperlinks are replaced
		Completeness	Completeness of available information	All components/topics mapped to respective documentation/articles; external expert review mechanism is implemented
		Usefulness	Usefulness of available information	Periodic reviews of the content are conducted; structured system for collecting user feedback is in place (e.g. comments, rating)
		Ease of Provision	Ease of provision of information for the KB	Regular maintenance of the system is scheduled; contribution methods are intuitive and documented

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

Data	Metadata Editors	Persistence	User input is persisted in static file	Ensure all values entered in the GUI are accurately saved in a JSON file.
		Alignment	Aligned metadata schema across project deliverables, Editor and metadata records	The Editor and the metadata records conform to the metadata schema documented in the project deliverable.
		Findability	Code lists findable and complete, sorted	Code lists (implemented as dropdown lists) contain all necessary options alphabetically sorted; a dictionary is provided
		Labeling	All labels in the GUI described	Every input field has a label and a clear description.
		Errors	Human-readable error messages	Error messages are clear and provide actionable information.
	Data Catalog	Findability	Findability of required data based on search criteria	Users can efficiently locate datasets using specific search criteria.
		Simple Search	Simple search	Basic search interface for quick dataset retrieval implemented.
		Advanced Search	Advanced search	Detailed search interface to search across all metadata fields implemented.
		Navigation	Ability to navigate to the identified dataset (either by link to original data download, when possible; or API endpoint with related API document for data access, e.g. endpoint for data in S3 bucket)	Data resources can be retrieved for use from their metadata records; access modalities clearly described (e.g. local download, cloud APIs…)
Data Access	WCS	Correct Response	Request returns correct responses in accordance with the WCS specification	Ensure all WCS requests return valid responses according to the specification.
		Measure Types	Measure types from SWE Common used according to the original specification	Ensure measure types are used correctly as per the SWE Common specification.
		XML Validity	XML responses valid	Validate all XML responses against the relevant schemas.
		Extent/Bbox	Extent/bbox information aligned with GML	Ensure extent and bounding box information follow GML standards.
		Subset Accuracy	Data returned is for the requested subset of the original source data	Ensure data returned matches the specified subset criteria.
		Format	Data is returned in the requested format	Ensure data is returned in the user-specified format.
	S3	Pricing	Transparent pricing scheme for data access	The different pricing schemes for data access are clearly and transparently described
		Optimisation	Data formats are optimized for cloud storage	Only cloud-optimized data formats (e.g. COG, zarr) are allowed on S3
		Authentication	Clear authentication requirements	Authentication options (e.g. external authentication) and requirements (e.g. subscription) are clearly and transparently described.
	Sentinel Hub	Pricing	Transparent pricing scheme for data access	The different pricing schemes for data access are clearly and transparently described
		Authentication	Clear authentication requirements	Authentication options (e.g. external authentication) and requirements (e.g. subscription) are clearly and transparently described.
	GitHub	Repository Content Overview	Provide users with a clear understanding of the repository's contents and structure.	Include a detailed README file that outlines the purpose, components, structure, and usage instructions of the repository.
		License	Clearly define the terms under which the repository's content can be used, modified, and shared.	Add a LICENSE file specifying the chosen open-source license (e.g., MIT, Apache 2.0).
		Findability	Ensure the repository is easily discoverable by users searching for related content.	Use relevant keywords, tags, and topics in the repository's metadata to optimize discoverability.
Processing	Lab	Execution	Scripts and Jupyter NB can be executed without error	Jupyter Lab and Python is up to date; essential libraries are installed and up to date (e.g. ipykernel)
		Libraries	Required libraries are available	Commonly used libraries for EO and ML are installed as part of the initial set up
		Library Addition	If libraries are missing, mechanism in place to add	Structured mechanism to request new libraries is implemented (e.g. through helpdesk)

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

		Scaling	Processing resources can be scaled to the requirements of a specific task	Ensure processing resources can be scaled as needed.
		Backup	Backup mechanism in place	Data and configuration of the Lab is regularly backup; the backup mechanism is known to the user; users can request restoring older backups
		Activation	Requested services activated and correctly set up	Structured way to request activation of additional services implemented; activated services are tested before being made available
WCPS		Response	Request provides a correct response	Ensure the request provides a correct response.
		Subset Accuracy	Data returned is for the requested subset of the original source data	Ensure data returned matches the requested subset.
		Accuracy	Processing steps numerically accurate	Ensure numerical accuracy of processing steps.
		Format	Data is returned in the requested format	Ensure data is returned in the requested format.
		Integration	Code can be integrated as UDF	Ensure code can be integrated as User-Defined Functions (UDFs).

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

2.5 Validation of WP5

Validation of WP5 (Ingest) is summarised in delivery D5.3. These deliverable details the validation checks performed on datasets ingested into the FAIRiCUBE system. This includes procedures for validating data integrity, quality, and comparing statistical and spatial characteristics with original sources for accuracy and consistency. The deliverable outlines procedures for ensuring data accuracy, completeness, and consistency, documenting the validation process and results to maintain data provenance and meet project standards. Any identified data issues are addressed to enhance reliability.

