

Overview of TIES system for use in the TIES Cancer Research Network (TCRN)

February 4, 2014

TIES is a system for automatically coding and indexing free-text clinical reports including. Surgical Pathology Reports (SPR) are clinical documents which contain diagnostic and prognostic information resulting from surgical and medical procedures in which tissue is removed for microscopic examination. SPRs typically contain patient identifiers within the free-text (names, accession numbers, locations, and dates). The system is capable of processing other kinds of clinical reports. But the current TCRN project only includes SPRs.

History

The TIES System has been in development at the University of Pittsburgh since 2003. The project has been funded as part of a U01 grant from the NCI to develop the Shared Pathology Informatics Network, through a series of contracts for the development of caBIG, as a project for the CTSA informatics core, and by an NCI R01 (1/1/09-12/31/14).

The current project to develop the TIES Cancer Research Network is funded by a U24 grant, with subcontracts to University of Pennsylvania, Roswell Park Cancer Institute, and Georgia Regents University. The funding period is September 25, 2013 - July 31, 2018.

Use of the system at University of Pittsburgh

TIES is currently in use at the University of Pittsburgh and has approximately 270 users. The TIES system is a UPMC-approved clinical system and updates its UPMC Security Plan on a yearly basis. The deployment of the TIES system has an approved IRB Exemption (PRO07050292). The Health Sciences Tissue Bank(HSTB) controls the operation of the TIES system, and creates and administers policies regarding its use. The HSTB Manager reviews user requests for access and authorizes users to specific 'protocols' for access to (1) aggregate de-identified data, (2) record-level de-identified data, or (3) remainder tissue ordering (with IRB approval). Once authorized, the helpdesk provides credentials to the user.

The TIES system relies on an Honest Broker model which is built directly into the software. Honest Brokers (HB) who use the TIES system are able to see identified information related to researcher requests, but only when the HB is within the confines of the University of Pittsburgh Medical Center firewall. From a policy standpoint, only HBs from IRB-approved HB systems may be given Honest Broker credentials in the TIES system. HBs who require Honest Broker credentials for TIES also apply through the HSTB. To be approved, the HB must be within an Honest Broker System that has an IRB Protocol that explicitly includes TIES as one of its data sources. HBs additionally agree to certain terms during the account request process including (1) that they have access to

the data as part of their daily workflow, (2) that they will use TIES only for the purposes of research, and (3) that they will follow all applicable HB policies. HBs with TIES credentials currently include Tissue Bank HBs, Cancer Registry HBs, and other HBs who retrieve tissue or provide other information for researchers using the TIES system. Users are assigned to a *research protocol* in TIES and each *research protocol* is also assigned to an Honest Broker. For any protocols that involve tissue distribution, an HB from the HSTB Honest broker service must be assigned.

Users who access the TIES system must agree to a Usage Agreement which appears at each sign-on to the system. The agreement specifies that (1) they are the individual to whom their credential belongs, (2) that they will not attempt to re-identify, (3) that they will quarantine any data that appears to have identifiers, and (4) that they will limit searches to the area specified in their HSTB or IRB approval. Research users only have access de-identified data.

Envisioned use of TIES across the TIES Cancer Research Network

Previous versions of TIES have had the functionality to use cross-institutionally, but this has never previously been used beyond demonstration purposes. In the TCRN, we envision using the TIES system to support cross institutional collaboration between collaborating investigators across institutions or to facilitate large scale multi-center programs such as the SPOREs.

- The network will be composed of four institutions - University of Pittsburgh, University of Pennsylvania, RPCI and GRU. They will be added to the network once all agreements are in place and TIES is successfully implemented.
- Users will be credentialed at their local institutions, possibly by the Tissue Bank at each institution (analogous to the Pitt implementation).
- At the point that a protocol is created which specifies the request to access data or tissue from another institution – the protocol is made visible to the TIES Administrators at the member institutions through the TIES system.
- Each TIES Administrator (typically the head of an HB service) examines the request, protocol, IRB documents, (DUA, MTA) and makes a determination as to whether the researchers on that protocol may be granted access to the institutions data and or to request tissue as a preliminary user, research user, or tissue user.
- If the TIES Node Administrator grants access, then the researchers on that protocol will be able to access de-identified data from that institution, and (if they have an MTA) they will be able to request tissue.
- Individuals with HB credentials to TIES will be able to view identified data only from their own institution and only when they are within their own firewalls.

Regulatory Compliance

Participation of institutions in the TCRN will require that institutions sign a Network Agreement that includes DUA, agreement to use the UBMTA, and other provisions. This agreement is currently being negotiated with the network sites participating in the pilot projects.

