



D2.3

Annotated ground truth data

CLEPIO

DATE : 2026.03.31

European
Innovation
Council



Funded by
the European Union

Document control page

Project name	3DPATH
Document	D2.3: Annotated ground truth data
Type (distribution level)	☒ Public ☐ Sensitive (confidential as specified in the Grant Agreement)
Due delivery date	31 03 2026
Date of delivery	31 03 2026
Status and version	1.0
Deliverable responsible	Clepio
Author(s)	Ludovico Silvestri (Clepio) Alessandro Grande (Clepio) Alessandra Franceschini (Clepio)
Official reviewer(s)	Taiana Deodato Reis (Clepio) Costanza Gori (Clepio) Claudia Capitini (Clepio)

PARTNERS:



Table of contents

1. Executive summary.....	3
2. Rationale	4
3. Images to be annotated	4
4. Annotation data	6
5. Data availability	7

1. Executive summary

This document provides a description of the annotated ground truth data produced within the 3DPATH project, and a link to the repository containing the actual files. These data will be used in the next 12 months for the training and the performance assessment of the cell detection software developed in WP2. In addition, this document provides guidelines that shall be followed in case future ground truth data would be needed to improve the performances of the algorithms.

Throughout this document, '**GT**' means 'ground truth', '**LSFM**' means 'light-sheet fluorescence microscopy', '**IGEA-3DS**' refers to the 3D scanner developed in the 3DPATH project, and '**IGEA Analytics**' refers to the software module dedicated to cell detection developed in the 3DPATH project. The latter two components belong to a larger solution that we named '**IGEA**'.

2. Rationale

Well-annotated GT is a mandatory step for the training and verification of image analysis tools aimed at extracting quantitative information from pixel images.

Within the workflow of Igea Analytics, we currently identify localization of nuclei (i.e., identifying the [x,y,z] coordinates of the center of mass of each cell nuclei) as the main image analysis goal.

To be useful, GT must be as representative as possible, including images from multiple samples, with large variability in luminosity and in shape / size of labeled nuclei. As a general point, it is better to include many small image chunks with large variability than few large images.

The GT produced in the 3DPATH project is a collection of 3D images, extracted from large-scale reconstruction obtained by the current (TRL5) prototype of IGEA-3DS, coupled with relevant annotations made by human experts. 3D images are saved in the same representation (e.g., unsigned 16 bit) as the original data that will be processed by the Igea Analytics pipeline.

As for now, we only include in the GT the relevant channel (i.e., the one corresponding to the marker of interest) out of the 4 generated by IGEA-3DS. In future, inclusion of multiple channels might be beneficial.

For localization of nuclei, annotations are stored as a .csv file with the list of the [x,y,z] coordinates of all manually labeled nuclei. There is no strict rule for the number of annotations. However, from previous experience (Frasconi et al., Bioinformatics 2014; Silvestri et al., Nature Methods 2021; Franceschini et al., Cell Reports 2023), we estimate between 5'000 and 10'000 nuclei a minimum number of annotations to get consistent results.

3. Images to be annotated

We annotated imaging datasets of $196 \times 196 \times 65$ voxels, tantamount to a physical volume of $128 \times 128 \times 130 \mu\text{m}^3$. The field of view in these datasets is sufficient to distinguish a large number of cells while providing a good overview of the tissue environment (see Figure 1).

Data are saved in 3D Tiff format. Each voxel is represented with an Unsigned 16-bit integer, little-endian. Files are stored in different sub-folders according to the tissue of origin, and named with a progressive number. A complete list of the annotated files is reported in Table 1.

File name	Min value	Max value	Mean value
Bladder-01.tif	280	8963	1148
Bladder-02.tif	155	8572	1767
Bladder-03.tif	547	11491	2529
Bladder-04.tif	284	8248	1185
Bladder-05.tif	403	7794	1747
Bladder-06.tif	397	7327	1355
Bladder-07.tif	858	8115	2046
Bladder-08.tif	813	8604	1945
LymphNode-01.tif	174	12454	1079
LymphNode-02.tif	697	5550	1660
LymphNode-03.tif	718	9021	3540
LymphNode-04.tif	306	5394	882

Table 1. Summary of 3D datasets used for annotation. The different files used for annotation are reported, together with their minimum, maximum and average value of pixel intensity.

In Figure 1 we report a few representative snapshots of different images, all displayed with a linear LUT where everything smaller than 150 is black and everything larger than 8000 is white. As evident from these images, the data used for annotations are highly heterogeneous in terms of luminosity, contrast, nuclei size and shape.