Table 5 : Validation checks WP5

Process	Check type	Characteristic	Description
Data Validation for Ingestion	Data validation checks	Duplicates	Check for duplicate entries.
		Date overlap	Check for overlaps in the date column, i.e., repeated dates with different values (only for timeseries).
		Date gaps	Check for missing dates between start and end date (only for timeseries).
		No data values	Verification of the correct use of no data.
		Value types	Check if data types are correct (string, integer, float, datetime format).
		Value encoding	Check if the encoding of the data is correct (e.g., character encoding is utf-8; point (.) is used as decimal separator).
	Statistical Comparison of Characteristics with Source Dataset	Completeness	Calculate and evaluate the ratio of not-NULL values.
		Count of distinctive values	Count and evaluate the number of distinct values in the dataset.
		Ratio of the most frequent value	Determine the number of occurrences for the most frequently repeated value, normalized by the batch size.
		Maximum	Maximum value of the dataset
		Mean	Mean value of the dataset
		Minimum	Minimum value of the dataset
		Standard deviation	Standard deviation of the dataset
		Number of records	Number of rows in the dataset (e.g., number of polygons)
		Date range	Start and end date (only for timeseries)
	Spatial Characteristic Comparison with Source Dataset	Grid boundaries	Top-left and bottom-right coordinates (only for gridded datasets)
		Data completeness	Verify data set is complete (total area, total number of features or pixel)
		Projection/CRS	Verify correct use of the projection/CRS.
		Pixel size	Verify that the pixel size is correct
		Number of bands	Verify all channels have been transmitted correctly.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

		Number of attributes	Verify all attributes of the table or vector dataset have been transmitted correctly.
		Datatype	Check if the data type is OK.
		Data format	Check if the ingested data follows the desired standard format (e.g., for raster, cloud optimized Geo Tiff)
		Centre coordinates and total area	For vector data, check if centre coordinates of polygons match (to find shifted polygons or duplicate polygons)

2.6 Validation of WP6

In WP 6 there is no formal validation task or validation-related deliverable included, but for communication and dissemination we have a number of KPI's that can be seen as validation measures. In the project proposal we have been provided a number of communication and dissemination KPI's which are listed in Table 6. The progress of the KPI's is reported in the Midterm review report and in D6.4.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

Table 6 : Overview of communication and dissemination (WP6) KPI's

Communication channel	KPIs	Description
EU policy	~ 3 events	Participate by presenting and attending events which target EU policy makers
Domain specific events	~ 5 events	• Participate by presenting and attending domain specific events e.g. related to use cases or WP specific tasks
Newsletter	~ 6 newsletters	Publish newsletters giving an overview of FAIRiCUBE related news items
Website	~ 2K views - > changed to 1K	Keeping the website up to date and attract viewers by using social media and newsletters to interest people for FAIRiCUBE
Social media	~ 1K followers -> changed to 500	Using Linkedin and X for posting to attract and update followers and to enthuse people to get to know FAIRiCUBE and invite them to FAIRiCUBE events.
Press releases and articles published in national/regional/European online media	~ 5 articles	Publish articles to give more in-depth information on a specific FAIRiCUBE topic.
Innovation workshops	~ 2 workshops	Organise workshops with the aim to generate ideas on innovations in relation to FAIRiCUBE to feed into the upscaling plan.
Trainings and master workshop	~ 3 training days	Organise training days and workshops for different stakeholders on FAIRiCUBE related topics, such as online seminars about the use cases or training days on the FAIRiCUBE Hub
Scientific papers	~ 3 papers	Publish peer reviewed scientific papers on FAIRiCUBE related topics

2.7 Validation of WP7

Formally, deliverable is D6.11 (AI ethics assessment, Ethics board review report) is part of WP6 due to the initial project setup and the unforeseen requirement to include an AI ethics assessment in the FAIRiCUBE project. Thematically however, D7.1 (general description of ethics aspects considered in FAIRiCUBE) and D6.11 (approach and implementation of the AI ethics assessment) belongs to the *WP7 Ethics requirements* which is focusing on validating ethical considerations in AI applications across each Use Case. It defines criteria and standards for fairness, transparency, accountability, and social impact of AI technologies used in the project. The deliverable outlines assessment procedures to ensure fairness and absence of bias in AI models, emphasizing transparency in their development and deployment. It documents ethical considerations and validation outcomes, providing guidelines for the ethical implementation of AI throughout project activities.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

Table 7 : Validation checks WP7 (AI ethics assessment)

