Segmed

The White Papers

February 2021

Table of Contents

Introduction **3**

Our Vision of the Future **4**

Eliminating Bias in AI 4

Where Healthcare Can Go 4

Our Customers **5**

Who Do We Work With? 5

Who Do We Not Work With? 6

Ethics & Quality Assurance **7**

Data De-Identification **8**

What Is De-Identification? 8

Why Does De-Identification Matter at Segmed? 8

Identifiers in DICOM Headers 8

How Does Segmed De-Identify Data? 10

Key Laws and Regulations 10

HIPAA 10

THE COMMON RULE 12

CCPA 13

GDPR 14

Introduction

Segmed was founded by a group of international physicians, researchers, and engineers from Stanford. We believe that healthcare providers have a responsibility to innovate and build upon the resources they have at their disposal in order to deliver better care. In order to efficiently create those next-gen health systems, Segmed is here to bridge the divide between technology developers with a need for data and health systems who want to put their data to work for their patients. We compiled this series of white papers to explain our intentions, standards, and how we build partnerships.

Connect with us by filling out our contact form at www.segmed.ai/connect.

To partner right away, email us at sales@segmed.ai. You can use our de-identified medical database to develop new solutions that will help patients and help us build the future of healthcare.

Our Vision of the Future

Eliminating Bias in AI

Currently, AI requires human programmers to feed the algorithms a variety of data so that they can respond to as many situations as possible. This means that bias can occur when the data isn't diverse enough. For example, a 2018 study found that an algorithm detected skin cancer with better accuracy than doctors. Unfortunately, the algorithm was trained on an easily accessible dataset that did not include very many dark-skinned people. As a result, the incredible accuracy for finding skin cancer only worked on light-skinned patients, and would potentially miss skin cancer in dark-skinned patients.[1]

We don't want bias to happen because AI companies have difficulty getting diverse data. That's why we work to collect data from different types of facilities, such as large medical research centers and small, private clinics.[2] While other companies are focused on the U.S., we also gather data from Asia, Africa, and Europe with the goal of reaching all modalities and a comprehensive variety of pathologies.

Where Healthcare Can Go

Today's healthcare is full of disparities. Rural areas suffer from lack of available healthcare and developing nations may not have the resources that developed nations do. At Segmed, we see a future where medical AI helps people access a better standard of care, no matter where they are in the world.

Medical AI can contribute to healthcare equity by improving physician efficiency and quality and also increasing the capabilities of other healthcare workers to fill in for physicians.[3] This starts with sharing data and contributing to medical databases to allow for greater accessibility and availability of quality data. With a collaborative environment and more people contributing to AI in healthcare, we can see progress in more accurate diagnosing, less expensive healthcare, and greater equality for long and healthy lives.

---

[1] Angela Lashbrook. 2018. AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind
[2] Read more on our blog post, Bias in Healthcare
[3] Jonathan Guo & Bin Li. 2020. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries

Our Customers

Who Do We Work With?

We work with the companies at the cutting edge of healthcare in order to advance innovation. We service customers in diverse sectors such as:

- Technology
- Medical Artificial Intelligence and Machine Learning
- Biopharmaceuticals
- Medical Device Manufacturers
- Medical Robotics OEMs
- Academic Researchers

Who Do We Not Work With?

Our mission at Segmed is to improve the base standard of care across populations. We aim to provide better, cheaper care to more people through aiding digital health development and innovation. To align with our goals and those of our numerous data partners, we do not work with any companies who may use data to the disadvantage of patients. We have identified these particular sectors as:

- Insurance Companies
- Actuarial Firms

Ethics & Quality Assurance

We understand that being entrusted with patient data, even de-identified, is a huge responsibility. That is why we developed a rigorous ethics and quality assurance program to provide our data partners with the utmost transparency and accountability:

- **Additional De-Identification.** Even after receiving anonymized data from our partners, we run our proprietary algorithms to check for identifying information or outliers, and remove anything that could link back to the patient or institution it came from.
- **Data Transparency for Partners.** We believe that data partners should be able to see how their patients' data is used and how that use is aligned with their own core values and missions. To that end, we provide reports quarterly as well as upon request to our data partners.
- **Regulatory Compliance.** All data activities pursued by Segmed are compliant with and conform to the requirements as listed by HIPAA, the Common Law, CCPA, and the GDPR.
- **Mission-Driven Team.** Every team member at Segmed understands that our ultimate responsibility is to the patients. We take our ethical roles seriously and constantly keep our mission in mind as we work together towards our goal.
- **Value-Add for Patients.** Most importantly, the goal of Segmed is to improve the standard of care for patients. This means we will NEVER share data with insurance or actuarial companies, because we believe in increasing access, not cost, to cutting-edge healthcare. Instead we have worked with AI developers, robotics OEMs, biopharmaceuticals, and researchers across the world to attain those common goals.

