Creation of an Open Source Master Person Index from Proprietary Code
The Open Source "Care Data Exchange" Project

Stuart Turner, DVM[1], Will Ross[2], Scott DuVall, BS[3], Odysseas Pentakalos, PhD[4], Jeff Wallace, MS[5]

[1]University of California, Davis, CA, [2]Mendocino Informatics, Ukiah, CA, [3]University of Utah, Salt Lake City, UT, [4]SYSNET International, Oak Hill, VA, [5]Carpe Occasio LLC, San Diego, CA

1 Problem

From 1998 to 2004 the "Care Data Exchange" (CDE) software was developed as a proprietary product by CareScience for the California HealthCare Foundation (CHCF). In 2005 CHCF asked Forrester Research to study the feasibility of releasing the CDE software assets under a free, open source license. The Forrester report articulated relationships between proprietary and nonproprietary components in the CDE Information Architecture (CIA).

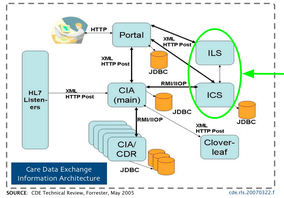
2 Goals

[A] Combine unencumbered CDE components with additional open source components as needed to assemble a Record Locator Service (RLS) built on an explicitly open source fork of the CDE, or the "OS-CDE"

[B] Deconstruct the monolithic CDE Information Architecture and assemble a successor OS-CDE with modular services to demonstrate extensibility and to provide adaptability

3 Project Plan

- Postpone complexity, sensitivity and specificity while focusing on RLS functionality
- Limit the investigation to CDE services for identity correlation and record location
- Remove proprietary platform dependencies (e.g., BEA WebLogic, Oracle)
- Remove proprietary application dependencies (e.g., Cloverleaf, Data Junction)
- Suppress unnecessary internal proprietary components (e.g., UPenn NeuralNet)
- Substitute open source services for proprietary depedencies
- Build on an open source stack (e.g., Linux, JBoss, PostgreSQL)
- Make all services modular for a plugable component system

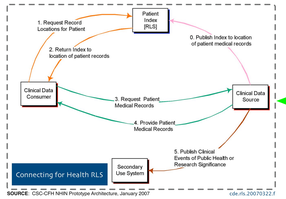


CDE Information Architecture

Together the CDE Identity Correlation Service (ICS) and Information Locator Service (ICS) perform the two key actions of an RLS:

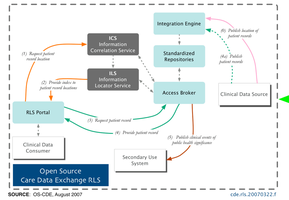
[1] *create a master person index*

[2] *operate a data retreival service*



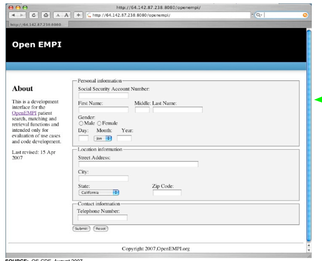
Record Locator Service

The classic Record Locator Service topology as defined by Connecting for Health



Open Source Care Data Exchange

The OS-CDE Record Locator Service stages the ICS and ILS modules (held over from legacy CDE code) in a framework of open source components



OS-CDE Identity Matching Service

The RLS demographic query screen for the OS-CDE Project

4 Results

Expert Panel

In March 2007 the project was initiated with a two day expert panel, featuring half day presentations by Shaun Grannis on RLS practices at the Indianapolis Network for Patient Care (INPC), and by Jim Hazen, the primary software developer on the CDE project for CareScience.

Intellectual Property Analysis

CHCF donated a comprehensive Intellectual Property (IP) Audit of the CDE software code by Palamida. The audit disclosed 27 packages with IP claims. Most were open source licenses, but 8 were commercial proprietary licenses. All encumbered code was either deleted or abandoned.

Patient Matching Algorithm

Dropping in a simple deterministic algorithm demonstrated the architectural modularity of the new OS-CDE. The matching algorithim used in the demonstration:

```
SSN + FN + LN
SSN + YB + MB + DB
SSN + FN + YB + ZIP
FN + LN + YB + MB + DB
```

OS-CDE Package

- Apache HTTP Server
- Apache Tomcat Server
- Java 2 SDK v1.4.2 with NetBeans 5.0
- JBoss
- PostgreSQL

Demonstration

The OS-CDE was demonstrated to CHCF in May 2007 with a custom built file of 3,000 synthetic patient identities. The file was split into two pools of 1,500 identities each. Ten specific patient identities were then loaded into both partitions for discovery by the matching service.

*The OS-CDE White Paper -- **An Open Source Record Locator Service Built with CDE Code** -- was published August 2007 at http://minformatics.com/news/chcf_os-cde_20070803q.pdf*

*This poster presented at **AMIA 2008 Spring Congress**, Phoenix, Arizona*

Correspondence: *swturner@ucdavis.edu, wross@minformatics.com*

Thank You
Shaun Grannis, Dave Minch, Jim Hazen, Jonah Frohlich, Tanya Laino, Mark Tolliver, Jeff Luszcz & Writch.

This project supported by a grant from the

CALIFORNIA HEALTHCARE FOUNDATION

[h]