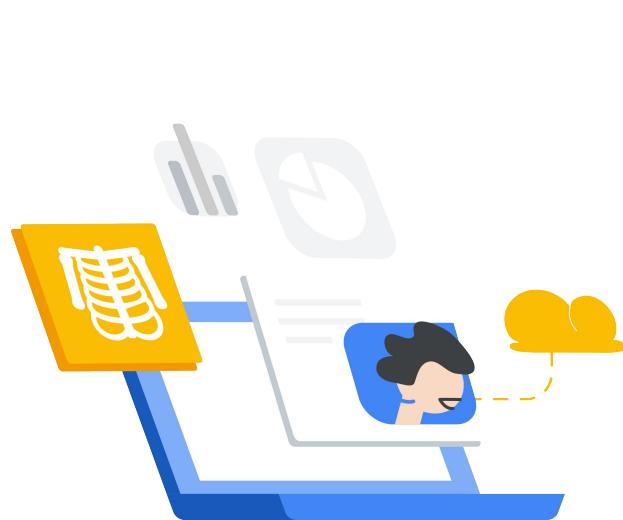Google Cloud

# Healthcare Data De-Identification

Remove Personally Identifiable Information (PII) from

Healthcare Data with the **Google Cloud Healthcare API**

**Bob Lou, MD**
Product Management
Google Cloud Healthcare & Life Sciences

# The HIPAA 18 identifiers

1. Names
2. Geographic locations smaller than a state (many 3-digit zip codes allowed)
3. Dates (year allowed in many cases)
4. Telephone numbers
5. Vehicle identifiers (ex. VINs, license plate numbers)
6. Fax numbers
7. Devices IDs and serial numbers
8. Email addresses
9. Web URLs
10. Social Security Numbers
11. IP addresses
12. Medical Record Numbers (MRNs)
13. Biometric identifiers (ex. fingerprints and voice prints)
14. Health plan beneficiary numbers
15. Full-face photographs and similar images
16. Account numbers
17. Any other unique identifying number, characteristic, or code
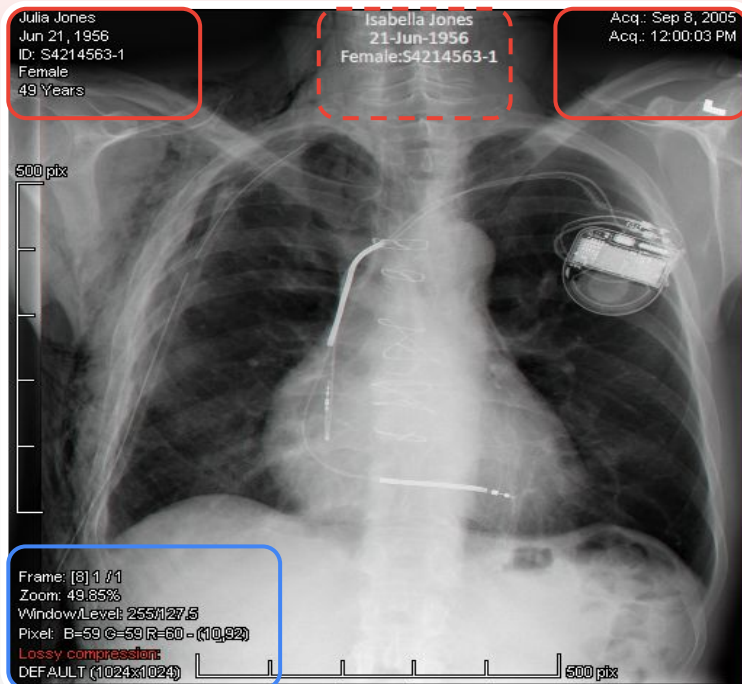18. Certificate / license numbers

Google Cloud

# Healthcare API DeID: Transformation options

1. **Redact/remove**

2. **Character mask** replace sensitive data with a fixed character, ex. "John" > "****"

3. **Replace with info type** replace input with a category label, ex. "John" > "[NAME]"

4. **Date shift** shift by a random number of days, with the option to maintain chronological order between dates belonging to the same patient or population

5. **Crypto hash** replace input with a hashed value generated using a (customer-defined) encryption key

