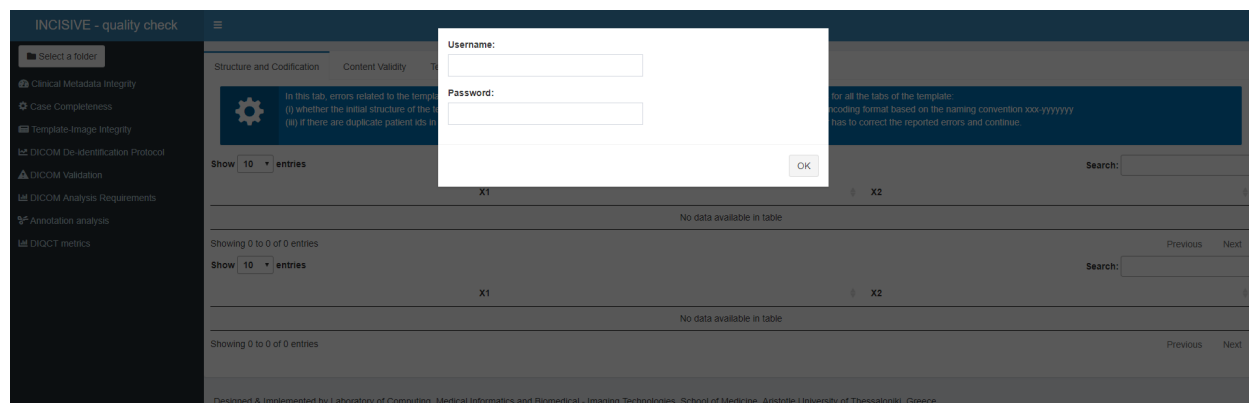
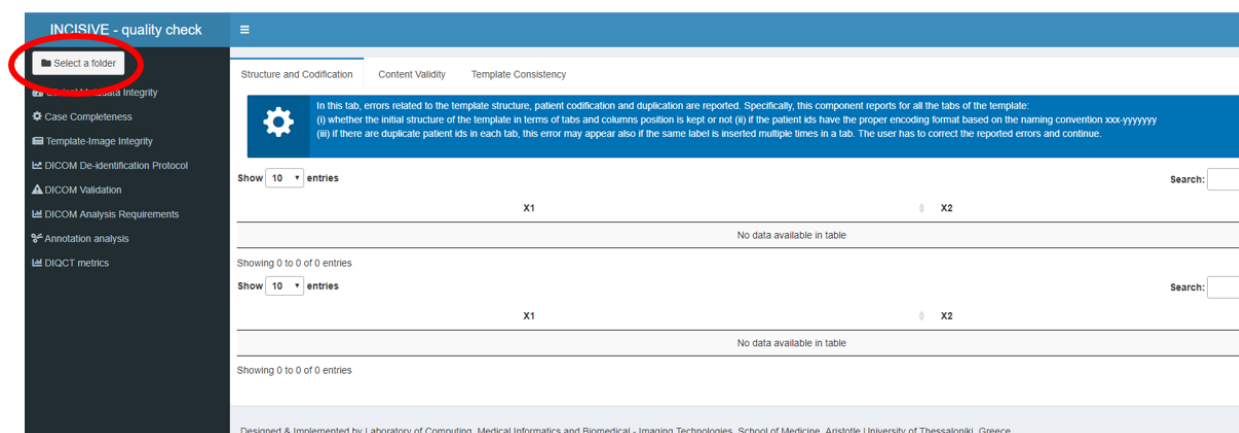


For the next steps, you can watch an 8-minute video on how to use the quality tool [here](#), or read the detailed instructions below:

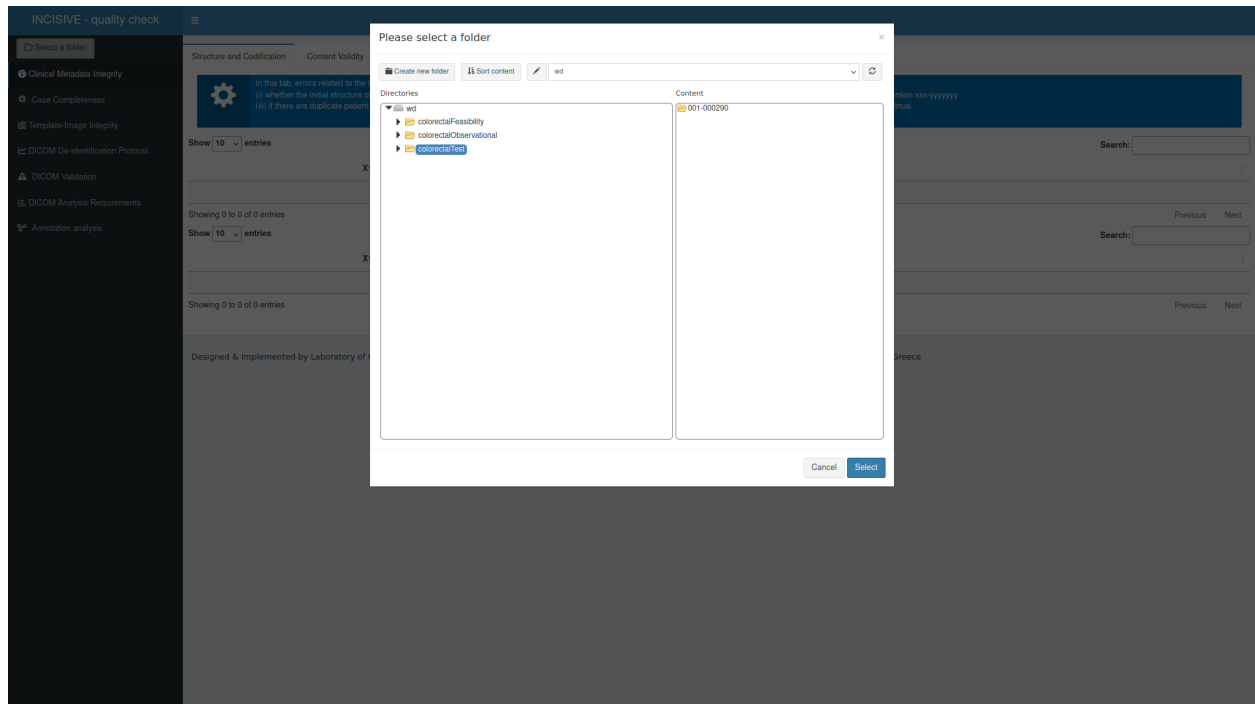
1. The quality tool is a web application and thus the user has to open a web browser (e.g. google chrome or Mozilla Firefox) and type: 127.0.0.1:3838 in the URL address. A web page appears with a login window. The user needs to insert the username and password that has already send to him and click on the 'ok' button:



2. In order to start the quality check of the data, the user must click on the button 'Select a folder' button located on the upper left side of the page, in order to select the data needed to be analyzed.



3. The directory tree of the computer appears, and the user needs to select the parent folder where the data lies, the images and the excel file (note: the excel file does not appear in this window) and click on the button 'Select'.



4. After that, the first component of the quality tool named *Clinical Metadata Integrity* starts to run. When it is finished the results appear in three tabs named *Structure and Codification & Content Validity*.

