# Anonymizing Data

## Obscuring Names for Demonstration

At times you will want to demonstrate the OpenMRS web application to people outside of your organization. Displaying real patient names would not be very appropriate.

OpenMRS has a built in way to simply obscure all patient's names in the system using your runtime properties:

```
obscure_patients=false
obscure_patients.given_name=John
obscure_patients.middle_name=P.
obscure_patients.family_name=Smith
```

## Complete Database Anonymizer

When you need to have (nearly) complete HIPAA compliant anonymity of patients and data you need to do a lot of scrubbing to the underlying database.

This archive:anonymizing sql script will scramble your patient names, patient addresses, locations, and any dates stored against patients. User names and passwords will be reset to username-"id" / test.

> Do not run this on a live database. Only run this on a copy. Accurate data is impossible to recover from this.

> This anonymization may not be fully HIPAA compliant. It is *very* difficult to truly anonymize (de-identify) data such that someone could not re-identify a patient. People (and a growing number of tools) can come up with clever ways of figuring out identities from supposedly anonymous data. Any free text (comments on observations, observations with text values, or any user-entered text anywhere else in the database) could accidentally reveal a patient's identity. Whenever possible, avoid sharing patient data publicly (even if you think you have anonymized it). If you want/need to share data publicly, be extremely vigilant in ensuring that the data are truly anonymized.

Ideally, all protected health information should be anonymized, including:

- Names
- All geographic subdivisions smaller than 20,000 people – e.g., address, city, county, precinct, exact zip code, and equivalent geocodes.
- All months and days of dates directly related to the individual. This includes birth date, admission dates, discharge dates, dates of death, encounter dates, and observation dates.
- Telephone numbers
- Fax numbers
- E-mail addresses
- National numbers (e.g., social security number, national ID)
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers
- Device identifiers
- URLs
- IP addresses
- Biometric identifiers
- Full face photographic images
- Any other unique, identifying number, characteristic, or code

In OpenMRS, this means that not only names and birth dates get scrubbed, but all dates (including dates on encounters & observations) should be stripped of month & day and any internal identifiers (e.g., patient_id, person_id, etc.) must be randomly changed. Identifiers and dates should **not** be adjusted relative to their original value (e.g., adding or substracting a specific number), since patterns (like differences between dates or identifiers) could be used to re-identify patients.