Google Cloud

# Sample DICOM Tags (Before DeID)

```
[
  {
    "00020002":{"vr":"UI","Value":["1.2.840.10008.5.1.4.1.1.7"]},
    "00020003":{"vr":"UI","Value":["1.2.276.0.7230010.3.1.4.8323329.78.1539083058.523695"]},
    "00020010":{"vr":"UI","Value":["1.2.840.10008.1.2.4.50"]},
    "00020012":{"vr":"UI","Value":["1.2.276.0.7230010.3.0.3.6.1"]},
    "00020013":{"vr":"SH","Value":["OFFIS_DCMTK_361"]},
    "00080005":{"vr":"CS","Value":["ISO_IR 100"]},
    "00080016":{"vr":"UI","Value":["1.2.840.10008.5.1.4.1.1.7"]},
    "00080018":{"vr":"UI","Value":["1.2.276.0.7230010.3.1.4.8323329.78.1539083058.523695"]},
    "00080020":{"vr":"DA","Value":["20110909"]},
    "00080030":{"vr":"TM","Value":["110032"]},
    "00080050":{"vr":"SH"},
    "00080064":{"vr":"CS","Value":["WSD"]},
    "00080070":{"vr":"LO","Value":["Manufacturer"]},
    "00080090":{"vr":"PN","Value":[{"Alphabetic":"John Doe"}]},
    "00081090":{"vr":"LO","Value":["ABC1"]},
    "00100010":{"vr":"PN","Value":[{"Alphabetic":"Ann Johnson"}]},
    "00100020":{"vr":"LO","Value":["S1214223-1"]},
    "00100030":{"vr":"DA","Value":["19880812"]},
    "00100040":{"vr":"CS","Value":["F"]},
    "0020000D":{"vr":"UI","Value":["2.25.70541616638819138568043293671559322355"]},
    "0020000E":{"vr":"UI","Value":["1.2.276.0.7230010.3.1.3.8323329.78.1531234558.523694"]},
    "00200010":{"vr":"SH"},
    "00200011":{"vr":"IS"},
    "00200013":{"vr":"IS"},
    "00200020":{"vr":"CS"},
    "00280002":{"vr":"US","Value":[3]},
    "00280004":{"vr":"CS","Value":["YBR_FULL_422"]},
    "00280006":{"vr":"US","Value":[0]},
    "00280010":{"vr":"US","Value":[1024]},
    "00280011":{"vr":"US","Value":[1024]},
    "00280100":{"vr":"US","Value":[8]},
    "00280101":{"vr":"US","Value":[8]},
    "00280102":{"vr":"US","Value":[7]},
    "00280103":{"vr":"US","Value":[0]},
    "00282110":{"vr":"CS","Value":["01"]},
    "00282114":{"vr":"CS","Value":["ISO_10918_1"]}
  }
]
```

Google Cloud

Original DICOM Image

Julia Jones
Jun 21, 1956
ID: S4214563-1
Female
49 Years

Isabella Jones
21-Jun-1956
Female:S4214563-1

Acq.: Sep 8, 2005
Acq.: 12:00:03 PM

500 pix

Frame: [8] 1 /1
Zoom: 49.85%
Window/Level: 255/127.5
Pixel: B=59 G=59 R=60 - (10,92)
Lossy compression:
DEFAULT (1024x1024)

500 pix

Metadata tags containing PII

Burnt-in text containing PII (i.e. pixel data)
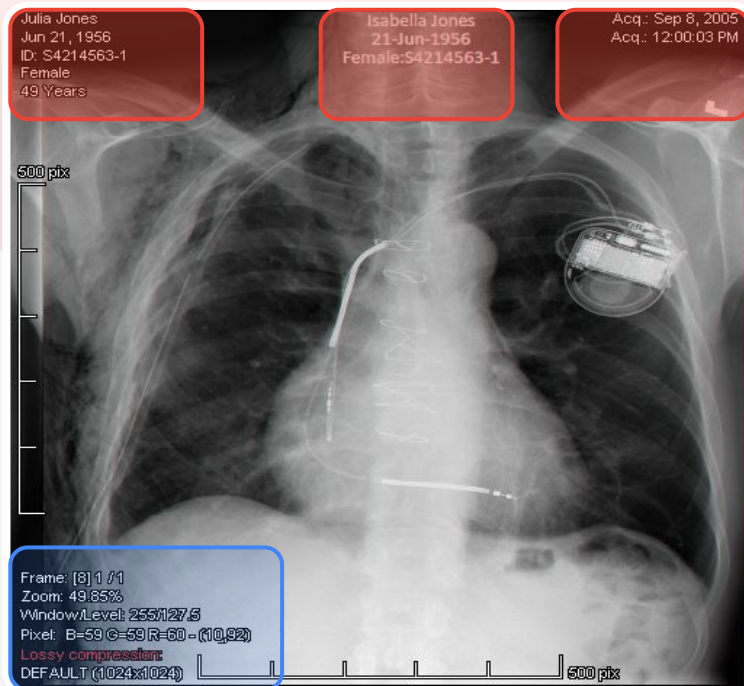
Metadata tags without PII