In the first tab, errors related to the template *Structure and Patient codification* are reported. Specifically, this component reports: (i) whether the initial structure of the template in terms of tabs and columns position in the template is kept or not and, (ii) if the patient ids have the proper encoding format and if there are duplicate patient ids in the template.

Structure and Codification

Content Validity

Template Consistency



In this tab, errors related to the template structure, patient codification and duplication are reported. Specifically, this component reports for all the tabs of the template:

- (i) whether the initial structure of the template in terms of tabs and columns position is kept or not
- (ii) if the patient ids have the proper encoding format based on the naming convention xxx-yyy-yyyy
- (iii) if there are duplicate patient ids in each tab, this error may appear also if the same label is inserted multiple times in a tab. The user has to correct the reported errors and continue.

Show

All

entries

Search:

Aspect

Result

All

All

1

Tabs Integrity

Tabs follow the proper sequence

Showing 1 to 1 of 1 entries

Previous

1

Show

All

entries

Search:

Tab from template

infoTab Columns

Columns

MutationsTab Columns

ResultsTab Columns

Correct Columns

Accuracy

info PatientIdEncoding

PatientIdEncoding

Mutations PatientIdEncoding

Results PatientIdEncoding

info DuplicateEntries

All

All

All

/

All

All

All

All

All

1

General

Columns: Case* are missing

01-000412

001-000431

2

BaselineTab

Columns: Weight Loss, Bowel obstruction are in the wrong order

3

TimepointsTab

All columns are in place

4

TreatmentTab

All columns are in place

Columns: 18q

Some common errors in this step might be:

- Line 3: A new column was inserted and has to be deleted
- Line 8: Wrong patient codification. The patient id should be 001-000026 instead of 01-000026.
- Line 16: A duplicate patient id was identified in Timepoints tab. This may also result from inserting the same label two times for a patient. The only label that can be inserted multiple times is label that represents 'other' option.

The user has to open the template and correct the above errors.

In the second tab, *Content Validity*, the user can see errors related to the template content are reported. Specifically, this component checks if the standards and terminologies proposed for the allowable values of all fields of the templates are followed. It reports, for each patient separately in a different row, the fields of each tab that do not comply with the proposed value range.

Structure and Codification										
Content Validity										
Template Consistency										
In this tab, errors related to the template content are reported. Specifically, this component checks if the standards and terminologies proposed for the allowable values of all fields of the templates are followed. It reports, for each patient separately in a different row, the fields of each tab that do not comply with the proposed value range. For each field a flag is reported to specify the type of the error: (i) T for data type, if the value inserted is inserted in the wrong format e.g. string instead of number, (ii) V for value, if the value inserted does not comply with the proposed value range. The user has to correct the reported errors and continue.										
Show All entries Search:										
Subset Selection										
Patient ID	General Info	Baseline	Timepoints	Treatment	Histology...Mutations	Lab.Results	TimepointsBoundaries	TimepointsOrder	Validity	
All	All	All	All	All	All	All	All	All		
1	001-000411			Label* in TP: 00, Type of CT in TP: 0 (V) 1 (V)		P: GT in TP: 2 (V) ALP in TP: 2 (V)		Order is correct	86.88	
2	001-000413			Label* in TP: 00		Label* in TP: 00		Order is correct	88.85	
3	001-000414			Label* in TP: 00, Date of surgery in TP: 0 (T), Type of CT in TP: 0 (V)		Label* in TP: 00			88.71	
4	001-000415			Label* in TP: 00		Label* in TP: 00, calcium in TP: 0 (V)			87.67	
5	001-000416			Label* in TP: 00, Type of CT in TP: 0 (V)		Label* in TP: 00		Order is correct	86.94	
6	001-000417			Label* in TP: 00, Type of CT in TP: 0 (V)		Label* in TP: 00		Order is correct	86.21	
7	001-000418			Label* in TP: 00, Type of CT in TP: 0 (V)	Colon Cancer Type in TP: 1 (V)	Label* in TP: 00	TP2 is out of boundaries / TP3 is out of boundaries	Order is correct	85.9	
8	001-000419			Label* in TP: 00, Type of CT in TP: 0 (V)		Label* in TP: 00			90.07	
9	001-000420			Label* in TP: 00		Label* in TP: 00		Order is correct	90.3	

For each field reported as erroneous, a flag is reported to specify the type of the error: (i) T for data type, if the value inserted is inserted in the wrong format e.g. string instead of number, (ii) V for value, if the value inserted does not comply with the proposed value range. The user needs to go back to the template and correct the errors reported by the tool. These errors mean that the inserted values do not comply with the proposed value ranges and guidelines. Examples given in the image below.

The accepted value is an ATC code, e.g. J07XA01 Decimal numbers must be inserted with '.' Insertions must follow the value range, e.g. T3a-b is not an accepted value and has to be replaced with T3		
Type of CT Terminology: https://www.whocc.no/atc_ddd_index/ CAPESITABMINA CAPESITABMINA CAPESITABMINA CAPESITABMINA CAPESITABMINA CAPESITABMINA CAPESITABMINA CAPESITABMINA	Max Tumor Diameter (in cm) 4,5 4,7 5 3,2 3,5 16 8 3,5	Ti Ti, T1, T2, T3, T4a, T4b T3 T4 T3a-b T2 T1 T4 T3 T3 T4 T3 T3 T2 Ni N0, N1a, N1b, N1c, N2a, N2b N0 N1 N0 N0 N0 N0 N1 N0 N0 N0 N0 N0 Mi M0, M1a, M1b, M1c M0 M0 M0 M1b M0 M0 M1b M0 M1b M0 M0 M0

5. Then the user must click on the second component named *Case Completeness* and the results appear in two tabs named “Clinical Metadata Completeness” and “Imaging Data Completeness”.
 - a. In the first tab, the component checks the completeness of the clinical metadata in the template for each case. Specifically, this component reports for each patient and for each tab the missing mandatory fields. It also reports the percentage of completeness for each patient. No action is required by the user in this stage, but this step needs to be executed because it is a dependency for the next step. Note that for the patients that there was an error in codification or were found as duplicate we don’t have a result in this step.

INCISIVE - quality check

Select a folder

- Clinical Metadata Integrity
- Case Completeness**
- Template-Image Integrity
- DICOM De-identification Protocol
- DICOM Validation
- DICOM Analysis Requirements
- Annotation analysis

Clinical Metadata Completeness Imaging Data Completeness

1. Clinical Metadata Completeness: This component checks the completeness of the clinical metadata in the template for each case. Specifically, this component reports for each patient and for each tab the missing mandatory fields. It also reports the percentage of completeness for each patient. The user should try to provide the missing information. 2. Imaging data completeness: This component presents an overview of the imaging data provided. Specifically, it reports what modalities are provided in each timepoint for each patient. No action is required by the user.