We have been lucky enough to work with 50+ data partners worldwide, ranging from large health systems to smaller imaging centers. We apply the same level of care and detail to each and every single partner, and are constantly learning and improving. If you have any questions regarding our ethical and quality assurance policies, feel free to reach out at info@segmed.ai.

Data De-Identification

What Is De-Identification?

“De-identification” generally means that the information, which identifies or is reasonably capable of identifying an individual, has been removed from a data set. However, the precise legal standard may vary based on the nature of the data and the jurisdiction. Most medical imagery is captured in accordance with the Digital Imaging and Communications in Medicine (DICOM) standard. Segmed processes and analyzes the standardized header data in DICOM medical images to conform it to the requirements of various key laws and regulations on de-identification, including HIPAA, the Common Rule, the CCPA, and the GDPR, as discussed below.

Why Does De-Identification Matter at Segmed?

Patients are at the core of our mission. We have designed our de-identification policies, procedures, contracts, and technical measures to safeguard patient privacy while simultaneously helping to improve algorithms for disease diagnosis and treatment. Segmed works hard each and every day to make this our priority.

Identifiers in DICOM Headers

The following table contains the different types of headers that are included in DICOM files. The direct identifiers are headers that are always removed by Segmed from DICOMs. These headers can easily identify the patient. Indirect identifiers alone may not be sufficient to identify a patient are headers; however, if combined, they could potentially lead to identification of a specific patient. Non-identifiers are headers that cannot reasonably identify the patient. Segmed maintains a limited number of indirect identifiers, which may vary depending on the data set. Non-identifiers cannot be used to reasonably identify a patient and are maintained in Segmed data sets.

DIRECT IDENTIFIERS

1. Names
2. Geographic subdivisions smaller than a state[4]
3. All elements of dates (except year)[5]
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers/numbers, including license plate numbers
13. Device identifiers and serial numbers
14. URLs
15. IP address numbers
16. Finger and voice prints
17. Full face images and any comparable images
18. Any other unique identifying number, characteristic, or code

INDIRECT IDENTIFIERS

1. Gender
2. Race
3. Ethnicity
4. Religion
5. Age
6. Marital status
7. Household composition
8. Number of children
9. Place of birth
10. Education
11. Major
12. Income
13. Job title
14. Place of work
15. Medical condition
16. Dates (of graduation, arrest, marriage)
17. Uncommon characteristics
18. Direct identifiers of household members

NON-IDENTIFIERS

1. Modality
2. Manufacturer
3. Age (rounded birthday)
4. Pregnancy
5. Slice thickness
6. State or country

[4] Except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000.

[5] Ages over 89 must be put into the category of age 90 or older

How Does Segmed De-Identify Data?

Typically, before Segmed receives medical data, it is de-identified by the data provider (usually a hospital). Then, we check the data and validate de-identification using our internally developed process, which also removes any remaining direct identifiers along with select indirect identifiers. During this step, Segmed also assigns each patient a unique randomly generated ID number. Although most data sources already provide us with data that has been assigned these randomly generated numbers, we use our own to help further reduce the risk of re-identification and to standardize our database. Segmed does this to be able to internally track medical data from different records and dates within a data set in order to have more information useful for research purposes, without unreasonably increasing the risk of identifying the patient. For example, it is extremely useful to be able to associate electronic medical records ("EMRs") with radiology data within the same data set, even while that data remains de-identified/pseudonymized consistent with the laws and regulations described above. Segmed also takes the additional step of analyzing image headers using an algorithm to check for and remove information about the patient, doctor, or hospital. Non-identifiers such as modality (the method of obtaining medical data, such as X-Ray, CT, MRI, or ultrasound) remain in the data set.

Segmed does not reidentify data as part of this process. Further, Segmed does not provide keys to customers to re-identify individuals within data sets, and re-identification is prohibited by Segmed's policies and procedures as well as in Segmed's contracts with customers.