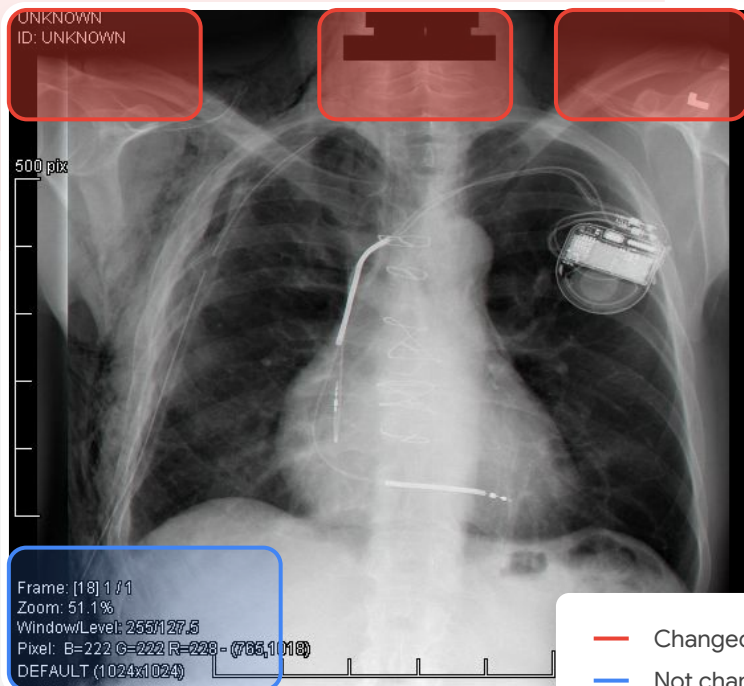Google Cloud

## Original DICOM Image
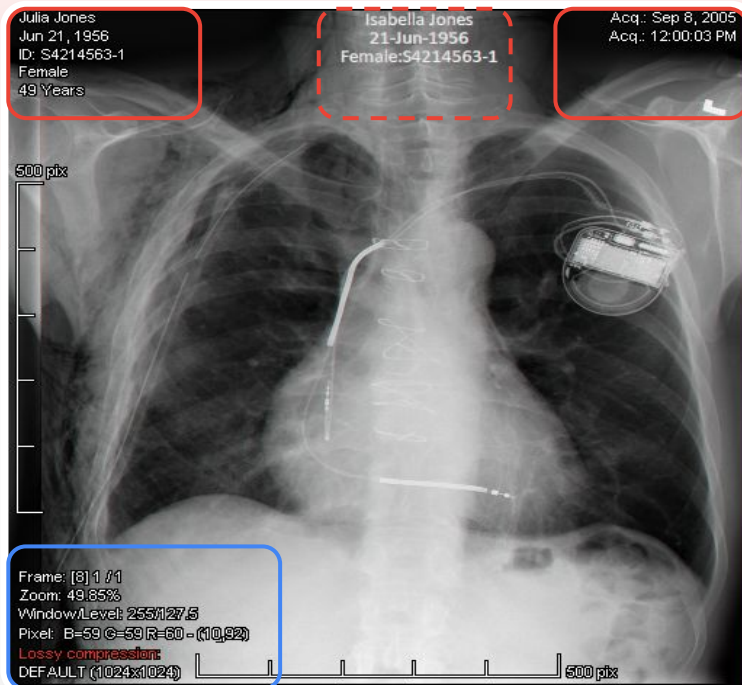
## De-identified DICOM Image
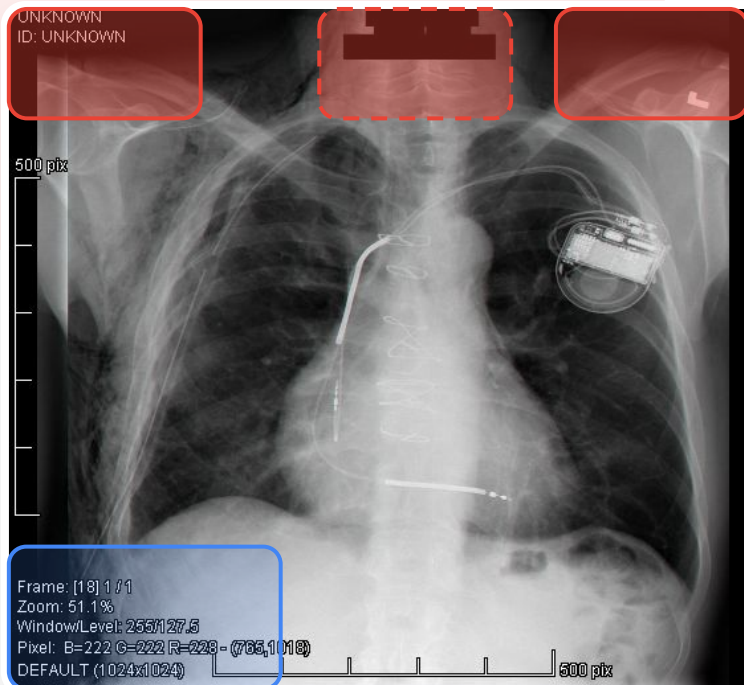
## Original DICOM Image
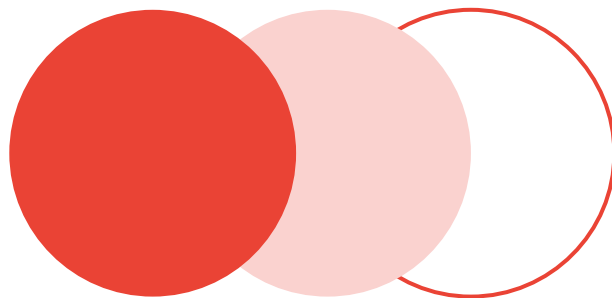
## De-identified DICOM Image
(using Cloud Healthcare API DeID)



Changed
Not changed

# FHIR

## Original (FHIR) record

```
{
  "address": [
    {
      "city": "Anycity",
      "district": "Anydistrict",
      "line": [
        "123 Main Street"
      ],
      "period": {
        "start": "1990-12-05"
      },
      "postalCode": "12345",
      "state": "CA",
      "text": "123 Main Street Anycity, Anydistrict, CA 12345",
      "type": "both",
      "use": "home"
    }
  ],
  "birthDate": "1980-12-05",
  "gender": "female",
  "id": "r77433dd-dkeuc-633743nfd-383nfdsjds732",
  "name": [
    {
      "family": "Smith",
      "given": [
        "Darcy"
      ],
      "use": "official"
    }
  ],
  "resourceType": "Patient",
  "text": {
    "status": "generated",
    "div": "<div><p><b>Patient</b></p><p><b>Name</b></p><p> Smith, Darcy </p><p>
  }
}
```

## De-identified FHIR record

```
{
  "address": [
    {
      "city": "",
      "district": "",
      "line": [
        ""
      ],
      "period": {
        "start": "1990-12-05"
      },
      "postalCode": "",
      "state": "CA",
      "text": "",
      "type": "both",
      "use": "home"
    }
  ],
  "birthDate": "1981-02-24",
  "gender": "female",
  ... r77433dd-dkeuc-633743nfd-383nfdsjds732",
  "meta": {
    "lastUpdated": "2018-01-01T2018-01-01T00:00:00+00:00",
    "versionId": "MTU0MDU4NTcxNjI2MTUxNDAwMA"
  },
  "name": [
    {
      "family": "",
      "given": [
        ""
      ],
      "use": "official"
    }
  ],
  "resourceType": "Patient",
  "text": {
    "div": "<div><p><b>Patient</b></p><p><b>Name</b></p>: [PERSON_NAME][PERSON_NAME][PERSON_NAME]
    "status": "generated"
  }
}
```

Google Cloud

# Thank you