Show: All entries Search:

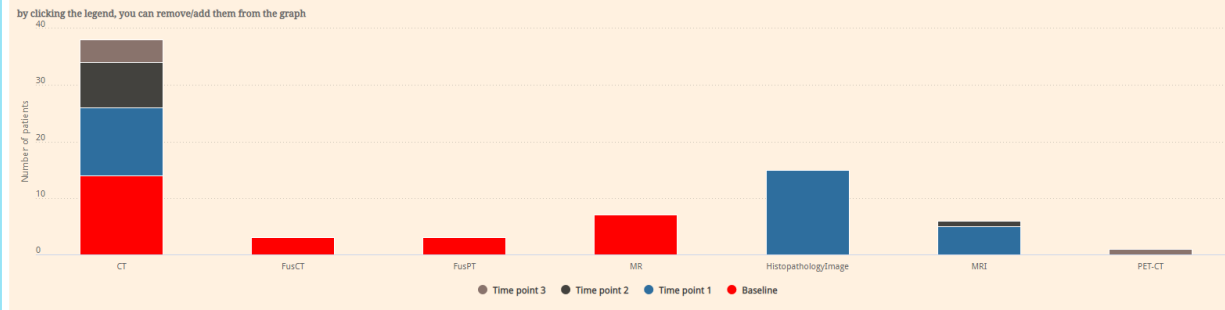
Subset Selection

	Patient Number	General info	Baseline	Timepoint1	Treatment	Histology	Mandatory Fields (%)
	All	All	All	All	All	All	All
1	002-000099		Max Tumor Diameter,Borders,T1,N1,M1			Lung Cancer Grade	38
2	002-000102		Solidity,Borders,T1,N1,M1			Lung Cancer Grade	28
3	002-000098		Solidity,Borders			Lung Cancer Grade	38
4	002-000095		Max Tumor Diameter,Solidity,Borders,Lung nodules,T1,N1,M1			Biopsy Date*,Lung Cancer Grade	25
5	002-000094		Lung,Solidity,Borders,Pleural Effusion,Lymph nodes,Lung nodules,T1,N1,M1			Lung Cancer Grade	16
6	002-000101		Stage,Max Tumor Diameter,Lung,Solidity,Borders,Pleural Effusion,Lymph nodes,Lung nodules,T1,N1,M1			Lung Cancer Grade	9
7	002-000103		Solidity,Borders,Pleural Effusion,T1,N1,M1			Lung Cancer Grade	28
8	002-000104		Solidity,Borders,Pleural Effusion,Lung nodules,T1,N1,M1			Lung Cancer Grade	25
9	002-000105		Solidity,Borders,Lung nodules,T1,N1,M1			Lung Cancer Grade	28
10	002-000106		Borders,T1,N1,M1			Lung Cancer Grade	34

- b. In the second tab, an overview of the imaging data is provided. Specifically, it reports what modalities are provided in each timepoint for each patient. The results are depicted as stacked bar charts and in tabular format, too. No action is required by the user.

This component presents an overview of the data provided, what modalities in each timepoint, as well as the percentage of mandatory fields that are present for each patient. No action is required by the user.

Imaging modalities per time point



Show All entries

Search:

Subset Selection

Patient.Number.	Baseline	Time Point 1	Time Point 2	Time Point 3	Time Point 3_1
All	All	All	All	All	All
1 001-000411	CT	CT	CT		
2 01-000412					
3 001-000413	FusCT/FusPT	CT			

- The third component named *Template-Image Consistency* checks if the file structure of the images is compliant to the template and performs a proper renaming of the studies' folders so they can be stored in a unified and commonly understandable way, which is a prerequisite in INCISIVE. In case that there is an inconsistency between the template and the folders a message appears for the specific patient. If the folders were renamed properly, no message appears.

INCISIVE - quality check

Select a folder

- Clinical Metadata Integrity
- Case Completeness
- Template-Image Consistency
- DICOM De-identification Protocol
- DICOM Validation
- DICOM Analysis Requirements
- DICOM Overall Patient Evaluation

This component checks if the file structure of the images is compliant to the template and performs a proper renaming of the studies' folders so they can be stored in a unified and predefined way. In case that there is an inconsistency between the template and the folders a message appears for the specific patient. If the folders were renamed properly, no message appears. The user has to correct the reported errors and continue.

Results of the Template-Image Consistency and Renaming check

Show All entries

Search:

patientID	path	error message
All	All	
1 001-000224	N/A	No valid timepoints index in csv template (or patient empty entry)
2 001-000225	/home/saloula/Documents/ColorectalProspective/001-000225	# of studies in template do not match # of folders in dir
3 01-000226	N/A	No folder with that patient ID
4 001-000227	/home/saloula/Documents/ColorectalProspective/001-000227	Folder already renamed 001-000227_CT_TP3

The errors that might appear are as in the image above:

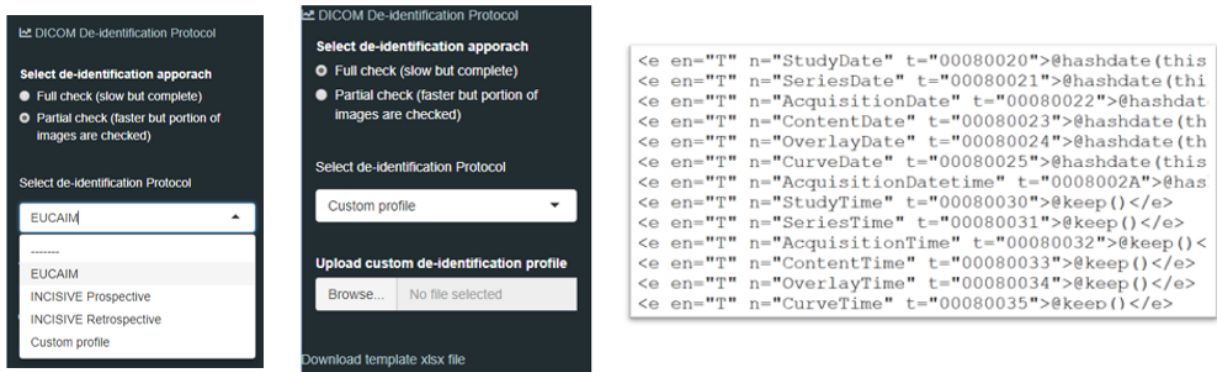
- No valid timepoints index in csv template (or patient empty entry): which means that the previous step did not produce a line for this patient, either because the patient id is not correct or nothing was inserted in the template.
- # of studies in template do not match # of folders in dir: which means that the insertions for the modalities in the template do not agree with the folders inside the patient directory. So, for the patient 001-000025, from the previous step the user can see that in baseline a CT and an MRI is inserted, however the folder contains only CT. The template must be corrected accordingly.
- No folder with that patient ID: which means that for this patient there is no available folder with images, either because the patientid is wrong or the folder is missing.

iv. Folder already renamed 001-000227_CT_TP3: Means that the renaming was done properly.

The user needs to go back to the data and correct the errors reported by the tool.

7. The user then must click on the fourth component named *DICOM de-identification Protocol*. In this component the errors appear in two tabs. The results are provided in three different tabs, each of them focuses on different aspects of the de-identification check.

