




A European Cancer Image Platform Linked to Biological and Health Data for Next-Generation Artificial Intelligence and Precision Medicine in Oncology

Deliverable D4.2: Cloud-based collaborative data annotation tool

Reference	D4.2_EuCanImage_CM RAD_v1
Lead Beneficiary	CMRAD
Author(s)	Mathias Engström (CM), Anders Nordell (CM), Pär Kragsterman (CM)
Dissemination level	Public
Type	Other / Cloud system
Official Delivery Date	2022-03-31
Date of validation of the WP leader	2022-03-30
Date of validation by the Project Coordinator	2022-03-31
Project Coordinator Signature	

EuCanImage is funded by the European Union's H2020 Framework
Under Grant Agreement No 952103

Version log

Issue Date	Version	Involved	Comments
2022-03-20	V0.1	Mathias Engström	First draft
2022-03-29	V0.2	Fred Prior, Karim Lekadir	Review and Feedback
2022-03-29	V0.3	Mathias Engström	Feedback incorporated.
2022-03-30	V0.4	Fred Prior, Karim Lekadir	Final Review
2022-03-31	V1	Isabell Tributsch, Karim Lekadir	Revised and corrected final version.

Executive Summary

A centralised, cloud-based collaboration solution for annotation, segmentation and classification of imaging data has been designed, developed and deployed with the purpose of enhancing EuCanImage data with additional input from clinical experts.

Each use case has its own separate workflow path in the system, providing customization and adaptation based on its requirements. Users are given access to the system based only on the tasks they need to complete. Currently three different tools to perform annotation and segmentation are available. A final decision has not been reached regarding the final tool and workflow for the colorectal MRI and colorectal CT use case. For the use cases covering live MRI, liver CT, breast MRI and mammography tools and workflow has reached a mature state.

Analyses and annotations are saved in a common standard, DICOM-compatible, format for easy access and use for AI developers.

Table of Contents

1	Introduction	4
1.1	Data flow	4
1.3	Upload procedure	6
1.4	Annotation and Segmentation	7
1.5	Export to EuroBioImaging / XNAT	8
2	Annotation and segmentation tools	8
2.1	Bounding box	8
2.2	Free Draw	9
2.3	SmartPaint	10
3	Use cases	13
3.1	Use case 1: Liver MRI	13
3.2	Use case 2: Liver CT	13
3.3	Use case 3: Colorectal CT	14
3.4	Use case 4,5: Colorectal MRI	14
3.5	Use case 7: Breast MRI	14
3.6	Use case 6,8: Mammography	14
4	Conclusion	15

Acronyms

Name	Abbreviation
Apparent Diffusion Coefficient	ADC
Artificial Intelligence	AI
Application Programming Interface	API
Collective Minds Radiology	CMRAD
Digital Imaging and Communications in Medicine	DICOM
Hypertext Transfer Protocol Secure	HTTPS
Perl Open Source DICOM Archive	POSDA
Extensible Neuroimaging Archive Toolkit	XNAT
Radiotherapy Structure Set	RTSTRUCT
Secure Hash Algorithms	SHA

1 Introduction

As part of the EuCanImage project Collective Minds Radiology will provide a centralised data annotation and segmentation solution. This will consist of a cloud-based medical research collaboration system with crowdsourced collaboration capabilities.

1.1 Data flow

The EuCanImage project includes clinical partners who share data. Pseudonymization of the data is performed at the clinical partners and the pseudonymized copy of the data is transferred to the central storage. The central storage consists of two repositories, the Euro-BioImaging archive and the Collective Minds Radiology central storage. The Collective Minds Radiology repository is used to annotate and segment imaging data. This data enhancement work is done by EUCanImage approved healthcare professionals in a collaborative fashion. The EuroBioImaging repository is the long term storage and access point for researchers and AI-developers.

