Operationalizing *Clinical Natural Language Processing* (CNLP) for Cancer Registries

The task of preparing a Cancer Registry for the use of Clinical Natural Language Processing (CNLP) is an important process that will have significant effect in the workflow processes of staff.

The addition of automation to managing text based records frees up CTRs from repetitive drudge work and increases record throughput by building a shared workflow.

Successful implementation should bring greater productivity and enjoyment to existing staff and pave the way for them to be engaged in more interesting and complex work.

Implementing CNLP in a Cancer Registry differs between Registries that are new to CNLP and those that already have experience in this area.

Treat operationalizing CNLP as a project and set up a project management team. *It is not just a matter of installing packaged software.*

Registries New to CNLP

Operational

- Activities include:
  - Commit a staff member to be the liaison officer with the CNLP service provider for coding negotiations
  - Setting up the communication mechanisms for data transfer to and from the CNLP service
  - Set up the architecture and implementation for the storage of the data sent in both directions
  - Ensure procedures are in place for CTR corrected data to be sent back to the CNLP service
  - Agree with the service provider on a schedule of language model revisions for ongoing tuning

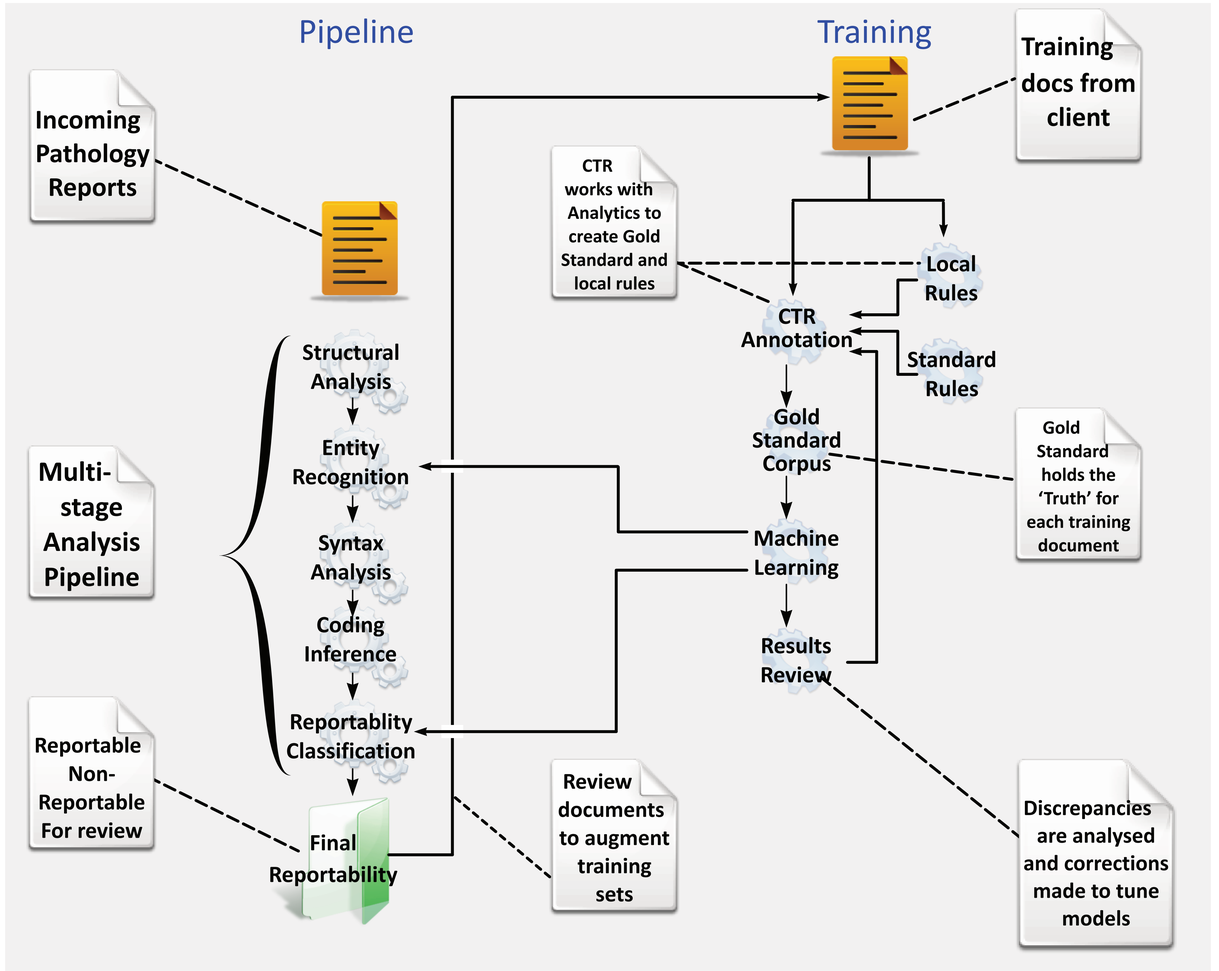
Tuning the Service

- Prepare training materials for the service provider
  - Compile a training corpus
  - Code the training reports accurately for:
    - Case identification
    - The five NAACCR data items
    - Any other required items

The service provider then needs to produce periodic reports on the progress of the tuning of the service e.g. weekly or fortnightly

Post Deployment

- Considerations include:
  - Strategy that defines actions to take if CNLP service is abandoned
  - Setting a regular schedule for auditing the results and providing feedback to the service provider



Registries Experienced in CNLP

How should a cancer registry initiate a new CNLP Process?

i) When the CDC Cancer Surveillance Cloud-based Computing Platform (CS-CBCP) becomes available these considerations will be ensured, but otherwise the Cancer Registry should explore these topics:

1) Ensure in-house CTR software has a suitable interface to receive and display in a suitable interface the results returned from the CNLP service.
2) Ensure the changes by CTRs are stored so validated and computed results are preserved.
3) Optionally, ensure changes created by CTRs are automatically feedback to the CNLP service.
4) Identify the nature of the training set used to create the CNLP service in the first place.
5) Ensure the training set is representative of your own data supply. Representativeness is defined over two major variables - Laboratory variety and Site+Histology variety, i.e. Schema ID.
6) Ensure the reporting of production statistics satisfies your needs and objectives.
7) Identify the quality control processes of the CNLP service and be comfortable with them.
8) Identify the feedback process and timetable to be used for revision of Machine Learning Models.
9) Check local rules for case identification and coding are implementable.
10) Check pathway and mechanism for classification and coding changes in the ML for revisions in NAACCR requirements.

ii) Check the access or availability for expanding data collection beyond current reporting requirements.

iii) Check how the accuracies offered by the service are determined.