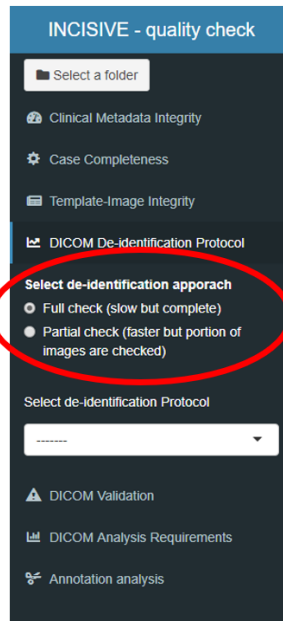
In the first tab, the component checks whether the de-identification protocol, already defined in INCISIVE, is properly applied in the metadata included on the imaging files. The user is given the ability to select the de-identification protocol that was applied and is related to the phase of the data collection (retrospective, prospective or custom). In case a custom profile is selected, then the user is prompt to upload the respective file describing the de-identification protocol, whereas a template file is also available to download. This file is the standards file that is used in CTP anonymizer and describes the action to be undertaken for each of the DICOM attributes.



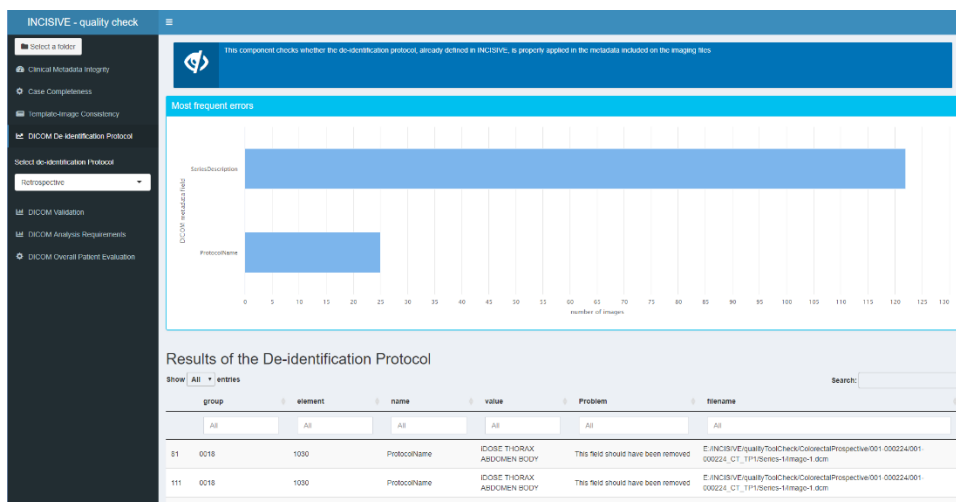
The image displays the 'DICOM De-identification Protocol' interface. It features two main tabs: 'Select de-identification approach' and 'Select de-identification Protocol'. The first tab offers two radio button options: 'Full check (slow but complete)' and 'Partial check (faster but portion of images are checked)'. The second tab includes a dropdown menu for 'Select de-identification Protocol' with options like 'EUCAIM', 'INCISIVE Prospective', 'INCISIVE Retrospective', and 'Custom profile'. Below this is a section for 'Upload custom de-identification profile' with a 'Browse...' button and a 'No file selected' status. A 'Download template xlsx file' link is also present. To the right, a separate window shows the XML content of the template file, which defines actions for various DICOM attributes using XSLT transformations.

```
<e en="T" n="StudyDate" t="00080020">@hashdate(this
<e en="T" n="SeriesDate" t="00080021">@hashdate(thi
<e en="T" n="AcquisitionDate" t="00080022">@hashdat
<e en="T" n="ContentDate" t="00080023">@hashdate(th
<e en="T" n="OverlayDate" t="00080024">@hashdate(th
<e en="T" n="CurveDate" t="00080025">@hashdate(this
<e en="T" n="AcquisitionDatetime" t="0008002A">@has
<e en="T" n="StudyTime" t="00080030">@keep()</e>
<e en="T" n="SeriesTime" t="00080031">@keep()</e>
<e en="T" n="AcquisitionTime" t="00080032">@keep()<
<e en="T" n="ContentTime" t="00080033">@keep()</e>
<e en="T" n="OverlayTime" t="00080034">@keep()</e>
<e en="T" n="CurveTime" t="00080035">@keep()</e>
```

Given that the de-identification check process involves the examination of all the images in the directory, this process might require some time to complete, which is proportional to the volume of images. For this purpose, the user can choose whether the check will be conducted on the entire directory or on a portion of data within each folder. The first option will be slower, but it ensures a comprehensive check, whereas the second option will expedite the process.



In the first tab, the user can see in a bar chart the 10 most common issues that are related to the de-identification of the images. In the Y axis the name of the DICOM metadata field is shown whereas the number of images which present the respective issue appears in the X axis. More details on the output of the de-identification check process are provided in a tabular format below. In that table, the exact metadata field in terms of group, element and name, where a problem was identified is mentioned. The current value of each field is mentioned in the column named “value” whereas in the column named “Problem” a short description of the issue is reported. Finally, in the last column, named “filename” the path of the image is mentioned (“filename”).



The user is also able to apply filters on this table, by typing the respective information in the first row of the table

Results of the De-identification Protocol

Show

All

entries

Search:

	group	element	name	value	Problem	filename
	All	All	AccessionNumber	All	All	All
20	0008	0050	AccessionNumber	1611293332707533	This field should have been removed	E:\NCIS\IVE\qualityToolCheck\ColorectalProspective\001-000224\001-000224_CT_BU\Series-1\image-1.dcm
201	0008	0050	AccessionNumber	1397548021897190	This field should have been removed	E:\NCIS\IVE\qualityToolCheck\ColorectalProspective\001-000224\001-000224_CT_TP1\Series-1\image-1.dcm
202	0008	0050	AccessionNumber	2026243997257304	This field should have been removed	E:\NCIS\IVE\qualityToolCheck\ColorectalProspective\001-000224\001-000224_CT_TP2\Series-1\image-1.dcm

If errors appear in this step, the user must communicate with the respective partner and resolve the issue. For example, a user by mistake adds a new DICOM file in a folder without having previously applied any de-identification process and the quality tool reports the issue and prompts the user to correct it.

Results of the De-identification Protocol

Show

All

entries

Search:

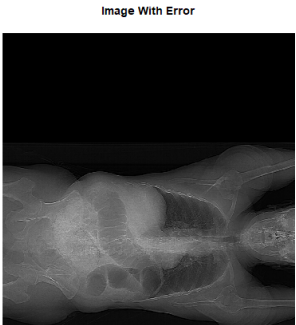
group	element	name	value	Problem	filename
All	All	All	All	All	All
20	0008	0050	AccessionNumber	3059906632743265	This field should have been removed
41	0010	1010	PatientsAge	034Y	This field should have been removed
40	0010	0040	PatientsSex	M	This field should have been removed

If, during the de-identification check, the application is unable to open a DICOM file, the paths of those files are provided in the second tab, indicating that the app was unable to verify the de-identification profile of those files. Thus, this component lists the files which were not able to be checked because of one of the following reasons.