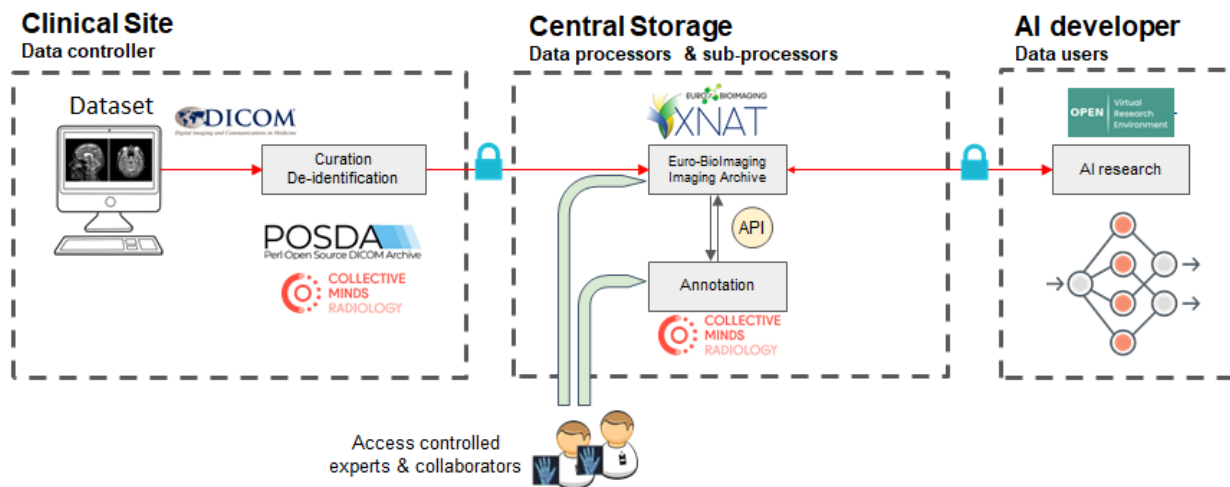


Figure 1 EuCanImage Data Flow

Data can be injected into Euro-BioImaging or Collective Minds Radiology and the two repositories communicate using an API.

1.2 EUCanImage Annotation Platform

In a highly controlled and regulated environment, Collective Minds Radiology provides a collaboration solution where data can be curated, annotated, segmented, and analysed. By using a well-established pseudonymization and de-identification method, data sharing can be performed both retrospectively and prospectively for all project participants. The solution is based on a subject centric approach where data layers are wrapped around the subject allowing for multi-modal data contribution.

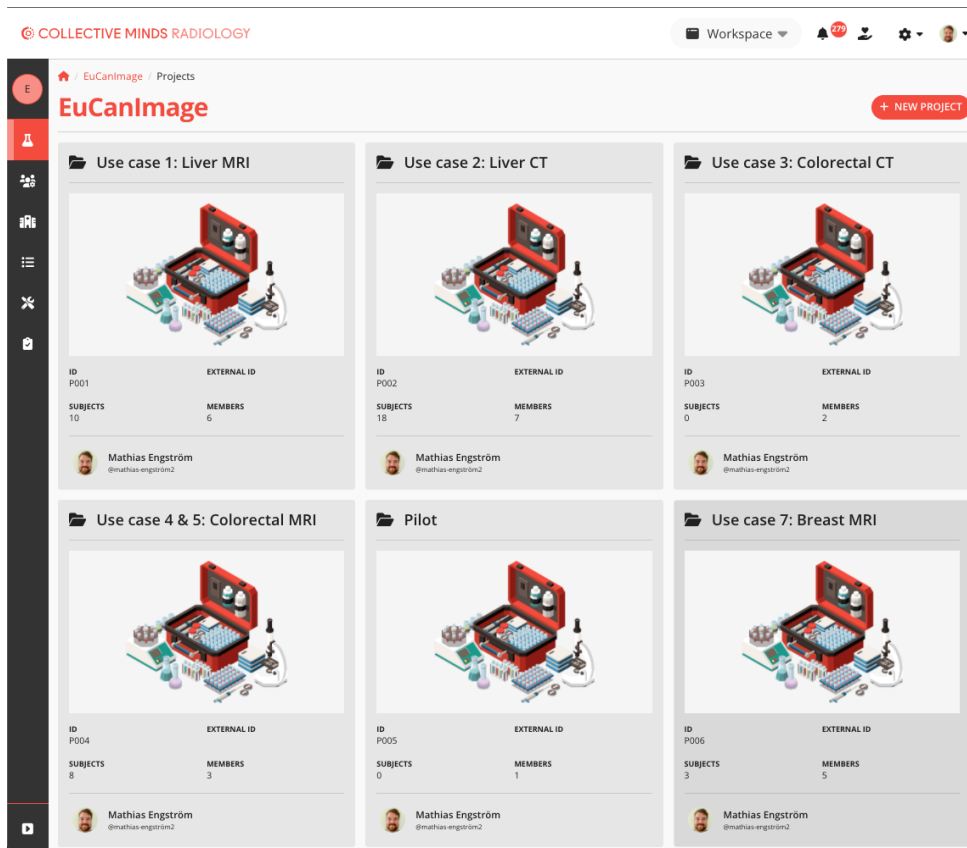


Figure 2 The EuCanImage repository where each use case is shown as its own project.