Once the Network Agreement is in place, it is expected that investigator access to data only will require at least completion of the Authorized User Agreement, and scientific review and approval of the providing institution (done on a case by case basis through the TCRN Oversight Committee at each institution). This should be able to occur quite rapidly, and will permit investigators to use data for studies done across network sites, and to determine quickly whether tissue studies are feasible based on cohorts across network sites.

Once the Network Agreement is in place, it is expected that investigator access to data and tissue will also require completion of a MTA based on the UBMTA. It is at this point that the cost recovery mechanism across institutions would be specified. The request is also expected to be vetted through existing institutional mechanisms such as Tissue Use Committee (TUC). IRB requirements may vary dependent on individual institutions, and whether tissue can be considered to be 'deidentified'. Although this process will take longer than data only access, it is expected to meet all regulatory requirements while minimizing burden on investigators (e.g. because the MTA will use the UBMTA per the network agreement, and because the investigator will already know that the study is feasible based on the data preparatory to research before the request is ever generated).

De-Identification

Researchers using the TIES system will gain access to deidentified data only. TIES uses the De-ID system to remove identifiers from free-text. University of Pennsylvania, RPCI, and GRU will also use De-ID. It is possible to use other de-identifiers with TIES.

Separation of Identified and De-Identified Information

TIES maintains a clean separation of de-identified and identified information. Identified and de-identified datastores are maintained in distinct databases. The databases are behind the institution firewall and are only accessible through web services. The identified data web services are only accessible from behind the institutional firewall.

The identified datastore includes patient name, date of birth, Social Security Number, Medical Record Number, date of acquisition, as well as the complete unprocessed clinical report. The de-identified datastore contains randomly generated identifiers for MRN and DOB, as well as the de-identified text, and associated document transformations.

Operationally, the TIES services are used to transform the identified data and populate the de-identified datastore. The TIES de-ID service polls the identified datastore for reports that have not been de-identified, calls the de-ID software to replace HIPAA identifiers within the text, creates the key pair linking the de-identified and identified documents and deposits this key pair in the identified datastore. The coding pipeline polls the de-identified datastore for reports that have not been coded, codes them, and populates the remaining fields of the de-identified datastore.

Encryption

Because of the issues related to incomplete de-identification, we prefer that all communications between institutions be encrypted. We use the Globus Security Infrastructure (GSI) which uses asymmetric public private key cryptography. The version of the system being deployed for TCRN uses RSA 1024 key encryption.

Role-based access to TIES

TIES uses role-based access. The client application is available to users through four “portals”, each of which provides access to the functionality required by a specific role:

- The Admin Portal is available to users who authenticate as TIES Local Administrators. It provides the functionality needed to manage local user accounts as described in the TIES use cases.
- The Honest Broker Portal is available to users who are authorized to be Honest Brokers. It provides the ability to search for identified data and to process orders.
- The Researcher Portal is available to users who are authorized to act as Cancer Researchers. It provides the ability to search for de-identified data, create case sets, and submit orders.
- The Preliminary User Portal is available to users who are authorized to act as Aggregate Data Users. It provides the ability to search for reports but limits the result set to aggregate information in the form of charts and tables. No individual record level data access is granted.

Ordering Tissue

The TIES software supports the ordering of tissue. Users will be able to select cases that they wish to obtain tissue from, and submit this request to the relevant individuals controlling access through the TIES system.

The honest brokers can then update the availability status of the order after considering factors such as:

1. Researcher is affiliated to the same organization as the honest broker.
2. Researcher is affiliated to an organization with which the honest broker’s organization has a materials transfer agreement.
3. Researcher has appropriate IRB approval.
4. Tissue being requested is available for distribution
5. Tissue being requested fits constraints of existing guidelines or work processes- such as approval by the Tissue Utilization Committee or other institutional mechanisms to be decided by each institution.

The researcher, after looking at the updated availability status of his order items can decide to either confirm the order for shipping, or modify the order and resubmit.

The system will support ordering of tissue across multiple organizations, however it will not make any guarantees of actual tissue being delivered to the researcher. The final control of dispersion of tissue lies with the organization's Honest Brokers/Tissue Banks.

Identity Provisioning

Provisioning of user identities is an important issue, because of the IRB requirements of each institution. It is expected that some form of IRB approval (which may simply be a 102F form – No Human Subjects Research Exemption) may be needed to gain access to the TIES system. A separate IRB approval (usually either exempted or expedited review at most institutions) will be required to gain access to tissue. Therefore it is imperative that the task of provisioning users be restricted to the home institution. All users will be required to sign an Authorized User Agreement when they gain access to the network.