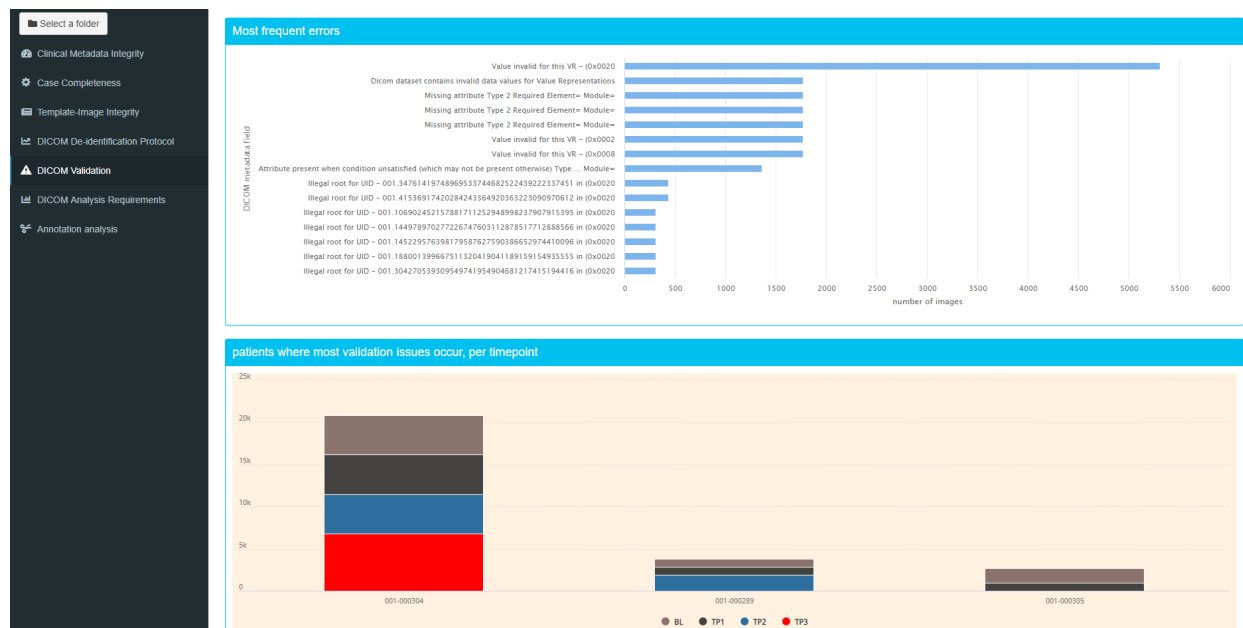
Reasons:

1. The size of the file is greater than 100 MB which implies the DICOM file includes the whole examination and not a single slice (which is a requirement in INCISIVE).
2. The DICOM file is corrupted and not able to be opened by the quality tool. The results are presented in a tabular format.
8. In the third tab, Burned-in Information, the tool checks the DICOM images for burnt-in information. If the image detected is a DICOM report, the image is flagged as 'Report' and needs to be removed. If the image detected is an actual DICOM image, the message 'Replace image' appears. The image should be replaced with a cleaned image placed in a folder named 'Temporary Folder' along with the rest of the data and the full path of the cleaned image is reported in the last column. The user, by selecting each of the rows in the table, can check the original image as well as the image after cleaning the burned in information.

Burned-In analysis results					
Show All entries		Search:			
patientID	Image Type	Path To Image With Error	Action	Path To Cleaned Image	
All	All	All	All	All	
1	001-000376	Report	E:/AAA_testSet/001-000411/001-000411_CT_BL/Series-1/Image-1.dcm	Replace Image	E:/AAA_testSet/001-000411/001-000411_CT_BL/Series-1/Image-1.dcm
2	001-000376	Report	/home/cheetara/Documents/inc/StorageProspClean/001-000376/001-000376_MR_BL/Series-999/Image-1.dcm	Remove Image	
Showing 1 to 2 of 2 entries					Previous 1 Next



9. The user then must click on the fifth component named DICOM Validation. This component checks whether the DICOM files are valid, in terms of missing any attributes or have any unexpected ones. The component is based on <https://dclunie.com/dicom3tools/dciodvfy.html>. The check is done by comparing the contents of the DICOM file against the modules and attributes required by the DICOM standard for the SOP class of the given dataset. This component does not check for consistency between multiple files (e.g., consistent values of Study or Series level attributes).
- The most common validation problems are depicted in a bar-chart (see figure below-upper) whereas in the bottom part of the figure below, the information of the validation issues is depicted for each patient, where the majority of the issues observed, for each timepoint. The user can interact with the chart and filter the visualization.



Finally, details are presented in a tabular format, and the user can filter the data based on the value of each column.

Details of the Validation Analysis

Show entries

Search:

	patientID	Modality	timepoint	series	filename	Problem
	All	All	All	All	All	All
1	001-000289	CT	BL	Series-1	Image-1.dcm	Illegal root for UID - 001.330364508950071549581689238347378176393 in (0x0002)
2	001-000289	CT	BL	Series-1	Image-1.dcm	Illegal root for UID - 001.330364508950071549581689238347378176393 in (0x0008)
3	001-000289	CT	BL	Series-1	Image-1.dcm	Illegal root for UID - 001.222061435994763281120083790630234607457 in (0x0020)
4	001-000289	CT	BL	Series-1	Image-1.dcm	Illegal root for UID - 001.266231414235415653149698118407666992619 in (0x0020)
5	001-000289	CT	BL	Series-1	Image-1.dcm	Illegal root for UID - 001.187667471752216542637444364011772581515 in (0x0020)
6	001-000289	CT	BL	Series-1	Image-1.dcm	Value invalid for this VR - (0x0002)
7	001-000289	CT	BL	Series-1	Image-1.dcm	Value invalid for this VR - (0x0008)
8	001-000289	CT	BL	Series-1	Image-1.dcm	Value invalid for this VR - (0x0020)
9	001-000289	CT	BL	Series-1	Image-1.dcm	Value invalid for this VR - (0x0020)
10	001-000289	CT	BL	Series-1	Image-1.dcm	Value invalid for this VR - (0x0020)
11	001-000289	CT	BL	Series-1	Image-1.dcm	Dicom dataset contains invalid data values for Value Representations
12	001-000289	CT	BL	Series-1	Image-1.dcm	Missing attribute Type 2 Required Element=<PatientBirthDate> Module=<Patient>
13	001-000289	CT	BL	Series-1	Image-1.dcm	Missing attribute Type 2 Required Element=<PatientSex> Module=<Patient>
14	001-000289	CT	BL	Series-1	Image-1.dcm	Missing attribute Type 2 Required Element=<ReferringPhysicianName> Module=<GeneralStudy>
15	001-000289	CT	BL	Series-2	Image-389.dcm	Illegal negative value - SpacingBetweenSlices = -1.5
16	001-000289	CT	BL	Series-2	Image-389.dcm	Illegal root for UID - 001.224603875142332890472033943802068065321 in (0x0002)
17	001-000289	CT	BL	Series-2	Image-389.dcm	Illegal root for UID - 001.224603875142332890472033943802068065321 in (0x0008)
18	001-000289	CT	BL	Series-2	Image-389.dcm	Illegal root for UID - 001.222061435994763281120083790630234607457 in (0x0020)

The validation of the DICOM files may be crucial for some applications, which do not operate correctly if an invalid DICOM file is provided. However, it is worth mentioning here, that the user is not able to perform any corrective actions in this step, since the editing of the DICOM tags is performed by the de-identification tool. This functionality was included to provide the user with a validation report in

those cases where some applications do not provide the expected results because of the provision of invalid DICOM files.

