Modeling the NPCR Restricted Access Data Set Release Process

2005 NAACCR Conference
Thursday, June 9th, 2005
Boston, MA
Contact Information

- David Lyalin, PhD – Presenter
  Northrop Grumman - CDC Information Technology Support, Atlanta, GA
  dil8@cdc.gov   (678) 530-3583

- KA Gerlach, MPH, CTR¹
- C Friedman, DO¹
- JD Rogers
- HK Weir, PhD¹
- L Almon, MPH¹

¹Centers for Disease Control and Prevention (CDC), Atlanta, GA
Background

- National Program of Cancer Registries (NPCR) registries report the data annually to CDC since 2001.
- NPCR receives data from 45 states, DC and 3 US territories.
- Following data submission, the registries receive a Data Evaluation Report.
- High quality incidence data are used for United States Cancer Statistics and State Profiles and other data release products.
- In order for NPCR to provide researchers with cancer incidence data, a due process for data release has to be established.
- Existing cancer data release processes, could provide various elements for possible re-use with the NPCR process under design.
Purpose

The workgroup has been formed to develop a plan to share NPCR-CSS (or USCS) data with qualified researchers and to develop recommendations for the data release process.
Challenges

Participants are:

- Very limited in time to dedicate to project
- Have diverse perspectives
- Have limited experience in using formal methods to analyze and document business processes
Approach

- Incremental, consensus-based analysis and recommendations development process
- Facilitated modeling sessions – six sessions (two hours each) in the two and a half months period
- Preparatory “off-line” work of a Business Analyst (analysis of processes and development of modeling components) and “on-line” work of the Experts (review and critiques of these components)
- Business Modeling techniques
- Facilitation techniques
- UML – Unified Modeling Language
Why Business Modeling?

- Business modeling and facilitation techniques:
  - Helps us to communicate and to reach a consensus.
  - Helps us to document our business in a standard engineering “blueprint”.
  - Helps us to reduce a complexity of our business by reflecting it in different types of diagrams and other artifacts.
  - Provides requirements for IT systems development, makes sure that business needs will drive technology solutions, and not the other way around.
How to put a model together: facilitated session

1. Discussing
2. Brainstorming
3. Reaching consensus

As-Is Business Model 1
As-Is Business Model 2
As-Is Business Model N

Cancer Surveillance Process, Registry 1
Cancer Surveillance Process, Registry 2
Cancer Surveillance Process, Registry N

Facilitation Team
SMEs Subject Matter Experts
Business Analysts / Modelers

JAD facilitated workshop session(s)

Contribute knowledge of cancer surveillance
Contribute expertise in systems modeling
Provides expertise in organizing and conducting JAD group activities

Best practice - based Cancer Surveillance Process

Capture consensus "best practice" in
Analyze at

To-Be Business Model
Results

- Workgroup developed recommendations for the NPCR RADS release process in the form of a business model:
  - Step-by-step process descriptions
  - Business rules
  - Visual presentation in the form of a process diagram
  - Comparison table for different data release processes

- This model identifies major functional areas of cancer data release process, specific sequence of steps, and responsibilities of process participants.

- These recommendations provide a set of general technology-neutral functional requirements and NOT a description of specific solution/design or implementation.
Analysis of Existing Processes

Major Functional Areas
- Submit request for data, including
  - Study data description
  - Get application forms
  - Track requests for data
- Approve request for data, including
  - Get registries’ approval (consent)
  - Appeal rejection of request
- Access cancer data
- Maintain data access licenses

Process participants
- Researcher
- Request Coordinator
- State Registry
- Approval Body (in case of NPCR process - Data Release Workgroup)
- IRB
- IT Contractor
Use case diagram: Release of Restricted Access Cancer Data (typical functionality)

Goal
(High-level use case)

Major Activities
(use cases)

Additional Activities
(use cases)

Responsible Parties
(Actors)

Revision: 04-05-04

Goal
(High-level use case)

Major Activities
(use cases)