Key Laws and Regulations

The laws and regulations described in this section provide the key legal frameworks in the United States and European Union for the confidentiality of medical imagery data and related information. The following laws and regulations also determine the ways in which medical data must be de-identified, and the repercussions if they are not properly de-identified. If you have questions regarding compliance with other laws or in other jurisdictions, please contact us at info@segmed.ai.

HIPAA

The Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, expanded by the Health Information Technology for Economic and Clinical Health Act ("HITECH") of the American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, and amended by the 21st Century Cures Act ("Cures Act"),

Pub. L. 114-255 (collectively, "HIPAA") is implemented through regulations at 45 C.F.R. Parts 160, 162, and 164 (together, the "HIPAA Rules"). The four regulations that make up the HIPAA Rules regarding privacy and security are:

1. **Privacy Rule:** Sets standards to protect patients' medical records and limits disclosures of PHI without patient consent.

2. **Security Rule:** Sets administrative, physical, and technical safeguards for electronic protected health information (ePHI) for covered entities and business associates.

3. **Data Breach Notification Rule:** Sets requirements for reporting unauthorized uses and disclosures of PHI to patients and to the Office of Civil Rights (OCR).

4. **Enforcement Rule:** Sets guidelines for how complaints and violations will be investigated and consequences for violations.

The Privacy Rule applies to uses and disclosures of "protected health information," which means individually identifiable health information held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. "Individually identifiable health information," with limited exceptions, is information, including demographic data, that relates to:

- the individual's past, present or future physical or mental health or condition,

- the provision of health care to the individual, or

- the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe it can be used to identify the individual.[6]

There are no HIPAA restrictions on the use or disclosure of de-identified health information.[7] De-identified health information neither identifies nor provides a reasonable basis to identify an individual. There are two ways to de-identify information; either under:

[6] 45 C.F.R. § 160.103
[7] 45 C.F.R. §§ 164.502(d)(2), 164.514(a) and (b)

1. the "Safe Harbor" method, (a) by removing 18 specified identifiers of the individual and of the individual's relatives, household members, and employers; and (b) only if the covered entity has no actual knowledge that the remaining information could be used to identify the individual; and

2. the "Expert Determination" method, which involves a formal determination by a qualified statistician that determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information.[8]

Segmed utilizes HIPAA's Safe Harbor method for de-identification by confirming removal of all 18 direct identifiers in DICOM header data of medical images in its data sets. Segmed has further implemented policies, procedures, and technical measures against re-identification, and also includes in its contracts a requirement that customers not seek to re-identify the specific individuals whose images were included in the data sets.

THE COMMON RULE

Heavily influenced by the Belmont Report, written in 1979 by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, the Federal Policy for the Protection of Human Subjects or the "Common Rule" was published by the Department of Health and Human Services ("HHS") at 45 CFR Part 46 in 1991, establishing the basic provisions for federally-funded human subjects research, including the requirement for institutional review boards ("IRBs") and informed consent. The Common Rule has been codified in separate regulations by, or affirmed by executive order for, most Federal departments and agencies, and it applies to human subject research conducted or supported by each federal department/agency that has published its own regulations signing on to the Common Rule.[9] A major revision of the Common Rule was published by HHS in 2018. Some federal departments and agencies have not yet updated their regulations to the new version of the Common Rule, so both the pre-2018 version and the 2018 version may be relevant.

The pre-2018 Common Rule provides that research involving anonymous or de-identified information is expressly exempt from regulation, and therefore the consent and institutional review board requirements under the Common Rule. Exemption 4 from the Common Rule applies to: "Research involving the

[8] 45 C.F.R. § 164.514(b)

[9] The FDA's regulations differ somewhat, but the agency is required to "harmonize" with the Common Rule whenever permitted by law pursuant to section 1002 of the 21st Century Cures Act, Public Law 114-255.

collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects."[10] Private information or specimens are "[not] individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding systems"[11]

OHRP Guidance provides that research involving only coded private information does not involve human subjects if (i) the investigator cannot "readily ascertain" the identity of the individual because the key has been destroyed before the research begins, (ii) the keyholder has agreed not to release the key to investigators under any circumstances, (iii) there are institutional review board (IRB)-approved written policies prohibiting release of the key until individuals are deceased, or (iv) there are other legal requirements prohibiting the release of the key to the investigators until the individuals are deceased.[12] Segmed does not release any codes or other keys to customers intended to allow re-identification.

