PIAF - Planungs-, Informations- und Auswertungssystem für das Feldversuchswesen

Rupert Pfeiffer, Münster
Gregor Zink, Münster
Dr. Karsten Block, Soest
Prof. Dr. Ulrich Streit, Münster

Abstract

PIAF is a computer-based system for the statistical analysis the results of field trials. It is designed in a Client-Server-architecture using the SAS-system for statistical functions and a SQL-database-system for data management. The two parts of the system are connected by an interface that follows the Open Database Connectivity -(ODBC)- standard.
The possibility of using the SAS-system together with any database-system that offers an ODBC-interface facilitates the use of different hardware-platforms as well as of different database-systems.

Zusammenfassung

Das Planungs-, Informations- und Auswertungssystem für das Feldversuchswesen (PIAF) ist ein Programmsystem in Client-Server-Architektur mit einer SAS-Applikation zur Datenauswertung und einem SQL-Datenbank-Server zur Verwaltung der Datenbestände. Die beiden Komponenten sind über eine Open Database Connectivity-(ODBC)-Schnittstelle miteinander gekoppelt.
Durch die Trennung der beiden Komponenten und durch die Verwendung von SAS bzw. eines beliebigen ODBC-fähigen Datenbanksystems wird eine maximale Flexibilität in Hinblick auf zu verwendende Hard- und Software erreicht.

1 Einleitung

Das Projekt PIAF ist ein Modellvorhaben, das vom Bundesministerium für Ernährung, Landwirtschaft und Forsten (BML) unter Beteiligung der Länder gefördert wird. Es hat eine Laufzeit von 2 Jahren, beginnend im Februar 1995.
Das Modellvorhaben wird von einer Arbeitsgruppe und einer Projektgruppe in Zusammenarbeit mit dem Bundessortenamt durchgeführt. Die Arbeitsgruppe besteht aus Experten des Feldversuchswesens der verschiedenen Bundesländer, ihre Aufgabe ist die Spezifikation des Systems. Die Koordination hat dabei der Vertreter Nordrhein-Westfalens.
Die Implementierung des Systems obliegt der Projektgruppe. Sie besteht aus vier teilzeit-beschäftigten Mitarbeitern des Instituts für Agrarinformatik an der Universität Münster. Der Austausch zwischen den beiden Gruppen erfolgt auf regelmäßigen Arbeitstagungen.
Ziel des Projekts ist es, ein Werkzeug zur Verwaltung von Feldversuchsdaten und zu deren statistischer Auswertung zu erstellen. Dieses Werkzeug soll bundesweit eingesetzt werden und so die Vergleichbarkeit von Auswertungsergbnissen und die Austauschbarkeit von Versuchsdaten gewährleisten.

2 Ausgangssituation und Anforderungen

Sinn von Feldversuchen ist es, Erkenntnisse über die Eigenschaften unterschiedlicher Sorten von Kulturpflanzen, über die Auswirkungen verschiedener landwirtschaftlicher Techniken oder über die Wirksamkeit unterschiedlicher Düngungs- oder Pflanzenschutzstrategien zu gewinnen. Sie werden im wesentlichen von Behörden auf Landesebene, z.B. Landespflanzenschutzämtern oder Landwirtschaftskammern durchgeführt, aber auch von Bundesbehörden wie dem Bundessortenamt (BSA) oder der Biologischen Bundesanstalt (BBA) sowie von der Industrie. Die Auswertung der Versuchsdaten erfolgt z.Zt. mit unterschiedlichen Hilfsmitteln. In einigen Ländern werden ältere Programme auf unterschiedlicher Basis (Fortran, Pascal, Clipper) eingesetzt, in anderen zusätzlich Standardsoftware (SPSS, SAS, PLABSTAT). Das einzige System, das in einem größeren Bereich, nämlich den neuen Bundesländern, eingesetzt wird, ist das Pascal-Programm EFDAS. Dieses Programm entspricht jedoch nicht den Anforderungen in Hinblick auf Modularität, Erweiterbarkeit und Wartbarkeit, die ein System erfüllen muß, wenn es die zukünftigen Aufgaben erfüllen soll.
Der Umfang von Feldversuchen steigt, weil z. B. zunehmend neue Sorten von Kulturpflanzen aus anderen EU-Ländern verfügbar werden und somit die zu überprüfenden Sortimente größer werden. Daraus folgt, daß auch die Werkzeuge zur Auswertung der größeren Datenmengen mächtiger werden müssen und daß eine koordinierte Durchführung und einheitliche Auswertung solcher Versuche durch mehrere Institutionen zu unterstützen ist. Außerdem bekommen Versuche mit integrierten Anbauverfahren und mit nachwachsenden Rohstoffen immer größere Bedeutung. Da diese Versuche nur mit hohem Aufwand durchgeführt werden können, wird eine Kooperation zwischen verschiedenen Versuchsanstellern immer notwendiger.
Durch einen Austausch von Versuchsdaten zwischen den einzelnen Versuchsanstellern werden folgende Wirkungen erzielt:

- Großangelegte Versuchsreihen können von mehreren Versuchsanstellern gemeinsam durchgeführt werden,
- Daten eines Versuchsanstellers können durch Daten einer benachbarten entsprechenden Institution ergänzt werden,
- überregionale Auswertungen können durch das Kombinieren von Daten unterschiedlicher Institutionen effizient durchgeführt werden,
- Die Kosten für das Versuchswesen lassen sich senken, weil die mehrfache Untersuchung gleicher Fragestellungen eingeschränkt werden kann.

Eine der wichtigsten Anforderungen an das PIAF-System ist es, die angestrebte Austauschbarkeit von Versuchsdaten zu ermöglichen. Dazu ist es jedoch notwendig, für alle Beteiligten eine gemeinsame Basis in bezug auf die Verwaltung und die Auswertung von Versuchsdaten zu schaffen. Dies wird erreicht, wenn die Versuchsansteller ihre Daten mit dem gleichen System verwalten und auswerten. Damit diese Standardisierung erreicht werden kann, ist eine bundesweite Initiative notwendig und dies ist der Grund für das des BML, sich für ein Projekt im Bereich des Feldversuchswesens zu engagieren.
Der Leistungsumfang des PIAF-Systems im Rahmen des Modellvorhabens beinhaltet die Erstellung eines Auswertungssystems, das den Anforderungen aller beteiligten Länder genügt sowie der dazugehörigen Datenverwaltung. Um ein sinnvolles Arbeiten mit diesem System sicherzustellen, sind außerdem Import- und Exportschnittstellen vorgesehen.
Nicht Bestandteil des Modellvorhabens sind Module zur Datenerfassung und zur Versuchsplanung sowie der Bereich des Informationssystems. Diese Komponenten könnten im Rahmen eines Folgeprojekts entwickelt werden. Es ist jedoch auch denkbar, daß einzelne Institutionen diese Komponenten für sich selbst ihren individuellen Ansprüchen gemäß entwickeln. Die dafür notwendigen Spezifikationen werden im Rahmen des Modellvorhabens verfügbar gemacht.

Um die von den verschiedenen Institutionen an das System gestellten Anforderungen erfüllen zu können, ist es notwendig, eine gemeinsame Spezifikation in Hinblick auf die Funktionalität des Systems zu treffen (z.B. welche Auswertungsverfahren implementiert werden müssen). Es müssen Standards bezüglich der Datenschnittstellen definiert und genormte Schlüsselsysteme festgelegt werden. Außerdem muß das System in der Lage sein, alle relevanten Varianten von Feldversuchen abzubilden und auszuwerten sowie sie zu harmonisieren bzw. gegebenenfalls ihre Unverträglichkeit festzustellen.
Eine ausreichende Leistungsfähigkeit des Systems kann nur sichergestellt werden, wenn hinreichend mächtige Werkzeuge ausgewählt werden. Dabei ist insbesondere zu berücksichtigen, daß durch die Möglichkeit des Datenaustausches sehr große Datenbestände entstehen können.
Um die Kosten der Anschaffung und des Betriebs von PIAF zu senken, soll möglichst auf bestehende Ressourcen von Hard- und Software beim Anwender zurückgegriffen werden.