10. The user must click on the sixth component named *DICOM Analysis Requirements*. This component checks whether some analysis requirements are met. In INCISIVE such requirements were defined for the thickness of the slices, or the availability of the information related to the pulse sequence use. The availability of this information allows the technical developers to use data of expected quality for analysis and the training of the algorithms.

INCISIVE - quality check

Select a folder

Clinical Metadata Integrity

Case Completeness

Template-Image Consistency

DICOM De-identification Protocol

DICOM Validation

DICOM Analysis Requirements

DICOM Overall Patient Evaluation

Image requirements

Annotations

This component checks whether some requirements are met with the goal of allowing the technical developers to use data of expected quality for analysis and the training of the algorithms

Results of the DICOM Analysis Requirements

Show: All ▾ entries

group	element	name	value	Problem	filename
1	0008	103E	SeriesDescription	Not able to check, field is missing from file	E:\INCISIVE\quality\toolCheck\ColorDial\respective\001-00024\001-00024_MR_T2\Series:10\image-1.dcm
2	0008	103E	SeriesDescription	Not able to check, field is missing from file	E:\INCISIVE\quality\toolCheck\ColorDial\respective\001-00024\001-00024_MR_T2\Series:10\image-10.dcm
3	0008	103E	SeriesDescription	Not able to check, field is missing from file	E:\INCISIVE\quality\toolCheck\ColorDial\respective\001-00024\001-00024_MR_T2\Series:10\image-11.dcm
4	0008	103E	SeriesDescription	Not able to check, field is missing from file	E:\INCISIVE\quality\toolCheck\ColorDial\respective\001-00024\001-00024_MR_T2\Series:10\image-2.dcm

Showing 1 to 1 of 1 entries (filtered from 685 total entries)

Previous 1 Next

Designed & Implemented by Laboratory of Computing, Medical Informatics and Biomedical - Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Greece

INCISIVE - quality check

Select a folder

Clinical Metadata Integrity

Case Completeness

Template-Image Integrity

DICOM De-identification Protocol

DICOM Validation

DICOM Analysis Requirements

Annotation analysis

Image requirements

Annotations

This component checks whether some requirements are met with the goal of allowing the technical developers to use data of expected quality for analysis and the training of the algorithms

Results of the DICOM Analysis Requirements

Show: All ▾ entries

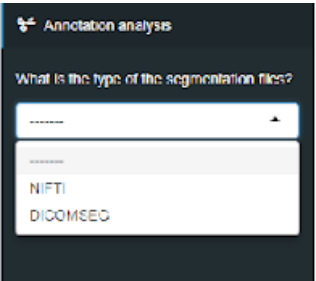
group	element	name	value	Problem	filename
885	0018	0050	SliceThickness	The value of this field is higher than expected	E:\AAAA_quality\test\001-000305\001-000305_FuacT_BL\Series-1\image-1.dcm

Showing 1 to 1 of 1 entries (filtered from 685 total entries)

Previous 1 Next

Designed & Implemented by Laboratory of Computing, Medical Informatics and Biomedical - Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Greece

11. The final component is the “annotation analysis”. This component examines the annotations provided for the images. The user can select the type of annotation files that are expected in the dataset (NIFTI or DICOMSEG). The result of the annotation analysis differs according to the type of the files.



12. In the case of the nifty files the following table is produced which lists the annotation files found in the specified path. For quality assurance of the annotation files, the component verifies whether each annotation file is located in the expected folder, specifically within the series folder. Additionally, if more than one annotation file is present in the folder, the corresponding

rows are highlighted in orange. Furthermore, each annotation file must include specific labels corresponding to the cancer type and imaging modality. The component checks for the presence of correct annotation labels and flags any invalid labels in the files by highlighting them in red. Moreover, the component ensures consistency between the Image Region of Interest (ROI) and the ROI provided by the annotation, as well as consistency in the number of images and their respective annotation slices.

Finally, the component verifies the agreement between the provided files and the information outlined in the template file. If a cell is highlighted in red a specific action must be made to ensure consistency.

1. If a cell in the series column is marked. This means that the annotation file was not placed inside a series folder and thus it is not possible to make the association between the annotation and the image folder.

2. If a cell in the inconsistency column is marked in red, this means that the annotation file does not match the images, and the user must remake the annotation. If ROIs inconsistency is written, then the resolution between the annotation mask and the images does not match. If slices inconsistency is marked, then the number of slices does not match.

Annotation Consistency
Annotation Details



This component examines the annotations provided for the images. The table below lists the annotation files found in the specified path. For quality assurance of the annotation files, the component verifies whether each annotation file is located in the expected folder, specifically within the series folder. Additionally, if more than one annotation file is present in the folder, the corresponding rows are highlighted in orange. Furthermore, each annotation file must include specific labels corresponding to the cancer type and imaging modality. The component checks for the presence of correct annotation labels and flags any invalid labels in the files by highlighting them in red. Moreover, the component ensures consistency between the Image Region of Interest (ROI) and the ROI provided by the annotation, as well as consistency in the number of images and their respective annotation slices. Finally, the component verifies the agreement between the provided files and the information outlined in the template file. If a cell is highlighted in red a specific action must be made to ensure consistency.

Actions:-

1. If a cell in the series column is marked. This means that the annotation file was not placed inside a series folder and thus it is not possible to make the association between the annotation and the image folder.
2. If a cell in the inconsistency column is marked in red, this means that the annotation file does not match the images and the user must remake the annotation. If ROIs inconsistency is written then the resolution between the annotation mask and the images does not match. If slices inconsistency is marked, then the number of slices do not match.

Annotation Files analysis results

Show
All
entries

Search:

