

Die Probandatenbank Chantal

Torsten Bronger, IEK-5*

13. August 2012

1 Überblick

Chantal ist eine Web-basierte Datenbankanwendung, die für das Datenmanagement von Proben in einem Forschungsumfeld optimiert ist. Daher wurde beim Entwurf von Chantal der Flexibilität eine besonders hohe Priorität eingeräumt. Weitere Design-Schwerpunkte sind einfache Bedienung und Sicherheit.

Hauptzweck von Chantal ist die Zusammenführung der anfallenden Daten der Forschungseinrichtung an einem zentralen Ort. Dabei werden zwei Aspekte abgedeckt. Zum einen werden die Daten gesammelt. Zum anderen werden die Daten für berechtigte Personen zur Verfügung gestellt.

Datensammlung

Die Datensammlung kann auf drei Arten erfolgen:

1. Die Software an der jeweiligen Anlage enthält Chantal-Anpassungen. Das versetzt sie in die Lage, neue Daten *aktiv* an Chantal zu senden.
2. Die Daten werden in Dateien auf einem Verzeichnis geschrieben, auf das Chantal Zugriff hat. In regelmäßigen Abständen scannt Chantal dort nach neuen Daten und pflegt sie ein.
3. Die Daten werden manuell von einem Mitarbeiter in Chantal eingegeben.

(1) ist die bevorzugte Methode, weil hier die Daten bequem, fehlersicher und latenzfrei in die Datenbank geschrieben werden. (2) hat u. a. den Nachteil der Latenz. Und (3) schließlich birgt neben der Unbequemlichkeit der manuellen Eingabe auch die Fehleranfälligkeit in sich. Dennoch wird man wegen der Heterogenität einer Forschungseinrichtung alle drei Methoden brauchen. Hier zeigt sich beispielhaft die Flexibilität von Chantal.

Verfügarmachung der Daten

Der Web-basierte Ansatz von Chantal hat folgende Vorteile:

1. Zugriff von jedem Arbeitsplatz und Betriebssystem aus möglich, wenn gewünscht sogar weltweit (externe Partner, Telearbeit, Konferenzen etc).

*t.bronger@fz-juelich.de

2. Die Kommunikation ist per HTTPS verschlüsselt; ferner kann pro Mitarbeiter die Berechtigung auf Anlagen bzw. Proben eingestellt werden.
3. Verbesserungen in der Software sind ohne Update sofort für jeden verfügbar.
4. Man kann Links und Bookmarks mit anderen Kollegen austauschen.
5. Per HTTP kann Anlagen-Software oder Programme von Mitarbeitern bidirektional mit Chantal kommunizieren.

2 Konzept

Im Kern besteht eine Chantal-Installation aus zwei großen Tabellen, den *Proben* und den *Prozessen*. „Prozeß“ ist hier ganz allgemein zu verstehen als etwas, das mit Proben gemacht werden kann. Das kann eine Deposition sein, eine Ätzung oder eine Messung.

Das Probendatenblatt ist dann die chronologische Auflistung der Prozesse, die mit der jeweiligen Probe durchgeführt worden sind. Dadurch wird übersichtlich dokumentiert, was mit der Probe geschehen ist. Das Splitting von Proben ist möglich; dabei enthalten die Probendatenblätter der Tochterproben auch die Daten der Mutterprobe bis zum Zeitpunkt des Splits.

Jede Prozeßart wird eigens programmiert. Man geht dabei von Schablonencode aus, so daß die Programmierarbeit auf ein Minimum reduziert wird. Zwar macht dies Programmierkenntnisse nötig, um neue Anlagen zu Chantal hinzuzufügen. Allerdings ist unsere Erfahrung, daß die enorm gesteigerte Flexibilität diesen Aufwand rechtfertigt, ja geradezu nötig macht.

3 Weitere Leistungsmerkmale

Statusmeldungen zeigen den Benutzern an, welche Anlagen zur Zeit stark belastet bzw. ganz ausgefallen sind und für wie lange.

Auftragslisten: Benutzer können Aufträge für z. B. Messungen in Chantal einstellen. Die Operateure sehen eine Übersicht über offene Aufträge und können die Benutzer über den Fortschritt auf dem laufenden halten.

Benachrichtigungen: Chantal stellt jedem Benutzer einen sog. Newsfeed zur Verfügung, damit dieser über Änderungen informiert ist. Der Newsfeed kann mit Email-Programmen, Browsern oder Smartphones abonniert werden und umfaßt Änderungen an Proben, Prozessen, Statusmeldungen und Aufträgen.

Externe Operateure sind z. B. Projektpartner oder Fremdfirmen, die im Auftrag Prozesse durchführen oder Proben liefern. Sie können in Chantal erfaßt werden.

Probenserien werden genutzt, um Proben zu gruppieren, aber auch, um Ergebnisse (z. B. Tabellen oder Plots) einer Serie von Proben zuordnen zu können.

Authentifizierung von Benutzern kann über einen zentralen Exchange/LDAP-Server, z. B. den vom FZJ, geschehen. Damit wird die Benutzerverwaltung nahezu vollständig an die zentrale IT abgegeben.

Quellcode: Der vollständige Quellcode von Chantal, den Hilfsprogrammen und den spezifischen Anpassungen wird der jeweiligen Forschungseinrichtung zur Verfügung gestellt, sowohl zur Einsicht, als auch zur weiteren Anpassung.

Elektronische Laborbücher für alle Anlagen.

Plots und andere Formen der Datenaufbereitung im Probandatenblatt.