3 Systemarchitektur und Werkzeuge

Um die Anforderungen zu erfüllen ist es sinnvoll, für PIAF eine Client-Server-Architektur zu entwerfen. Auf diese Weise ist es möglich, mit verschiedenen spezialisierten Werkzeugen den Server für die Datenverwaltung und den Client für die Datenauswertung getrennt zu entwickeln. Will man vorhandene Strukturen nutzen, bietet es sich an, im Bereich des Servers auf Standards zurückzugreifen. Einen solchen Standard stellt die ODBC-Schnittstellenkonvention mit einer an SQL angelehnten Abfragesprache dar.
Beschränkt man sich auf diesen Standard, so kann als Grundlage des Servers jedes Datenbankmanagementsystem (DBMS) verwendet werden, das als ODBC-Server fungieren kann. Bei einem Betrieb im Netz ist außerdem weitgehende Hardware-Unabhängigkeit sichergestellt.
Zur Implementierung des Clients für die Datenauswertung ist es sinnvoll, ein Werkzeug zu wählen, das bereits im Bereich der Auswertung von Feldversuchsdaten verbreitet ist, das die Erstellung von einfach zu bedienenden Applikationen erlaubt und das auf verschiedenen Hardware-Plattformen verfügbar ist. Ein solches Werkzeug ist in Form des Statistikpakets SAS gegeben.
Auch die Schnittstellen zum Im- und Export sollten einem Standard folgen, damit Kopplungen zu anderen Systemen leichter hergestellt werden können. Der Schnittstellenstandard, der im PIAF - Projekt Verwendung findet, folgt der 'Agricultural Data Interchange Syntax' (ADIS), die als ISO - Norm vorliegt.
Die Systemarchitektur von PIAF ist aus Abb. 1 ersichtlich. Man erkennt die getrennten Bereiche des Clients und des Servers, die über die ODBC-Schnittstelle miteinander kommunizieren und gemeinsam mit ihr das PIAF-System bilden. Der Kontakt zu Fremdsystemen erfolgt über eine ADIS-Schnittstelle. Ein Fremdsystem kann dabei natürlich auch ein anderes PIAF-System sein.
Da Server und Client getrennt entwickelt werden, mit unterschiedlichen Werkzeugen und

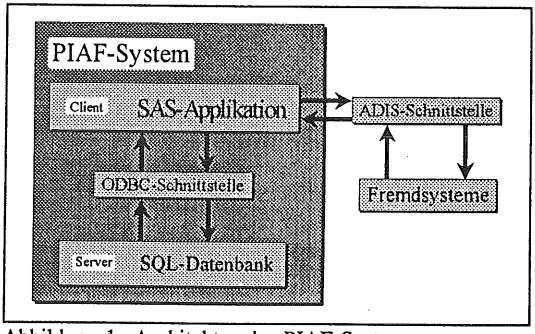

Abbildung 1: Architektur des PIAF-Systems

jeweils eigener Spezifikation, sollen sie im folgenden getrennt beschrieben werden. Auch der Bereich der Im- und Exportschnittstellen ist in einem eigenen Kapitel behandelt.

3.1 Das PIAF-Datenbanksystem

Das Datenbanksystem (DBS) innerhalb von PIAF muß in der Lage sein, Daten von unterschiedlichen Versuchsanlagen zu verwalten und spezifische Eigenheiten bestimmter Versuche abzubilden. Außerdem muß es die Bildung von Gruppen von Versuchen, sog. Serien ermöglichen und dem Benutzer die Eingabe von 'persönlichen' Versuchsparametern erlauben, die nicht mit den gesetzten Standards konform sind. Auch Updates der standardisierten Schlüsselsysteme müssen vorgesehen werden.
Aus diesen Anforderungen resultiert ein komplexes Datenmodell, das in ein relationales Schema umgesetzt wird und aus dem externe Sichten (z.B. für die Datenauswertung oder für den Im- und Export) durch SQL-Abfragen erzeugt werden.
Die Relationen des Datenmodells werden bei der Installation des PIAF-Systems durch SQL-Scripts erzeugt.
Durch die Festlegung, daß man jedes DBMS, das als ODBC-Server verwendet werden kann, auch zur Implementierung des PIAF-DBS benutzen kann, sind viele verschiedene Produkte einsetzbar. Am Institut für Agrarinformatik wird die Entwicklung beispielhaft mit Hilfe von Oracle 7.0 auf einer Sun-Maschine durchgeführt.