The EuCanImage consortium will be using a secluded repository hosting a set of projects that reflect the multiple use cases that constitutes the annotation and segmentation portion of the project. Each project has strict access control and a custom processing chain of events, referred to as a pipeline. All participants will be given a role and affiliation that dictates and controls the data access that is required. The role and site affiliation will also control what tasks that can be performed by individual users. The system provides a full audit trail of all actions and provides traceability via digital signatures. The solution also enforces strict forward versioning, which enables multiple executions of tasks with backward traceability.

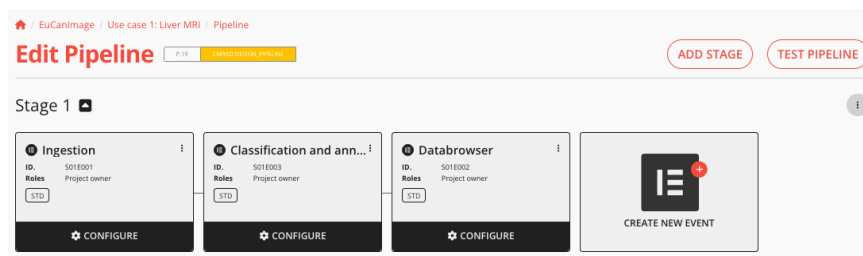


Figure 3 Example of a pipeline within a project where the progress is indicated per subject.

Each use case will have a common pipeline design including an ingestion, annotation/segmentation, and a data review event. It should be noted, however, that annotation, segmentation and structured reporting tasks will be tailored individually to each use case, as discussed in section 3 of this document. Active radiologists will be provided with a list of tasks to perform and not be exposed to the full structure of the pipeline design. Embedded in the system is a control mechanism that guarantees that only ready to run tasks are being displayed, blocking users from working on the same task simultaneously.

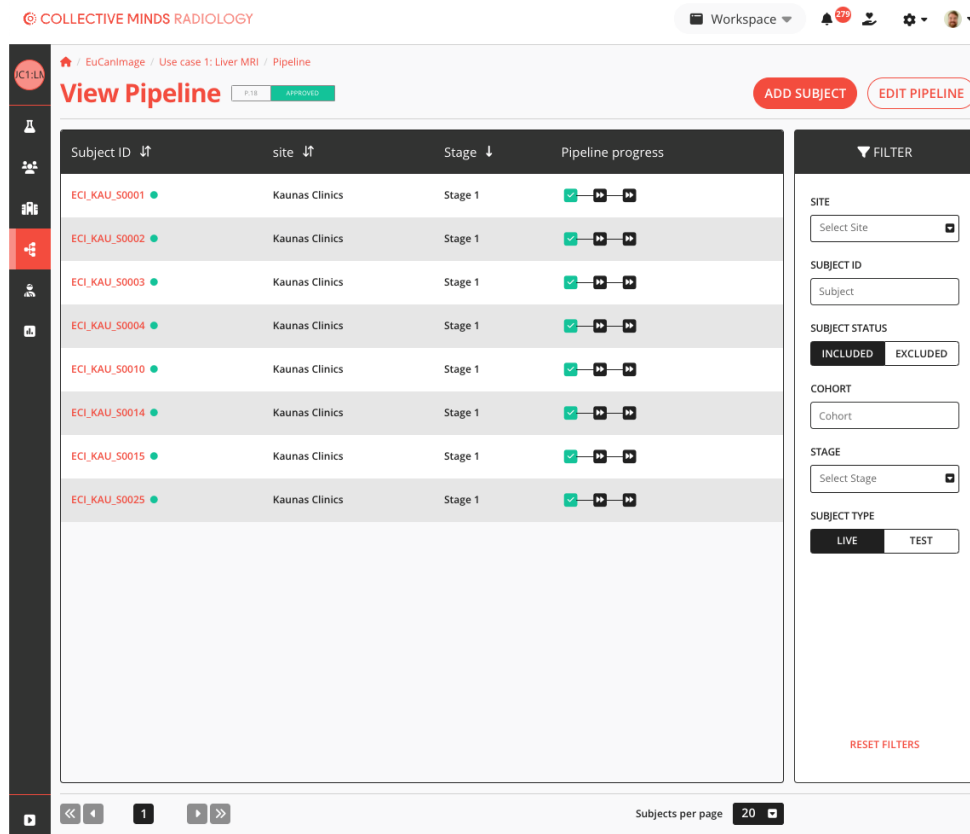


Figure 4 Example of a pipeline design where each task is clearly visible.

