Summary of the Medical Imaging Data Anonymization Program

For any questions and comments regarding usage and technical details, please send correspondences to *Daniel Eftekhari* at daniel@16bit.ai

In fulfilment of the requirements for a secure and robust approach to anonymizing medical imaging data, 16 Bit has developed a reliable anonymization software program. The program removes patient health information (PHI) from digital imaging and communications in medicine (DICOM) files, as well as any associated physician reports, prior to exporting the data to non-hospital research centers. Specifically, the anonymization program removes all patient, physician, and institutional identifiers from a DICOM instance and its associated medical reports, while preserving the integrity of the underlying medical imaging data. The software can be executed on any operating system (Windows, Linux, MacOS) and is designed efficiently for use on even the largest of medical datasets. The remainder of this document will provide a brief, high-level description of the program and its functionality.

The program's anonymization protocol is motivated by the following universal DICOM standard requirement. All medical images must be uniquely identified by four values: the patient id, study id, series id, and instance id. As demonstrated in Figure 1, each patient may have many associated studies, each study may have many associated series, and each series may have many associated instances. Therefore, prior to exporting any DICOM files and their associated physician reports, the software replaces each of these identifiers with new values. These values preserve the patient-study-series-instance ordering between files, whilst eliminating all traces to the original identifiers. Other sensitive information in the DICOM instances and physician reports are also removed.

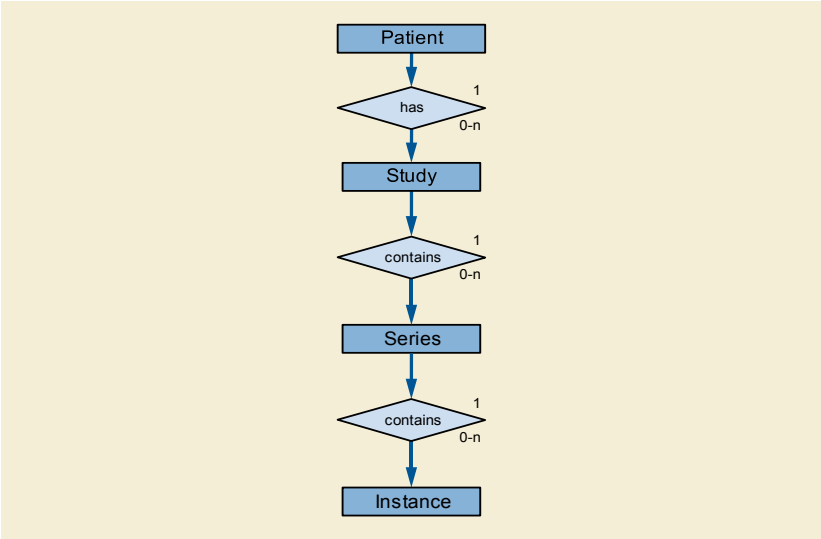

Figure 1. All DICOM instances can be uniquely identified by the following four values: the patient, study, series, and instance ids. (*image source: https://book.orthanc-server.com/dicom-guide.html*)

To improve the end-user's experience, the program contains many additional and easy-to-use features. For example, anonymized DICOMs can either be saved to a single output folder or grouped into subfolders based on their (anonymized) patient or study ids. As a second example, the user need not group DICOM instances and their associated reports into a specific folder structure prior to program execution, improving ease of use.

Furthermore, the program has been designed to facilitate anonymization on large datasets, typical of large medical centers. For example, the program can be paused if the dataset size is expected to exceed the space available on any one hard drive, then resumed once additional storage is available. Additionally, the program can be executed even in cases where the data is made available in a streaming fashion (for example as new cases are added to the hospital database system), without needing to re-anonymize all previously analyzed instances. For more information on the program's specifications and features, and for any additional feature requests, please contact 16 Bit.