3.2 Das PIAF-Auswertungssystem

Der Auswertungsteil des PIAF-Systems wird mit Hilfe des Statistikpaketes SAS 6.10 entwickelt. Dieses Produkt bietet für das PIAF-Projekt verschiedene Vorteile:

- Es können Applikationen mit eigener Benutzeroberfläche erstellt werden. Diese Benutzeroberfläche bietet die wesentlichen Elemente einer graphischen Oberfläche wie Schaltflächen, Menüs usw.
- Diese Applikationen können mit nur geringen Änderungen auf verschiedenen Hardwareumgebungen laufen. Das SAS-Paket ist für verschiedene Hardwareplattformen und Betriebssysteme verfügbar. Erstellte Applikationen sind Source-Code-kompatibel und somit auch in verschiedenen Umgebungen lauffähig. Die Notwendigkeit von Änderungen ergibt sich in Programmbereichen, die außerhalb der SAS-Umgebung operieren. Solche Bereich sind z. B. der Zugriff auf andere Programme (z.B. den ODBC-Manager) oder auf Dateien. Für diese Bereiche müssen für jede Umgebung spezifische Lösungen gefunden werden.
  Für die Entwicklung am Institut für Agrarinformatik wird SAS unter Windows verwendet.
- SAS bietet den Zugriff auf die ODBC-Schnittstelle und zusätzlich den direkten Zugriff auf unterschiedliche gängige DBMS.
- SAS beinhaltet die meisten der benötigten statistischen Auswertungsverfahren in Form von fertigen Prozeduren.

Neben diesen Vorteilen birgt die Verwendung des SAS-Systems jedoch auch Nachteile. Diese sind z.B. in Mängeln der Programmiersprache begründet, in Laufzeitproblemen und im komplizierten Lizensierungsverfahren.
Der Aufbau des Auswertungssystems ist nach Versuchsanlagen gegliedert, besteht also aus einer Sammlung von Teilsystemen für die einzelnen Anlageformen. Durch diese Struktur wird eine schrittweise Entwicklung begünstigt, weil zunächst ein überschaubares System für die

Auswertung von Versuchen nur einer Anlageform erstellt werden kann, das dann sukzessive um weitere Anlageformen ergänzt wird.
Welches dieser Teilsysteme zu verwenden ist, ist i.d.R. durch die auszuwertenden Daten festgelegt.

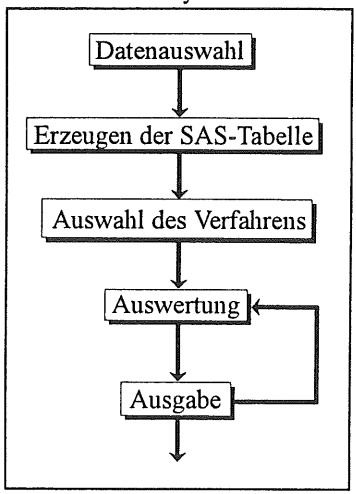

Abbildung 2: Ablaufschema des Auswertungssytems

Für das Auswertungssystem ergibt sich ein Ablaufschema wie in Abb. 2 dargestellt.
Im ersten Schritt wird festgelegt, welche Daten ausgewertet werden sollen. Dazu werden Teile der Datenbank ausgewertet. Durch die Auswahl der Daten ist auch festgelegt, welche Versuchsanlage vorliegt und damit, welche Auswertungsverfahren zur Anwendung kommen dürfen.
Der zweite Schritt besteht aus dem Erzeugen einer SAS-Tabelle, welche die Datengrundlage für die späteren Auswertungen darstellt. Ihre Struktur ist ebenfalls von der Versuchsanlage abhängig. Die SAS-Tabelle wird mit Hilfe von automatisch generierten SQL-Befehlen an das Datenbanksystem erstellt.
Danach folgen die Schritte der Auswertung und der Ausgabe. Die Auswertung beinhaltet dabei die Auswahl eines der möglichen Verfahren. Nach der Ausgabe der Ergebnisse kann ein weiteres Verfahren ausgewählt und durchgeführt werden oder das System wird verlassen.