Mehrsprachigkeit: Zur Zeit ist Chantal in Englisch und Deutsch verfügbar. Die Infrastruktur für weitere Sprachen ist vorhanden, es müßten sich nur Übersetzer finden.

4 Sicherheit

Das Berechtigungskonzept von Chantal stellt einfache Regeln auf, wer Zugriff auf eine bestimmte Probe hat. Proben und Benutzer werden auf sog. „Themen“ verteilt, und nur innerhalb der Themen kann Zugriff stattfinden. Da ein Benutzer in beliebig vielen Themen sein kann, ist so eine flexible und transparente Kontrolle möglich. Privilegierte Mitarbeiter (z. B. alle Gruppenleiter) haben lesenden Zugriff auf alle Proben.

Chantal ist so geschrieben, daß Sicherheit gegenüber böswilligen Benutzern gewährleistet ist. Zum einen werden sämtliche Daten, die an die Datenbank übersendet werden, auf Validität und Berechtigung überprüft, zum anderen kommt eine Programmbibliothek zum Einsatz, die bewußt Sicherheit vor Performanz stellt.

5 Recherche

Es existieren mehrere Möglichkeiten, in der Probandatenbank nach Dingen zu suchen. Zentrale Funktion ist die sog. „erweiterte Suche“. Hier kann man komplexe Suchanfragen absetzen, die nahezu beliebige Verschachtelung zulassen. Beispielsweise kann man nach Proben suchen, die in Meßapparatur A *und* in Meßapparatur B dieses Jahr gemessen wurden, und die in der Deposition zusammen mit einer Probe auf Aluminium-Substrat hergestellt wurden.

Von-Bis-Angaben sind möglich, ebenso ein Export der Suchergebnisse in Form einer Excel-Tabelle.

Da diese erweiterte Suche sehr umfassend ist, ist es möglich, für bestimmte, immer wiederkehrende Anforderungen spezielle Suchfunktionen zu implementieren. Außerdem können programmierfreudige Mitarbeiter Chantals HTTP-Schnittstelle für Recherchen in der Datenbank nutzen.

6 Server

Theoretisch kann die Chantal-Software auf jeder Hardware installiert werden, auf der ein HTTP-Server laufen kann. Im folgenden sei eine sinnvolle Installation skizziert, nämlich die des IEK-5.

Der Server ist ein Hochverfügbarkeits-Cluster aus zwei gewöhnlichen Arbeitsplatz-PCs. Wenn einer ausfällt, übernimmt der andere innerhalb von 15 Sekunden. Der Server wird einmal pro Minute von einem anderen Gebäude aus auf Erreichbarkeit getestet. Ferner tun zwei USVs ihren Dienst. Sollte irgendetwas ausfallen, bekommt der Administrator eine

Email zugesandt. Damit haben wir über einen Zeitraum von zwei Jahren eine zeitliche Verfügbarkeit von 99,96 % erreicht.

Da ausschließlich Freie Software eingesetzt wird, belaufen sich die Sachkosten einmalig auf ca. 200 € und monatlich auf 100 €. Wirklich signifikant sind also nur die Personalkosten von 3–10 Mannwochenstunden, um den Serverbetrieb zu gewährleisten. Allerdings muß der Administrator, der das verrichtet, sich auch im Quelltext von Chantal auskennen.

Die Performanz von Chantal ist sehr gut. Die Antwortzeit im Browser ist fast immer kleiner als eine Sekunde. Der IEK-5-Server könnte unseren Schätzungen zufolge 5 000 Mitarbeiter bedienen und wäre bei Bedarf sehr leicht auszubauen. Durch einen dreistufigen Cache ist in vielen Fällen die Reaktionszeit in schnellen Browsern quasi instantan. Unseren Messungen zufolge liegt die Cache-Effizienz bei 40–60 %.

Um Datenverlust zu vermeiden, werden stündlich Backups der Datenbank gemacht und auf ein externes System kopiert, welches nächtlich zentral gesichert wird. Die Backups der letzten 48h werden getrennt aufbewahrt.

7 Migration

Um ein Gefühl zu bekommen, wie aufwendig die Migration einer Forschungseinrichtung auf Chantal ist, seien hier die Zwischenschritte dazu kurz dargestellt.

In der ersten Phase sind intensive Vorgespräche nötig, um die typischen Workflows herauszufinden. Chantal wird dann an diese Workflows und sonstigen Besonderheiten der Forschungseinrichtung angepaßt. Exemplarisch dafür sei die Besonderheit des IEK-5 erwähnt, daß hier Proben nach bestimmten Prozessen einen neuen Namen bekommen.

In der zweiten Phase der Migration werden die Anlagen eine nach der anderen zu Chantal hinzugefügt. Das ist keine alles-oder-nichts-Situation. Chantal kann ohne Reibungsverluste produktiv genutzt werden, auch wenn dauerhaft etliche Anlagen ohne Chantal-Integration weiterbetrieben werden.

In der Chantal-Installation im IEK-5 haben wir bislang 35 Prozeßarten (Anlagen, Meßsysteme etc) implementiert; das sind ca. 50 %. Eine neue Anlage hinzuzufügen nimmt wenige Tage bis 2 Wochen in Anspruch. Sollen auch Altdaten importiert werden, hängt der Arbeitsaufwand sehr von der Qualität der sog. Metadaten jedes Datensatzes ab (Zeitstempel, Operateur, Probenname, Parameter). Hier ist mit einer bis mehreren Wochen zu rechnen.