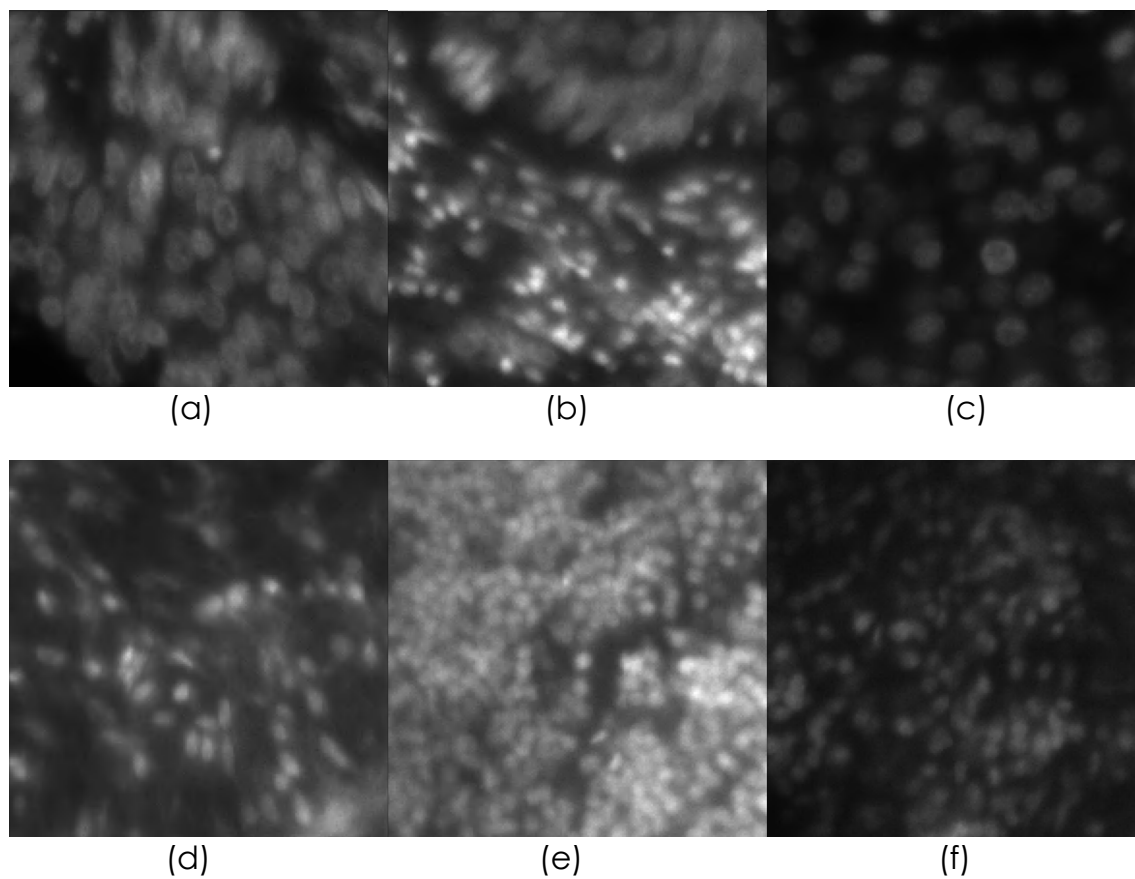


Figure 1. Example of nuclei images used for GT generation. Snapshots from selected slices from 6 different 3D tiffs: Bladder-02 (a), Bladder-03 (b), Bladder-04 (c), Bladder-07 (d), LymphNode-03 (e), LymphNode-04 (f). All images are displayed with a linear LUT between 150 and 8000.

4. Annotation data

Centroids of nuclei were manually labeled using Napari (<https://napari.org>), an open-source software for image visualization and annotation. The coordinates of annotated centroids were saved in a .csv file with the same name of the original image, in the same folder.

Manual annotations were performed by skilled personnel expert in bi microscopy. All data were reviewed by a second annotator before release. Table 2 reports the name of annotator and reviewer for each imaging dataset, and the number of annotated nuclei therein.

File name	Annotator	Reviewer	# of annotated cells
Bladder-01.csv	LSI	AFR	708
Bladder-02.csv	AFR	CCA	750
Bladder-03.csv	CCA	CGO	850

Bladder-04.csv	CGO	LSI	566
Bladder-05.csv	LSI	AFR	372
Bladder-06.csv	AFR	CCA	467
Bladder-07.csv	CCA	CGO	434
Bladder-08.csv	CGO	LSI	980
LymphNode-01.csv	LSI	AFR	1224
LymphNode-02.csv	AFR	CCA	1426
LymphNode-03.csv	CCA	CGO	1338
LymphNode-04.csv	CGO	LSI	1092

Table 2. Annotation files. List of annotation files with the names of annotators, reviewers and total number of annotated cells. LSI is Ludovico Silvestri, AFR Alessandra Franceschini, CCA Claudia Capitini, CGO Costanza Gori.

The total number of annotated cells is 10207.

5. Data availability

Raw imaging data together with the annotation .csv files are publicly available on Zenodo, with the DOI <https://doi.org/10.5281/zenodo.19351592>.