1.3 Upload procedure

All participant sites will be provided with a dedicated lightweight virtual server, hereafter referred to as a proxy, for transfer of DICOM imaging data between the site and Collective Minds Radiology. Effectively allowing mass ingestion of pseudonymized and de-identified local data into the repository. Pseudonymization and de-identification is performed within the proxy and is guaranteed unidirectional using industry standard hashing procedures (salted SHA512/256). A unique hash secret will also be assigned to each site installation, further strengthening the pseudonymization process.

Due to variations in patient tracking between collaboration participants, as well as the high impracticality of editing DICOM metadata before submitting, Collective Minds Radiology provides a pseudonymization verification tool that enables the linkage between the original DICOM Patient ID, Collective Minds Radiology hash and EuCanImage ID. By using this method,

it is possible for participants to keep track of the data contributions effectively. Please note that Collective Minds Radiology never will know, be aware, store or keep track of the local patient identifiers, only the hash and the EuCanImage ID.

Upon request Collective Minds Radiology will provide an end-to-end encrypted batch upload option using the secure copy protocol, following the secure shell mechanisms and with a public/private key procedure. Please note that this option is only available following that data has been locally pseudonymized using the consortium agreed upon pseudonymization and de-identification procedure.

1.4 Annotation and Segmentation

Annotation and segmentation work will be performed within the Collective Minds Radiology platform. For each subject, in each use case, a segmentation or annotation task will be prepared as soon as data is ingested. The task will be readily available to all users with the necessary role permissions and can be started and executed by whoever is invited to the repository with such permissions.

After starting a task, the user will be presented with a data browser that includes a list of all series of DICOM data that has been submitted. In addition a custom report form will be provided that is specific to the use case and the suggested workflow. More information about these workflows can be found in section 3. All DICOM images will be shown in an embedded medical grade zero-footprint DICOM viewer, accessible from the data browser window. The DICOM viewer will host the segmentation and annotation tools which will be available according to use case specifications.

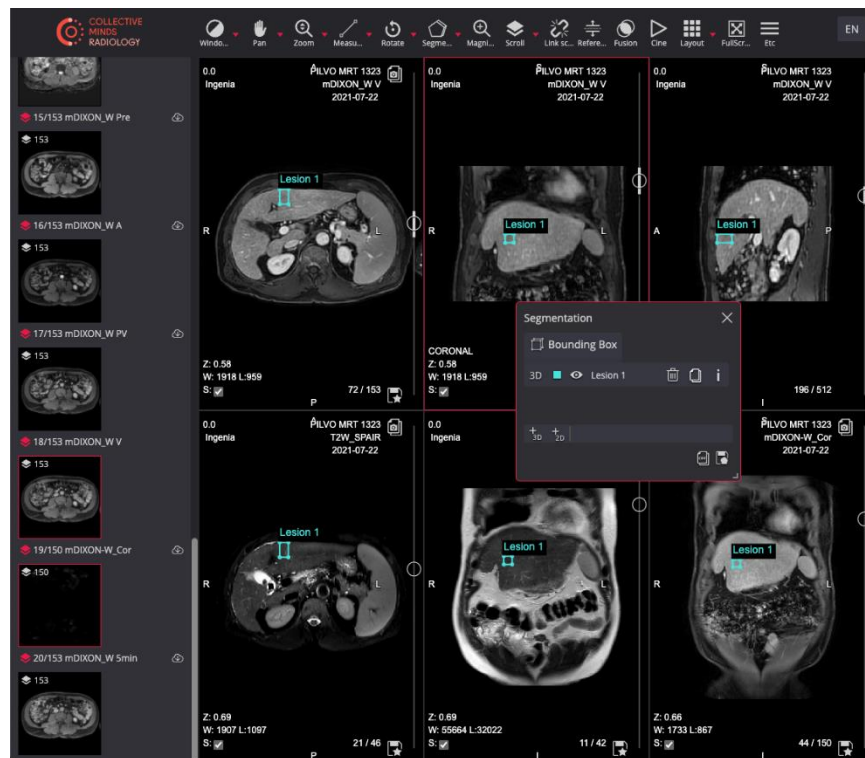


Figure 5 Example of the embedded viewer with a Bound Box annotation.