3.3 Schnittstellen

Schnittstellen, die vorrangig implementiert werden sollen, sind die Im- und Exportschnittstelle nach dem Standard des BSA, damit dort möglichst bald Daten für umfassende Auswertungen gesammelt werden können und damit über diese Schnittstelle Daten aus bestehenden Systemen wie dem DAWEP in das PIAF-System geladen werden können. Weiterhin wird eine neue Schnittstelle implementiert, die zunächst hauptsächlich der Kommunikation zwischen verschiedenen PIAF-Systemen dient.
Bei der Realisierung dieser Schnittstelle wird der ADIS-Standard verwendet. Dieser Standard liegt als ISO-Norm vor und wurde für die Kommunikation zwischen Verwaltungs- und Prozessrechnern entwickelt. Er beruht auf ASCII-Dateien, in denen neben den Daten selbst eine Datenbeschreibung übermittelt wird, deren Bedeutung dem Datenlieferanten und dem Datenabnehmer bekannt sein muß. Das Vokabular dieses 'Data Dictionary' muß von den Teilhabern an der Schnittstelle selbst definiert werden und zwischen ihnen normiert sein. Durch den ADIS-Standard sind nur syntaktische Normen und einige Klassen von Daten festgelegt.

4 Stand der Entwicklung

Die bisherigen Arbeiten im PIAF-Projekt (bis Juni 1995) sind im wesentlichen konzeptioneller Natur. Das Datenmodell wurde entworfen und eine Liste von Auswertungsverfahren, die im System enthalten sein sollen, wurde erstellt. Ein System für die Auswertung von einfaktoriellen, vollständig randomisierten Versuchsanlagen mit einer SAS-Tabelle als Datengrundlage wird

gerade implementiert. Das Datenbanksystem konnte bisher wegen lizenzrechtlicher Schwierigkeiten mit dem Vertreiber des SAS-Systems noch nicht eingebunden werden. Diese Schwierigkeiten verhindern z. Zt. den Zugriff auf die ODBC-Schnittstelle von SAS aus, sollten jedoch bald geklärt sein.

5 Ausblick

Bis zum Anfang des Jahres 1997 wird ein Auswertungssystem fertiggestellt sein, das die Auswertung von Daten aus verschiedenen Versuchsanlagen und außerdem die Verwaltung von Daten einer großen Zahl von Versuchen erlaubt. Die Kombination von Versuchen zu Serien und deren Auswertung wird ebenfalls möglich sein. Es werden Schnittstellen für den Im- und Export von Daten zur Verfügung stehen.
In anderen Bereichen wird das PIAF-System jedoch noch unvollständig sein. Die Komponenten zur Planung von Versuchen und Serien sowie Module zur Erfassung von Versuchsdaten vor Ort sowie der gesamte Bereich des Informationssystems werden fehlen.
Für alle diese Bereiche existieren jedoch bereits jetzt konzeptionelle Ansätze; so sind durch die Festlegung des PIAF-Datenmodells die Datenstrukturen für Datenerfassungs- und Versuchsplanungsmodule schon jetzt vordefiniert.

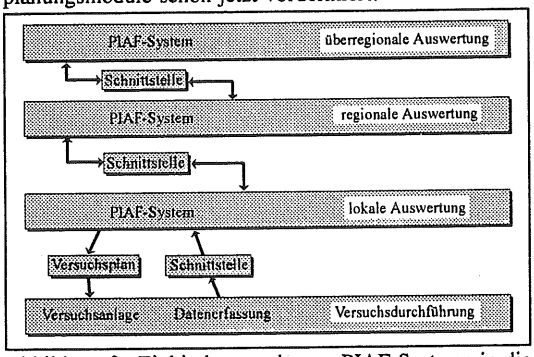

Abbildung 3: Einbindung mehrerer PIAF Systeme in die Versuchsauswertung verschiedener Institutionen

Nach einer Vervollständigung des PIAF-Systems und bei entsprechender Verbreitung ist eine Struktur wie sie Abb. 3 zeigt denkbar. Das PIAF-System wird hier für die Auswertung auf lokaler, regionaler und überregionaler Ebene verwendet. Die Kommunikation geschieht hierbei über Schnittstellen nach dem ADIS-Standard.
Auf der untersten Ebene gibt es zusätzliche Werkzeuge zur Erstellung von Versuchsplänen und zur Erfassung von Versuchsdaten. Für solche Zwecke sind auch spezielle Hardwarekomponenten wie Pen-Computer zur Datenerfassung im Feld und Transponder zur Identifizierung von Parzellen usw. einsetzbar.
Die Informationssystemkomponente kann zum einen für die Verwaltung akkumulierter Versuchsergebnisse benutzt werden. Im Zusammenhang mit mehreren PIAF-Systemen kann sie außerdem als Meta-Datenbank dienen, die Informationen darüber enthält, welche Versuchsdaten bei welchen Institutionen verfügbar sind.