Additional Activities
(use cases)

Responsible Parties
(Actors)

Revision: 04-05-04
## Comparison of Cancer Data Release Processes (fragment 1)

<table>
<thead>
<tr>
<th>Functional Area</th>
<th>NPCR (under design)</th>
<th>NAACCR</th>
<th>SEER-Medicare</th>
<th>SEER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submit request for data</td>
<td>Request is submitted to Request Coordinator. ~90% of Request Coordinator’s responsibilities will be fulfilled by a Contractor, and ~10% - by CDC employee.</td>
<td>Request is submitted to NAACCR DEP Committee. Forms are available on-line.</td>
<td>Request is submitted to Admin (Contractor), who c/c to Coordinator. Forms are available on-line.</td>
<td>Request is submitted to SEER via the form at the web site. A personalized SEER Public-Use Data Agreement will be created for each request.</td>
</tr>
<tr>
<td>Track requests for data</td>
<td>Yes, it is necessary to keep a log of all actions pertinent to a request. Requests tracking IT system will have to be developed.</td>
<td>?</td>
<td>MS Access database, run by Contractor, accessible by Coordinator</td>
<td>In place; details unknown.</td>
</tr>
<tr>
<td>IRB Approval</td>
<td>External IRB approval or “exempt research” papers required for non-CDC researchers.</td>
<td>NAACCR IRB review is a part of the approval process</td>
<td>IRB approval is not required, but almost all requests have it anyway</td>
<td>Not required</td>
</tr>
</tbody>
</table>
### Comparison of Cancer Data Release Processes (fragment 2)

<table>
<thead>
<tr>
<th>Functional Area</th>
<th>NPCR (under design)</th>
<th>NAACCR</th>
<th>SEER-Medicare</th>
<th>SEER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approve request for data</td>
<td>Data Release Workgroup</td>
<td>1) Data Evaluation and Publication Committee (CINA Deluxe sub-Committee?)</td>
<td>1) Request is sent to a single registry selected in revolving order (as per telephone interview)</td>
<td>SEER; details unknown.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) NAACCR IRB</td>
<td>2) Representatives of NCI, SEER, and CMS review and approve (as per web site)</td>
<td></td>
</tr>
<tr>
<td>Appeal process</td>
<td>Through Request Coordinator</td>
<td>?</td>
<td>Not formally defined; need for appeal never aroused.</td>
<td>?</td>
</tr>
<tr>
<td>Get permissions from registries</td>
<td>Not required for many (44?) registries. For the rest of registries – Researcher can be tasked to obtain consents (?) Need to survey states about willingness to participate in RADS and give authority to Data Release Group. The other option is to get permission project by project (unacceptable, by Group’s opinion)</td>
<td>Each registry gives consent for release of data on a project-by-project basis.</td>
<td>A single registry approves a request in behalf of all registries</td>
<td>Not required</td>
</tr>
</tbody>
</table>
## Comparison of Cancer Data Release Processes (fragment 3)

<table>
<thead>
<tr>
<th>Functional Area</th>
<th>NPCR (under design)</th>
<th>NAACCR</th>
<th>SEER-Medicare</th>
<th>SEER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization of cancer data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Charging for data</td>
<td>No</td>
<td>No</td>
<td>Yes – SEER contractor has to be compensated for preparing custom data by outside researchers</td>
<td>No</td>
</tr>
</tbody>
</table>
| Access cancer data         | Electronic access to data thru SEER*Stat | Electronic access to data thru SEER*Stat | Data file is sent to Researcher | 1. Use SEER*Stat to access the data through your Internet connection (SEER*Stat's client-server mode)  
2. Have CDs containing the data and SEER*Stat software shipped to you. These include the binary and text versions of the data.  
3. Download two compressed files containing the CD images. SEER*Stat, the binary data, and the text data are available for download. |
Comparison of Cancer Data Release Processes (fragment 4)