As soon as a segmentation or annotation task has been completed, the data is saved and a confirmation message will be sent to the user. A continuous saving mechanism is used in order to store intermediate results of the segmentation process in case the segmentation task has been terminated and then need to be reattempted at a later time or date. After the annotation or segmentation has been completed, saved, and the custom report form filled out, the event is electronically signed, indicating that it is ready for export. Previously signed tasks can not be altered, as part of the forward versioning concept, but subject to review for tracking deltas. A previously signed task can however be re-run in order to update a segmentation. In the scenario of a re-run, a new version number will be provided indicating the update.

The annotations and segmentations associated with any DICOM study reside within the DICOM study itself, and there is no requirement for a proprietary file format or external file management.

The segmentation and annotation workflows for each use case have been established by the EuCanImage dataflow and annotation workgroup together with the clinical use case workgroups. Each use case workgroup is responsible for providing the training and user manual material. Local super users have been identified from each participant site for training and will provide information about procedures to their colleagues. Platform onboarding and user account creation is facilitated by Collective Minds Radiology. Repository, project and data access will be assigned in agreement with the consortium requirements and demands. Technical and platform support is provided by Collective Minds Radiology.

1.5 Export to Euro-BioImaging / XNAT

Collective Minds Radiology and EuroBioImaging provides an API towards XNAT for export and synchronisation of completed segmentation and annotation tasks. This API will only be accessible to consortium agreed upon members central to driving the forward propagation of curated data. The data transfer is encrypted using HTTPS and as a result of exposing the API, the data transfer is automated, which will further improve the efficiency of the consortium's data flow.

2 Annotation and segmentation tools

As part of the Collective Minds Radiology platform, three tools for annotation and segmentation work are available to EuCanImage: BoundingBox, FreeDraw, and SmartPaint. As part of the centralised platform, all tools are included in the zero footprint DICOM viewer.

All segmentations and annotations are saved in the format of radiotherapy structure set objects (DICOM-RTSTRUCT). Among the benefits of this approach are version control, non-proprietary read and write capabilities, as well as broad accessibility and inspection. Additionally, all segmentations and annotations are saved and managed along with the DICOM image data in a standard DICOM database without the need for any external storage.

The DICOM-RTSTRUCT format also provides a patient space mapping of all segmentations and annotations which facilitates the use across series within a study.

2.1 Bounding box

A Bounding box is a rectilinear selection, which can be either 2-Dimensional (single slice) or 3D-dimensional (multiple slices) in shape. The Bounding box can be added in any orientation

or plane. This application is intended for fast annotation work and for quickly capturing areas and volumes of interest. The activation of a BoundingBox annotation will, by default, initiate a three plane multi-planar overview mode as a reference to the annotation. BoundingBoxes are rendered in all active viewports where spatially overlapping data from the same study is available, if there is a need for reference series, e.g. for detection purposes.

A new Bounding Box can be added to a study by clicking on a point of interest and dragging the mouse outward to set an initial size. Once the initial size has been set all sides of the Bounding Box can be individually adjusted.

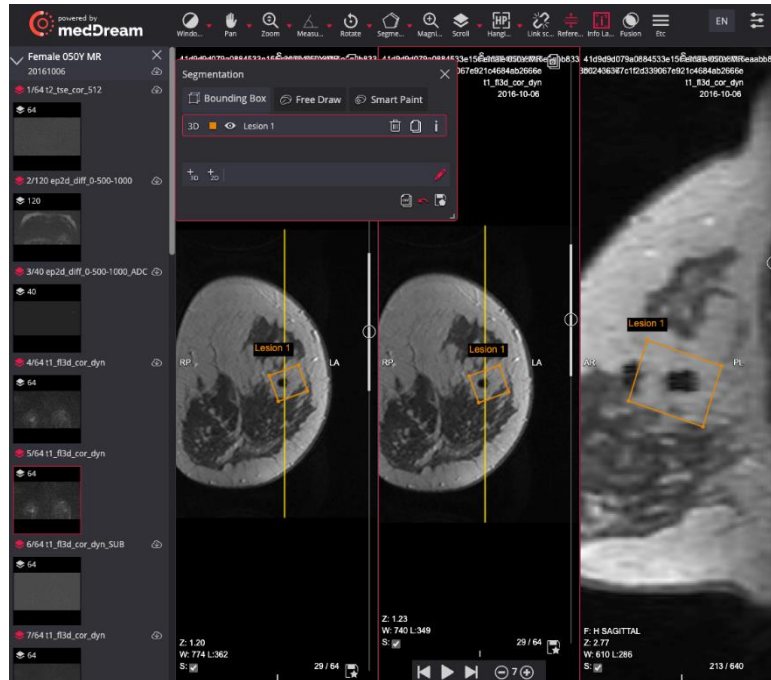


Figure 6 Example of a Bounding box annotation.