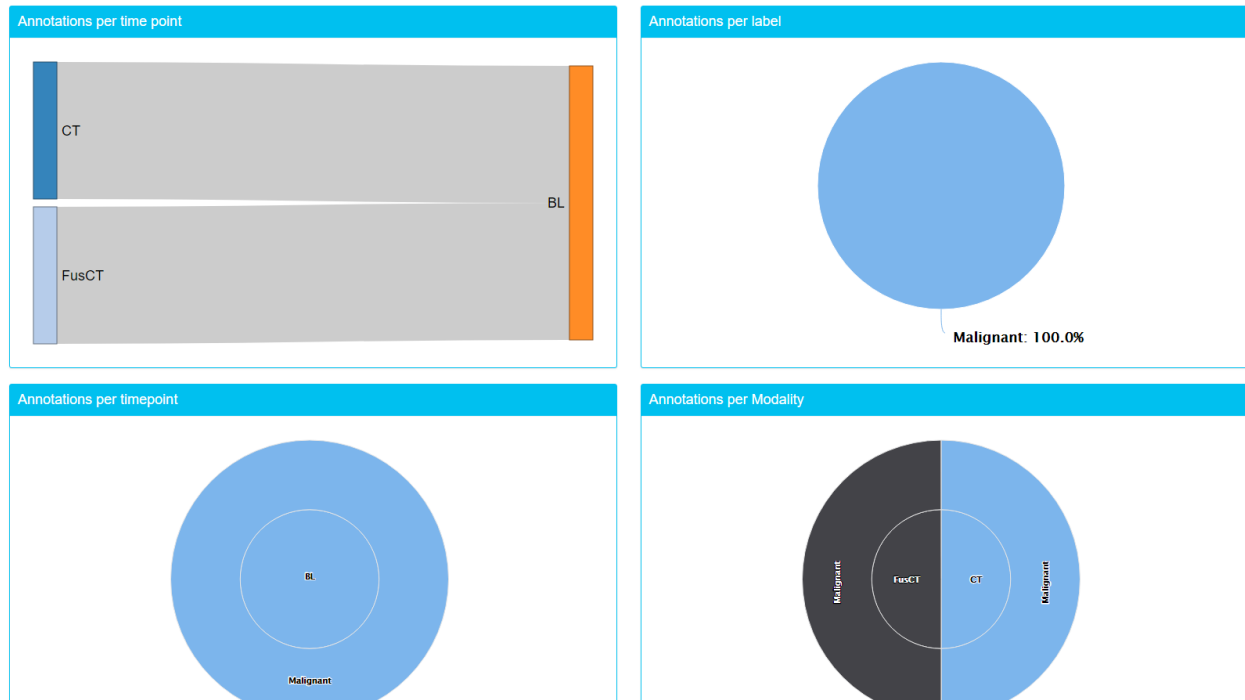
	patientID	timepoint	series	filename	inconsistency	labels	annotationRows	annotationColumns	annotationSlices
	All	All	All	All	All	All	All	All	All
1	001-000289	001-000289_CT_BL	Series-2	Image-1.nii.gz	ROI inconsistency		512	512	443
2	001-000304	001-000304_MR_BL	Series-902	Image-1.nii.gz	ROI inconsistency		400	400	32
3	001-000305	001-000305_FusCT_BL	Series-3	Image-1.nii.gz	ROI inconsistency		512	512	343

Showing 1 to 3 of 3 entries

Previous
1
Next

Designed & Implemented by Laboratory of Computing, Medical Informatics and Biomedical - Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Greece

In the second tab, named “Annotation details”, the component examines the annotations provided by the user. The images provide more details on the provided annotation files, but they do not include any information regarding the consistency of the files



The user needs to go back to the data and place the respective file into the correct folders.

Then the user needs to run the pipeline again to ensure that all errors are corrected.

In case the annotation files are in DICOMSEG format, the information that is stored in the files metadata differs and thus the results that are provided to the user differ, too. The association between the imaging data and the annotations is based on StudyInstanceUID attribute and thus in the table below the folder with the respective imaging data is associated with one or more DICOMSEG files that exist in the repository. In addition, the last column of the table checks the dimensions of both the images and the masks in the X and Y direction to check if they coincide. However, it is worth mentioning that even if the dimensions are not identical, the mask is still useful for analysis.

Association of images with DICOMSEG files

Show

All

 entries

Search:

ParentFolder	StudyInstanceUID	DICOMFiles_DifferentUID	SEGFiles	DimensionCheck
All	All	All	All	All
1	E:/INCISIVE/ProstateX-0004/study1/10.000000-tfidynfastra1.5x1.5l3.5sec-83832	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
2	E:/INCISIVE/ProstateX-0004/study1/11.000000-tfidynfastra1.5x1.5l3.5sec-10660	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
3	E:/INCISIVE/ProstateX-0004/study1/12.000000-tfidynfastra1.5x1.5l3.5sec-14427	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
4	E:/INCISIVE/ProstateX-0004/study1/13.000000-tfidynfastra1.5x1.5l3.5sec-45920	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
5	E:/INCISIVE/ProstateX-0004/study1/14.000000-tfidynfastra1.5x1.5l3.5sec-75847	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
6	E:/INCISIVE/ProstateX-0004/study1/15.000000-tfidynfastra1.5x1.5l3.5sec-65592	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm	Different
7	E:/INCISIVE/ProstateX-0004/study1/16.000000-tfidynfastra1.5x1.5l3.5sec-85576	1.3.6.1.4.1.14519.5.2.1.7311.5101.170561193612723093192571245493	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-3.549/1-1.dcm; E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcm; E:/INCISIVE/ProstateX-	Different

In addition, the tool checks also the existence of duplicated annotation files and if such cases are identified, they are provided in a tabular format as shown in the image below.

Annotation Image Consistency

Duplicated DICOMSEG

DICOMSEG tag

Results of Dublication check analysis

Show

All

 entries

Search:

File	Duplicates
All	All
1	E:/INCISIVE/ProstateX-0004/study1/300.000000-Segmentation-7.701/1-1.dcmE:/INCISIVE/ProstateX-0004/study1/duplicated_segfile/1-1.dcm

Showing 1 to 1 of 1 entries

Previous1Next

For the data utility and based on the requirements set, a table is provided, informing the user whether each patient includes data that fulfils the requirements, e.g. for the slice thickness and the pulse sequence (in case MRI is needed).

image requirements

Overall Evaluation

This component checks if there are duplicate images in the dataset and if some prerequisites for the analysis are present, modalities in each cancer type, slice thickness and pulse sequence

Results of the DICOM overall patient evaluation

Show

All

entries

Search:

patientID	modalities	sliceThickness	PulseSequence	
All	All	All	All	
1	001-000289	at least one required imaging modalities are provided	Slice Thickness requirement is covered	No MR data are provided
2	001-000304	at least one required imaging modalities are provided	No requirement for the provided modalities	MR data are provided and required pulse sequences are not found
3	001-000305	none of the required imaging modalities are provided	No requirement for the provided modalities	No MR data are provided

Showing 1 to 3 of 3 entries

Previous

1

Next

Designed & Implemented by Laboratory of Computing, Medical Informatics and Biomedical - Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Greece

Finally, the tool performs a validation of DICOM SEG files by checking conformance with the attribute requirements of the DICOM standard and the minimum annotation metadata defined by the project. This validation ensures that the essential tags are present, correctly formatted and semantically consistent.

The results are presented in a structured table format, highlighting for each patient whether the segmentation files comply with the defined standards. Each row corresponds to a specific DICOM SEG file containing all missing or non-compliant attributes as well as identifiers such as patient ID, series ID and study ID for traceability.

Annotation Image Consistency

Duplicated DICOMSEG

DICOM SEG validator



Validates DICOM SEG files for compliance with DICOM standards and project-specific metadata requirements. Reports missing, invalid or non-compliant attributes to support data quality and interoperability

Results of DICOM SEG validation

Show

All

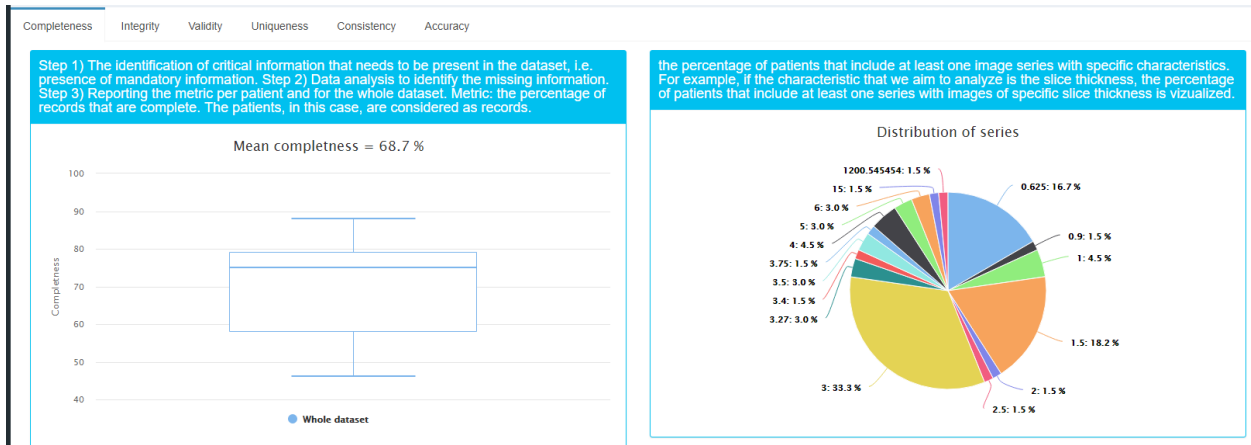
entries

Search:

	PID	SeriesID	Study	Issue	Path
	All	All	All	All	All
1	LUNG1-006	1.2.276.0.7230010.3.1.3.2323910823.21456.1597260540.379		Compliant	/content/drive/MyDrive/EUCAIM/annotator/input/300.000000-Segmentation-0.379/1-1.dcm
2	LUNG1-006	1.2.276.0.7230010.3.1.3.2323910823.21456.1597260540.379		Tag (0020,9157) is required but not present or empty.	/content/drive/MyDrive/EUCAIM/annotator/input/300.000000-Segmentation-0.379/1-1.dcm
3	LUNG1-006	1.2.276.0.7230010.3.1.3.2323910823.21456.1597260540.379		Tag (0020,9167) is required but not	/content/drive/MyDrive/EUCAIM/annotator/input/300.000000-

13. Metrics

In this ta



The definition of the integration rules, which in this case is the link between images and clinical metadata through the template. Metric: The percentage of records that is properly integrated. At this case, records are considered the modalities provided in the dataset.

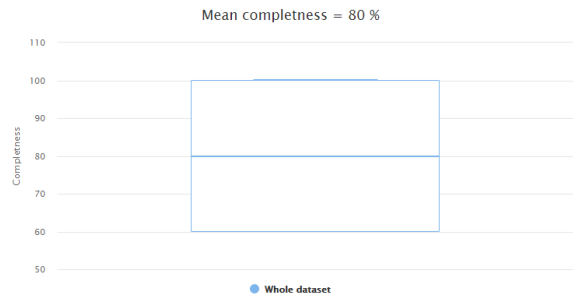
Show **All** entries Search:

Subset Selection

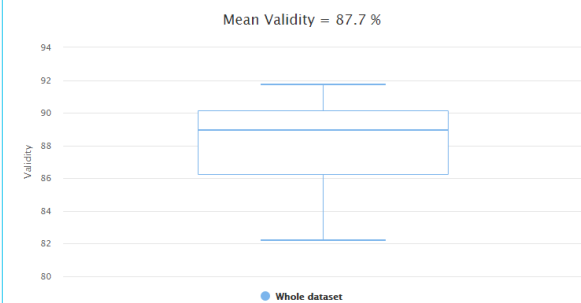
	patientID	errors	total_checks	integrity_percentage
	All	All	All	All
1	001-000003	2	5	60
2	001-000034	0	4	100

Showing 1 to 2 of 2 entries Previous **1** Next

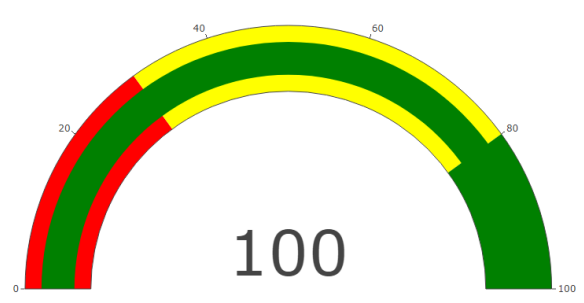
The percentage of series in which the segmentation mask is relevant



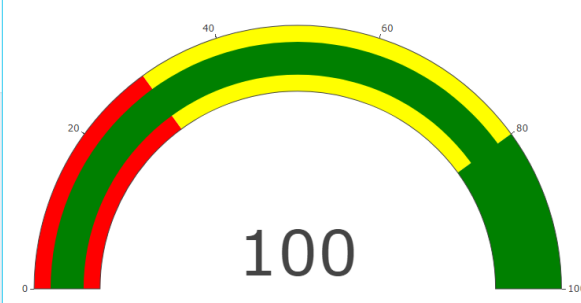
The percentage of records in which all values are valid.



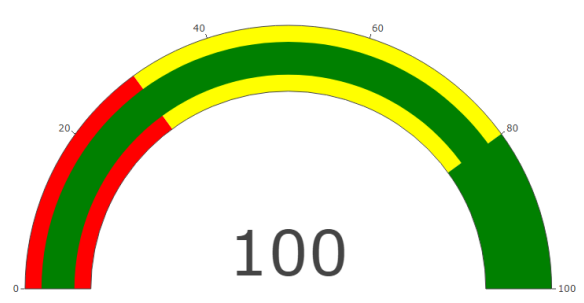
Correctly de-identified images (%)



Correctly de-identified series (%)

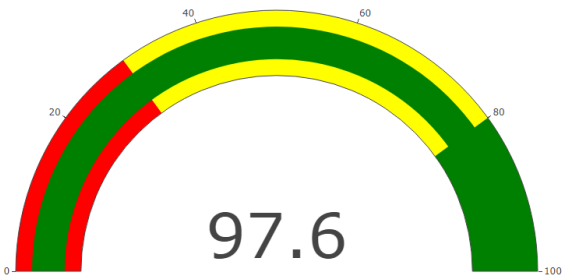


Correctly de-identified patients (%)



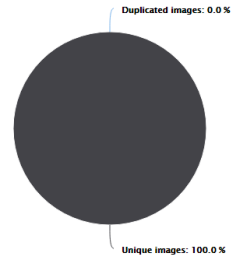
Completeness Integrity Validity **Uniqueness** Consistency Accuracy

The absolute number of duplicate records identified for the dataset. At this case, a record is considered a patient entry in the dataset.



The total number of identified duplicate image files or series within the dataset.

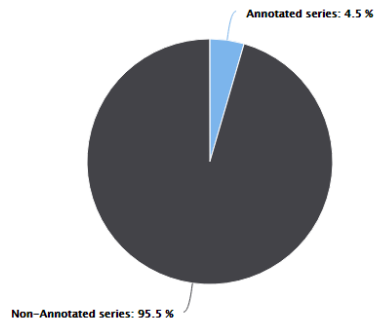
Percentage of duplicated images = 0 %



Completeness Integrity Validity **Uniqueness** Consistency Accuracy

Percentage of correctly Annotated and non-annotated series

Percentage of annotated series = 4.5 %



Completeness Integrity Validity Uniqueness **Consistency** Accuracy

This metric measures how well the data structure conforms to a specific template (e.g. sheets and columns in an excel file, columns in a csv file, etc.), mainly focusing on semi-structured data, or cases where the structure is not pre-imposed. Metric: The absolute number of inaccuracies in the examined file.