<table>
<thead>
<tr>
<th>Functional Area</th>
<th>NPCR (under design)</th>
<th>NAACCR</th>
<th>SEER-Medicare</th>
<th>SEER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain data access licenses</td>
<td>Revoked after one (1) year. Request for extension (continuation form) can be filed after that.</td>
<td>Revoked after one (1) year. Request for extension can be filed after that.</td>
<td>Don’t have a requirement to destroy the data after certain period of time (used to be 5 years)</td>
<td>?</td>
</tr>
<tr>
<td>Establish Data Research Center</td>
<td>No. Could be considered as a last resort.</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Who can be served?</td>
<td>Stage 1. (Year of 2005) NCI - no SEER-Stat software in use at this point.</td>
<td>NAACCR Researcher (from the NAACCR registry)</td>
<td>Outside and inside (from a SEER Registry) Researchers. No requests for data from outside of USA were made.</td>
<td>Restrictions unknown.</td>
</tr>
<tr>
<td></td>
<td>Stage 2. Researchers from US cancer registries and national partners.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage 3. Any researcher representing a USA institution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requests from outside of the USA will be dealt with if/as they come.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NPCR RADS release process Domain diagram (draft)

- **Qualified Researcher**
  - a type of
  - gets access thru
  - provides ID and password

- **Data Requestor**
  - communicates with
  - responsible for
  - presents request to

- **SEER*Stat Interface**
  - maintains

- **NPCR RADS file**
  - submits

- **NPCR Contractor**
  - reviews and approves/rejects

- **Request Coordinator**

- **Request for Data**
  - log of activities

- **NPCR RADS Requests Form**

- **Research Proposal**

- **IRB Approval**

- **Data Release Workgroup**
<table>
<thead>
<tr>
<th>BP #</th>
<th>Theme / Area</th>
<th>Business Rule statement</th>
<th>High-level business motivation aimed for by the Best Practice</th>
<th>Remarks / Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR01</td>
<td>Available Data</td>
<td>No customization of NPCR RADS file is offered, only pre-build files are available.</td>
<td>Minimize efforts, increase security.</td>
<td></td>
</tr>
<tr>
<td>BR02</td>
<td>Access to data</td>
<td>Data is not given away, only password-protected limited access to data thru a software like SEER*Stat.</td>
<td>Increase security.</td>
<td></td>
</tr>
<tr>
<td>BR03</td>
<td>Access to data</td>
<td>Only Qualified Researchers can get access to NPCR RADS file.</td>
<td>Serve only valid requests for data, reduce amount of requests.</td>
<td></td>
</tr>
<tr>
<td>BR04</td>
<td>Access to data</td>
<td>Qualified Researcher is defined as a Researcher who: a) Can provide a research proposal with approval of non-CDC Federally approved IRB board or “exempt research” papers (for external researchers) or from CDC IRB (for internal – CDC researchers). b) Represents NCI (during the first stage - 2005). Represents cancer registries or national partners (during the second stage). Represents a USA institution (during the third stage).</td>
<td>Serve only valid requests for data, reduce amount of requests.</td>
<td></td>
</tr>
<tr>
<td>BR05</td>
<td>Access to data</td>
<td>Access to NPCR RADS data is granted for a limited period of time – one (1) year.</td>
<td>Increase security.</td>
<td></td>
</tr>
<tr>
<td>BR06</td>
<td>Data use agreement</td>
<td>Researchers may not publish or present findings in which the number of cases in a cell, by registry, is less than five.</td>
<td>To eliminate the potential for re-identification of patients</td>
<td>Based on SEER-Medicare: <a href="http://healthservices.cancer.gov/seermedicare/obtain/use.html">http://healthservices.cancer.gov/seermedicare/obtain/use.html</a></td>
</tr>
</tbody>
</table>
Process participants:
1. Researcher
2. Request Coordinator
3. Data Release Workgroup
4. NPCR Contractor

Preconditions:
What must be true in order for process to be possible to start.
1. RADS is created and ready for release.
2. RADS data description is available to Researchers for review.
3. Forms necessary for requesting the data are developed and available to Researchers.
4. Permissions are obtained for re-release of State registries data in NPCR RADS.
1. The process starts when Researcher, after reviewing RADS data description (see Business Rule BR01), decides to initiate a request for access to NPCR RADS.