By selecting the centre of a Bounding Box and moving it, a translation is accomplished. Selecting the corner of a Bounding Box and pressing the "alt"-key allows the box to be rotated. Translations are performed in the in-plane direction of the scan plane that the Bounding Box is currently rendered on. Rotations are performed around the centre of mass of the Bounding Box along a vector that is defined by the cross-product of the vectors that span the plane it is currently rendered on. Changing the size, position, or rotation of a Bounding Box will directly affect all viewports that are currently active.

A user may select both the colour and a free text label when adding a Bounding Box to a study. The colour as well as the label is saved as properties of the DICOM-RTSTRUCT object, along with the annotation location. It is not necessary to use a unique label to separate Bounding Boxes and they can be spatially overlapping.

2.2 Free Draw

FreeDraw is an interactive tool that can be used to accurately segment and delineate lesions and structures with high precision in both 2D and 3D. A semi-pixel perfect technique is utilised to allow users to draw their desired regions by following the contours of objects or by defining their own based on their preferences. With this tool, the user is able to generate an annotation

gradually by adding and removing areas of interest incrementally, which is an ideal way for depicting complex anatomy accurately. There is also an option to hide any FreeDraw annotation on the image in order to get a clear view of the underlying anatomy.

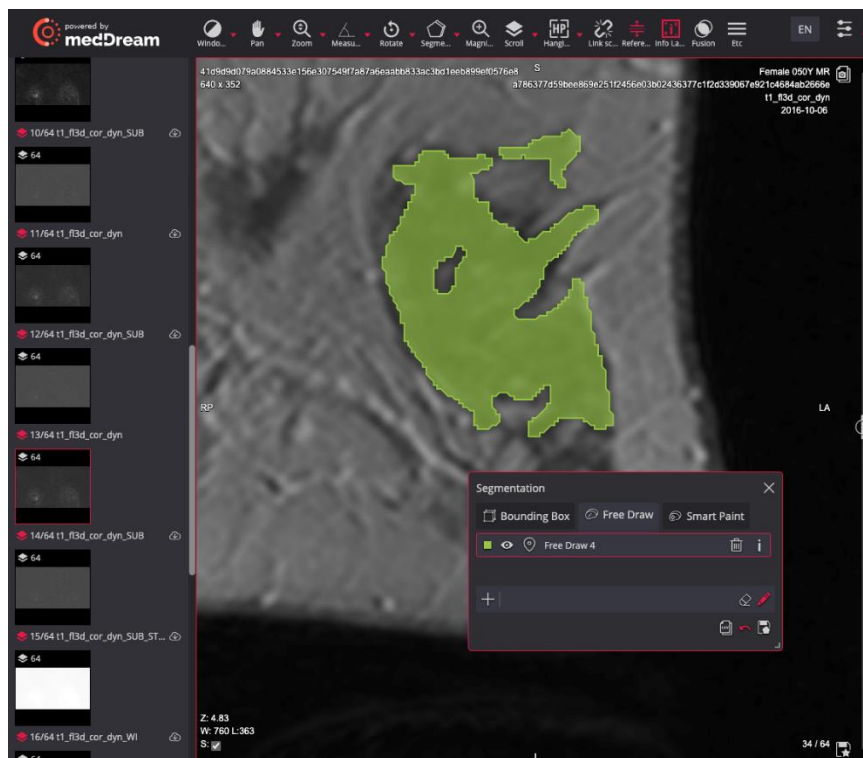


Figure 7 Example of a Free Draw segmentation.

As with the Bounding Boxes, multiple Free Draw annotations can be added to a series for capturing all the structures that are desired. Essentially, the annotator works through a series, slice by slice, to generate the complete segmentation. This technique lends itself well to both mouse and keyboard as well as stylus and drawing pad input devices.

Free Draw segmentations are saved using a closed planar geometric contour type in DICOM-RTSTRUCT where areas, holes and islands are captured using boolean operators. Post-processing reconstruction can be carried out using readily available third-party libraries for DICOM-RTSTRUCT reconstruction. As a FreeDraw segmentation is based on slices, the slice direction resolution of the FreeDraw segmentation will be determined by the slice thickness of the underlying image series.

As part of adding a FreeDraw segmentation to a series, the user has the option to select both the colour and the free text label. The colour as well as the label will be saved as properties of the DICOM-RTSTRUCT object, along with the annotation location. It is not necessary to use a unique title to separate FreeDraw segmentations and they can be spatially overlapping.