Component	Process	Check type	Characteristic	Description
Ethics (Trustworthy AI)	Fundamental rights	Fundamental rights	Safety, health, non-discrimination, freedom of association	How are you dealing with the effect of the application on the rights to safety, health, non-discrimination, and freedom of association?
	Privacy and data protection	GDPR compliance	Personal data protection	How are you implementing the GDPR to safeguard the personal data protection rights of those you collect personal data from?
	Transparency rights	User rights	Notification, access, control, explanations, audit	Do you include the following user rights to: be notified that their data is being processed/collected, access information on which personal data are collected, control their data, access explanations of results produced by the system, be informed of who, when and how the system can be audited?
	Accessibility	Inclusivity	Demographics, language, disability, digital literacy, financial accessibility	Can your app/system/resource be used by all regardless of demographics, language, disability, digital literacy, and financial accessibility?
	Education and tutorials	User capability	Information, correct use	Do you ensure that users are informed and capable of using the system correctly?
	Data management	Data minimization	Local and temporary storage, encryption, necessity	Do you comply with the data-minimization principle, i.e., usage of local and temporary storage and encryption, based on principles of data protection by design? Do you ensure that only strictly necessary data are captured and processed?
	Security	User authentication	Access, modification, disclosure, identifiers	Do you have user authentication in place to prevent risks such as access, modification, or disclosure of the data? Do you use unique and pseudo-random identifiers, renewed regularly and cryptographically strong?
	Ease to deactivate/remove	System removal	Deactivation, removal ease	How easy is it to deactivate or remove the system and data once users are no longer interested or need the system?
	Ease to access services without using the AI system	Service alternatives	Non-system alternatives	In the case of AI systems aimed to replace or complement (public) services, are there full non-system alternatives?

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

	Open-source code	Development access	Participatory, multidisciplinary	Is the development participatory and multidisciplinary? What kind of access to the code and development is there?
	Ownership	Resource clarity	Clear ownership	Is the ownership of the resource clear?
	Openness about Data governance	Data sharing	Open data license	How open is Data governance? Do you have a policy and actions to share data under an open data license?
	Legislation and Policy	Relevant policies	Legislation, other policies	Are there explicit legislation and/or other policies relevant to your system/resource?
	Design Impact Assessment and Open Development Process	Process transparency	Design process accessibility	How publicly accessible is the information about the design process leading to this resource?
	Right to contest/liability	User rights	Contest decisions, human intervention	How are users able to contest decisions/actions or demand human intervention?
GDPR applicability	List of data	Data types	Processed data types	List the types of data you will process in your pilot
	Personal data	Data processing	Personal data definition	Will personal data be processed for the proposed pilot? Personal data is understood as data that may directly or indirectly identify a natural person.

Funded by the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.

3 Outlook on validation services

The section provides a high-level overview of the planned validation process. By taking a flexible and pragmatic approach, the team aims to implement a validation system that is both effective and adaptable to future needs. The exact details of the implementation will be refined over time, ensuring that the system evolves to meet the project's requirements and constraints.
The future steps and plans for the validation process, providing a tentative roadmap for gathering all validation information is covered below. The goal is to ensure that the process is both systematic and adaptable to evolving project requirements and resource constraints.

3.1 System for validation

The team has identified the need for a system where responsible parties can efficiently enter and manage validation checklists. The system should ideally support the automatic or semi-automatic entry of checklist items. One option being considered is using GitHub for forms, which would streamline the entry process, ensuring consistency and reducing manual errors. However, the specific tools and technologies to be used are still under discussion, and the implementation details remain to be determined.

3.2 Searchable Interface

It is essential that the validation system be searchable by other users. This functionality is desired to enable stakeholders and use cases to easily locate specific validation information based on keywords or predefined criteria. While the exact implementation details are yet to be decided, the goal is to enhance the accessibility and usability of the validation system, making it easier for users to find the information they need.

3.3 Linking Validation to Metadata

A critical aspect of the validation process is the aspiration to link checklists to metadata records. This linkage aims to provide a clear connection between the validation process and the specific data sets or processing resources being validated. The methods for achieving this linkage are still being explored and will be refined as the project progresses, ensuring that the system is effective and user-friendly.

Funded by
the European Union

This project has received funding from the Horizon Europe research
and innovation programme under grant agreement No. 101059238.

4 Summary

The Validation Report (D1.2) of FAIRiCUBE consolidates and integrates validation activities across its diverse work packages. This report serves as a comprehensive framework outlining validation processes for ensuring the accuracy, consistency, and applicability of data science models and methodologies. Key deliverables such as D2.5, D3.6, D4.6, D5.3, and D6.11 address thematic validation of use cases, data processing and machine learning applications, FAIRiCUBE Hub components, dataset integrity checks, and ethical considerations in AI applications. By documenting methodologies, tools, and outcomes, the report ensures coherence and reliability across all project components, supporting overall project integrity and success.

In conclusion, D1.2 underscores FAIRiCUBE's commitment to rigorous validation practices, essential for mitigating risks associated with data-driven decision-making and enhancing project impact. By establishing a structured approach to validation, FAIRiCUBE aims to uphold high standards of accuracy, ethical compliance, and operational effectiveness throughout its implementation.

Funded by
the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.