2. Researcher obtains NPCR forms necessary for requesting the access to RADS. These forms can be obtained from the NPCR web site or thru the communications with Request Coordinator.

3. Researcher fills out and submits NPCR RADS access forms (BR06) and additional materials, such as IRB (external to CDC) approval of the research project, “exempt research” papers, etc.
1. The process starts when Researcher, after reviewing RADS data description (see Business Rule BR01), decides to initiate a request for access to NPCR RADS.

2. Researcher obtains NPCR forms necessary for requesting the access to RADS. These forms could be obtained from the NPCR web site or thru the communications with Request Coordinator.

3. Researcher fills out and submits NPCR RADS access forms and additional materials, such as IRB (external to CDC) approval of the research project, "exempt research" papers...

4. Request Coordinator receives and processes a data release request package, establishing a log for tracking purposes. All future activities pertinent to the request are logged in by the Request Coordinator.

5. Request Coordinator validates the request based on the Business Rules BR03, BR04. If the request does not pass a validation test, then Request Coordinator rejects the request for access to the NPCR RADS (a standard letter). Process terminates. If additional information is necessary to evaluate a validity of the request, then Request Coordinator inquires additional information from Researcher.

6. Request Coordinator presents a validated request to Data Release Workgroup for approval.

7. Data Release Workgroup considers the request for the access to RADS and approves or rejects it.
4. Request Coordinator receives and processes a data release request package, establishing a log for tracking purposes. All future activities related to the request are logged by the Request Coordinator.

5. Request Coordinator validates the request based on the Business Rules BR03, BR04. If the request does not pass a validation test, then Request Coordinator rejects the request for access to the NPCR RADS (a standard letter). Process terminates. If additional information is necessary to evaluate a validity of the request, then Request Coordinator inquires additional information from Researcher.

6. Request Coordinator presents the validated request to Data Release Workgroup for approval.

7. Data Release Workgroup considers the request for access to RADS and approves or rejects it.

8. Request Coordinator notifies Researcher about the decision regarding the request for access to NPCR RADS (standard letter). If the request is approved, then the process continues. If the request is rejected, then Researcher can appeal.

9. Request Coordinator directs NPCR Contractor to provide Researcher with access to NPCR RADS.

10. NPCR Contractor provides Researcher with user ID and password to access NPCR RADS via standard SEER*Stat software interface (see Business Rule BR02).

11. When the access period is expired, NPCR Contractor revokes the Researcher's access to NPCR RADS (see Business Rule BR05). The process ends.
Overview of the NPCR RADS data release process implementation tasks

**Endorsements and buy-ins**
- Get Endorsement at CDC
- Get endorsements at state registries

**Resources**
- Put in place resources

**Information Technology**
- Deploy SEER-Stat Client/server system
- Implement requests initiation and tracking system

**Timeline**
- 6-9 months
- 18-24 months

**Deployment**
- Deploy the data release process
- Three stages, 1st stage - 2005

Revision: 05-06-04
Conclusions

- A modern approach based on business engineering techniques was used to develop recommendations for the NPCR RADS release process.
- Developed business model facilitates analysis and documentation of the data release process and provides a basis for information technology systems requirements.
- Presented materials are work in progress; at this time not all internal details of SEER and NAACCR processes are known and available for analysis.
Take Home Messages

1) Designed RADS release process expressed in standard engineering formats that support better communication, understanding, and consensus building.

2) Business modeling and facilitation techniques:
   - Provided an adequate support to participants.
   - Helped to reach a consensus among participants.
   - Helped to produce “technology neutral” solutions and recommendations.
Questions?

David Lyalin
dlyalin@cdc.gov

The End