2.3 SmartPaint

SmartPaint is a multi-purpose, pixel perfect, segmentation tool that can be used to effectively create accurate 2D and 3D segmentations using both a standard mouse and keyboard input system, as well as a stylus and drawing pad setup. The user is able to improve their

segmentation by sweeping the mouse, or stylus, over an object or background in a gradual manner, much like how an airbrush is used to paint. The included areas are painted with semi-transparent colours, and the contours are indicated by a thicker line, providing immediate feedback on the chosen interaction plane. The tool has the distinctive feature of adapting to the area or volume in which it operates and is intended to understand what the user wants to depict.

The underlying algorithm that drives this behaviour takes into consideration both the euclidean distance to the mouse cursor and the composition of the local image content. Image content in this context referring to image intensity values. There is a certain probability for every voxel to belong either to the foreground (included) or the background (excluded). When the brush passes through the vicinity of a voxel, the probability that the brush will interact with it can be increased or decreased, and the amount of interaction depends on the distance between the brush centre and the voxel. There is a threshold type of probability which, when it passes, it switches from background to foreground or vice versa.

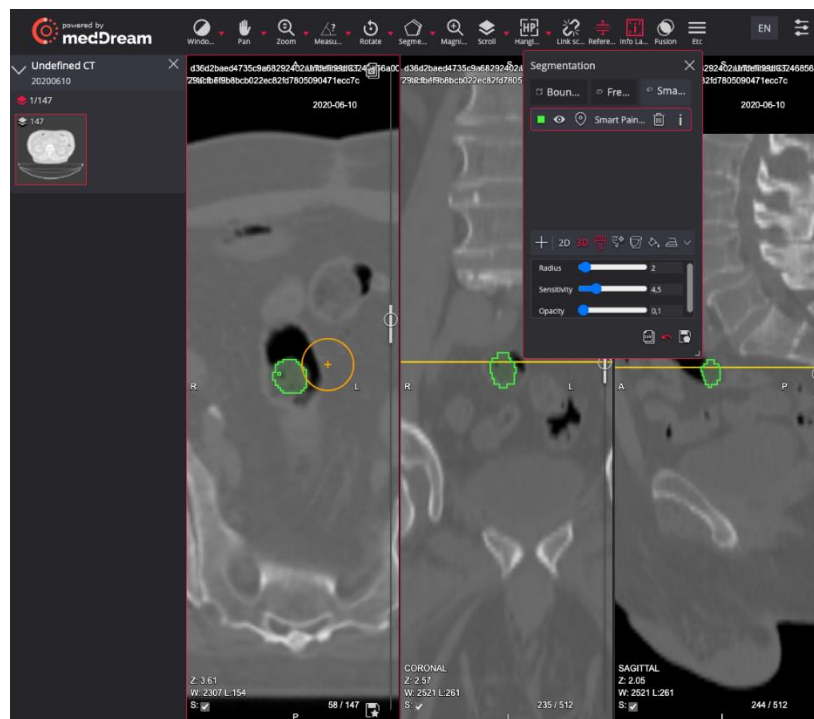


Figure 8 Example of a SmartPaint segmentation.

During a segmentation task, Smart Paint can be modified to perform 2D and 3D operations interchangeably, and the behaviour of SmartPaint can be altered during the segmentation process. That is, the user can initiate a segmentation by a coarse delineation in 3D and then gradually refine the selection by a series of 2D and 3D updates based on preference. Much like the Free Draw tool, Smart Paint allows for an iterative process to accurately capture complex anatomical structures.

In addition to capturing areas and volumes, Smart Paint enables the user to create exclusion zones and islands. SmartPaint has a workflow advantage in the ability to monitor the expansion of a segmentation in a three-plane multi-planar view port setup.

Smart Paint segmentations are, like Free Draw, saved using a closed planar geometric contour type in DICOM-RTSTRUCT and area, holes and islands are captured using boolean operators. Post-processing reconstruction can be carried out using the same third-party libraries for DICOM-RTSTRUCT reconstruction as for Free Draw.

Similarly to Bounding Boxes and Free Draw annotations, Smart Paint objects reside in patient space which offers the ability to map segmentation between series for accurate overlay independent of image series acquisition plane and location, assuming there is a spatial overlap.

3 Use cases

To provide the best possible quality of annotations and segmentations in conjunction with the most resource-efficient workflows, various approaches have been selected for the several use cases and use cases that are part of the EuCanImage project. It should be noted that this is the current suggestion of procedure and modifications can be introduced as a measure to even further strengthen the input that is desired from the AI development work groups that are part of the consortium.