The 2018 Common Rule maintains the pre-2018 exemption for information that cannot be identified directly or through identifiers [codes] linked to the subjects, with the added requirement, similar to HIPAA, that the investigator will not re-identify the subjects and will not contact the subjects. Notably, these exemptions do not apply to human subjects research involving prisoners or children, and Segmed does not utilize medical imagery from these populations in its data sets.

Although the Common Rule may apply to information more broadly than HIPAA's coverage for PHI, Segmed's approach to de-identification removes direct identifiers, does not include keys for customers to re-identify data, includes contractual requirements against customer re-identification of Segmed data sets, and is designed to address the HIPAA Safe Harbor standard and the Common Rule's requirements and guidance (pre- and post-2018) on de-identification, in addition to the other key laws and regulations discussed below.

CCPA

The California Consumer Privacy Act ("CCPA") was signed into law on June 28, 2018 and became effective on January 1, 2020, with implementing regulations that took effect August 14, 2020. The CCPA gives individuals more control over their personal information that businesses and covered entities

[10] 45 C.F.R. § 46.101(b)(4)
[11] Office of Human Research Protections Guidance on research involving coded private information or biological specimens. 2008. http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm.
[12] 45 C.F.R. § 46.101(b)(4)

collect about them. “Personal information” is defined broadly under the CCPA to mean “information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household. Personal information includes, but is not limited to, the following if it identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household,” including a number of enumerated categories and exceptions under the CCPA.[13] Some of the privacy rights for California residents include:

1. The right to know about the personal information a business collects about them and how it is used and shared;
2. The right to delete personal information collected from them (with some exceptions);
3. The right to opt-out of the sale of their personal information; and
4. The right to non-discrimination for exercising their CCPA rights.

The CCPA excludes de-identified information from its definition of personal information; however, the CCPA definition of “de-identified” information differs, and arguably applies more stringent requirements than HIPAA or the Common Rule. Under the CCPA, for the information to be deidentified, the information “cannot reasonably identify, relate to, describe, be capable of being associated with, or be linked, directly or indirectly, to a particular consumer;” technical safeguards and business processes must be used to prohibit reidentification of the information; business processes must be implemented to prevent inadvertent release of de-identified information; and the business must make no attempt to re-identify the information.[14] Segmed has implemented business policies, procedures, technical measures, and language in its customer contracts designed to address CCPA’s requirements for de-identification.[15]

GDPR

The General Data Protection Regulation (“GDPR”) is a regulation in the European Union (“EU”) on data protection and privacy which took effect on May 25, 2018. The GDPR applies to companies that are based

[13] Cal. Civ. Code 1798.140(o)
[14] Cal. Civ. Code § 1798.140(h)
[15] Note: The CCPA and its implementing regulations do not expressly state that data de-identified in accordance with the HIPAA De-Identification Standard or the Common Rule is “de-identified” for the purposes of the CCPA. As of August 2020, 5he California legislature is considering clarifying such changes, among others, in a pending bill, AB 713.

in the EU and to companies outside the EU that target or monitor EU individuals. The GDPR applies to "personal data," defined broadly as "information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."[16]

The seven core data protection principles pursuant to GDPR Article 5 that apply to the personal data of individuals located in the European Union include:

1. **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.

2. **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.

3. **Data minimization** — You should collect and process only as much data as absolutely necessary for the purposes specified.

4. **Accuracy** — You must keep personal data accurate and up to date.

5. **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.

6. **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).

7. **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

The GDPR does not apply to "anonymised" data.[17] This occurs only if the personal data is rendered anonymous in such a manner that the data subject is not or no longer identifiable, analyzed through all means reasonably likely to be used.[18] This is generally considered to be a high bar, requiring removal of

[16] GDPR Art. 4(1)
[17] GDPR Recital 26
[18] See, Article 29 Working Party, Opinion No. 05/2014 on Anonymization Techniques, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

all direct and most to all indirect identifiers. Segmed does not anonymise data within the meaning of the GDPR, because some indirect identifiers are maintained within the data set. Instead, Segmed's de-identification process is consistent with what the GDPR terms "pseudonymisation."

Pseudonymisation under the GDPR means "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."[19] Pseudonymised data is subject to the GDPR, since it may still be possible to re-identify the data and associate it with the patient, including for example, through indirect identifiers or coded data sets. This approach differs from HIPAA, the Common Rule, and the CCPA, which all exclude data that has been de-identified in accordance with their requirements, even if some indirect identifiers remain.

[19] GDPR Art. 4(5)