3.1 Use case 1: Liver MRI

For liver MRI a representative single lesion will be selected for annotation. Due to heterogeneity of image acquisition between consortium partners, it will be up to the annotating radiologist to select the most representative series. Although data homogenisation is desired there are regional differences in acquisition strategies which renders an inevitable variation, e.g. a delayed phase 3D T1w acquisition can not be guaranteed. Note that the contributed annotations can be spatially mapped between series.

The annotation tool used will be 3D Bounding Box. A single Bounding Box will be placed in 3D, tightly around the selected lesion and spatial position verified across series using visual inspection. If gross motion is found between series, post processing motion correction can be applied to correct for the spatial mismatch. A default display protocol is applied to aid the workflow. The display protocol will be continuously updated to accommodate participant variations in acquisition strategy, naming convention and device manufacturer differences.

To complement the Bounding Box, a standardised custom report form will be used to provide both a classification and a description of the lesion.

3.2 Use case 2: Liver CT

An annotated liver CT will be based on a representative single lesion.

The annotation tool used will be 3D Bounding Box. A single Bounding Box will be placed in 3D tightly around the selected lesion and spatial position verified across series using visual inspection. A display protocol will be applied, using a three by three grid showing axial, sagittal and coronal planar reformats, to aid the workflow.

To complement the Bounding Box, a standardised custom report form will be used to provide both a classification and description of the lesion.



Figure 9 Example of a use case 2 Bounding Box annotation with the viewer to the right and the custom report form to the left.

3.3 Use case 3: Colorectal CT

To be decided.

3.4 Use case 4,5: Colorectal MRI

To be decided.

3.5 Use case 7: Breast MRI

For the breast MRI multiple lesions (up to three) will be annotated. The annotations will be performed on subtraction of contrast enhanced 3D T1w series, T2w and ADC. Due to the discrepancy between the contributed image series across the various consortium partners, it will be up to the discretion of the annotating radiologist to select the most representative series. The contributed annotations can be spatially mapped between series.

The annotation tool used will be 2D SmartPaint and Bounding Box. A 2D SmartPaint segmentation will be performed on the first subtraction series. This segmentation will be duplicated into the last subtraction series in a post-processing step. A 2D Bounding Box will be placed tightly in the T2w series, as well as in the ADC. Standardised custom report form will be used to provide both a classification and a description of the lesions. The interlinking between the placed 2D SmartPaint and two 2D Bounding Boxes and the custom report form will be done using the title/label of the annotations in combination with a special entry of the custom report form. To account for possible spatial mismatch post processing motion correction (outside of the Collective Minds Radiology platform) can be applied.

By default a hanging protocol will be applied to aid the workflow. It is expected that the hanging protocol will be continually updated to accommodate variations in acquisition strategy, naming convention, and device manufacturer differences between participants.

3.6 Use case 6,8: Mammography

For mammography multiple lesions (up to three) will be annotated. The standardisation of mammography data collection will minimise the expected inter consortium variability and a high homogeneity of data is expected.

The annotation tool used will be 2D Bounding Box. Multiple 2D Bounding Boxes will be placed in the different projections to complete the picture of the lesion presence. Given that a standard mammography examination consists of multiple projections of the same volume the Bounding Box title/label will be used as an indicator to show equivalence between annotations, i.e. the same label will be reused if the same lesion is depicted across instances.

A standard mammography hanging protocol will be used showing a two by two grid with the left and right breast projections on the left and right side respectively.

Standardised custom report form will be used to provide both a classification and a description of the lesions. In order to connect the 2D Bounding Boxes to the custom report form, the annotation title/label will be combined with a special entry on the custom report form.

4 Conclusion

With a cloud-based solution for classification, annotation and segmentation multiple healthcare professionals, regardless of geographic location, can work simultaneously, securely and effectively on curating data. As each use case is handled separately, radiologists can focus immediately on the task at hand with minimal distractions. In addition, users only see information that applies to their task. By performing pseudonymization locally on the clinical partners' sites, the data can be transferred throughout the project in a very safe manner.

Through the use of a standard, DICOM compatible, for storing both annotations and segmentations, data can be easily shared between the project's two storage repositories. Additionally, it enables easy access for subsequent AI research.

For colorectal use cases, the emerging work involves completing the selection and adaptation of the segmentation tool as well as establishing its role in the corresponding workflows. This work is conducted in parallel with other use cases.

Activating the large user group for the liver and mammography use cases requires only a few final adjustments.