

Grundlegende Anforderungen an das Datenmanagement in interdisziplinären Forschungsprojekten

Claus Mücke¹, Gabriel A. Schachtel¹, Alexander Wehrum²,
Jens Nieschulze³, Wolfgang Köhler¹

¹ SFB 299 Giessen

² Connecta AG Wiesbaden

³ SFB 552 Göttingen

c/o Biometrie und Populationsgenetik

Justus-Liebig-Universität

Heinrich-Buff-Ring 26-32, D-35392 Giessen

claus.mueckschel@agr.uni-giessen.de

Abstract: Set-up and maintenance of central data management systems in interdisciplinary research projects face some typical problems and challenges. Data management should be regarded as an integral and indispensable constituent of any multi-group project and must therefore be included from early on into the conceptual planning. Continuous and close communication between the system developing team and all other subprojects about the needs and options of either side facilitates an efficient and smooth data management, e.g. the collection, storage, and access of project data. The user's integration into involvement in the processes of system development supports the implementation of intuitively understandable interfaces and promotes system acceptance.

1 Datenmanagement: Ziele, Aufgaben und Anforderungen

Der SFB 299 (www.sfb299.de) ist ein von der DFG gefördertes (IS) Forschungsprojekt, das im Jahre 1997 startete und eine maximale Laufzeit von 12 Jahren umfasst. Die Aufgabe des Datenmanagements im SFB 299 besteht in der Unterstützung der Forschungstätigkeit und leitet sich aus einem der Hauptziele eines jeden SFB ab, nämlich Gewinnung von Synergien aus einem interdisziplinären und langfristigen Forschungsansatz. Die Unterstützung des interdisziplinären Aspekts eines laufenden SFB verlangt, die Daten aus allen Teilprojekten nach klaren Regeln zu strukturieren und zu filtern, sie trotz ihrer Vielfalt in einem einheitlichen Format abzubilden und den einzelnen Projekten schnell und leicht zugänglich zu machen. Der langfristige Ansatz eines SFB bewirkt u. a. erhebliche personelle Fluktuationen und methodische Weiterentwicklungen. Diese Veränderungen verlangen vom Datenmanagement eine in die Zukunft gerichtete, dauerhafte Sicherung und korrekte Interpretierbarkeit aller in einem Projekt gesammelter Daten, um möglichst jeden Daten-, Informations- und damit Wissensverlust zu vermeiden. Eine übergreifende Interpretation der meist komplexen Datensätze, die aus unterschiedlichen Disziplinen stammen und für verschiedene Zeitskalen und Raumebenen anfallen, ist eine nicht zu unterschätzende Herausforderung.

Ihr adäquat zu begegnen braucht es ein entsprechendes System von Funktionen und Werkzeugen, die sorgfältig auf die besonderen wissenschaftlichen Fragestellungen abgestimmt sind [E98], [Mü04]. Dabei ist zu beachten, dass sich die modularen Systembestandteile, die miteinander in Beziehung stehen, ergänzen und durch ihren Zusammenschluss eine neue Qualität erhalten. Ebenso ist zu beachten, dass die Möglichkeit der problemlosen Erweiterung um beliebige Parameter gegeben ist.

Neben einer ausgefeilten Datenhaltung und Bereitstellung muss bei der Konzeption und bei jeder Weiterentwicklung ständig die größtmögliche Anwenderfreundlichkeit angestrebt werden, von der die Akzeptanz durch die Nutzer wesentlich abhängt. Dieses Ziel der einfachen Handhabung muss trotz der den Daten innewohnenden Komplexität verwirklicht werden. Während die Inhalte des Informationssystems (IS) zentral verwaltet werden, müssen sie dezentral aus den Projekten eingepflegt werden können. Plattform-unabhängige, browserbasierte Systeme sind gegenwärtig die effektivste Methode den ortsunabhängigen Zugang zum System sicherzustellen. Der vorliegende Beitrag stellt eine Auswahl allgemeiner konzeptioneller und technischer Aspekte und Grundsätze zusammen, die bei der Planung, Einrichtung und im laufenden Betrieb eines projektweiten Informationssystems berücksichtigt werden sollten.

2 Systemarchitektur und Dynamik der IT-Entwicklung

Die lange Laufzeit von SFB stellt wegen der raschen Entwicklung im IT-Bereich besondere Anforderungen an die Systemarchitektur. Wegen sich ändernder Ansprüche und Schwerpunkte der Teilprojekte müssen die verwendeten IT-Systeme entsprechend angepasst werden, was zwangsläufig zu einer hohen Dynamik innerhalb des Datenmanagements führt. Ein einmal in Betrieb genommenes IS muss über die gesamte Laufzeit operabel und so flexibel sein, dass Schnittstellen für Benutzer im Prinzip konstant genutzt werden können, auch, wenn die Systemarchitektur sich grundlegend ändert. Die strategische Entscheidung für eine Systemarchitektur muss zunächst die momentane Leistungsfähigkeit berücksichtigen. Neben der Leistungsfähigkeit ist auch die Investitionssicherheit wichtig, hierbei gilt, dass sowohl Open-Source (OS) als auch proprietäre Produkte vom Markt verschwinden oder in anderen Produkten aufgehen können [WEM04]. In der Regel gibt es kein Software-Produkt, welches allein allen speziellen Rahmenbedingungen eines Großforschungsprojekts gerecht wird. Informationssysteme entstehen meist im Rahmen speziell geförderter Forschungsprojekte mit sehr spezifischen Fragestellungen und Datenaufkommen. Deshalb haben sich insbesondere Großforschungsprojekte in der Vergangenheit kaum dazu entschlossen, fertige System-Lösungen einzukaufen [SR94].

Ein rein monolithischer Ansatz ist somit nicht umsetzbar, jedoch muss über den Grad der Modularität jeweils neu entschieden werden, oft stehen sowohl proprietäre als auch Open-Source-Produkte zur Wahl. Neben den Implementierungs- und Unterhaltungskosten werden im SFB 299 folgende Entscheidungskriterien herangezogen: Skalierbarkeit, Stabilität, Modifizierbarkeit und Kompatibilität mit anderen eingesetzten oder geplanten Produkten. Das IS des SFB 299 basiert auf LAMP-Systemen (*Linux-Apache-MySQL-PHP/Perl*) mit unterschiedlichen Erweiterungen. Der Einsatz dieser Open-Source-Produkte ist seit der Implementierung weitestgehend kostenneutral und hat sich bewährt.

Die Offenheit der Systemkomponenten erleichtert die häufig notwendigen technischen Anpassungen. Durch die Bereitstellung von Erfahrungen und Tipps durch die „Community“ ist ein umfangreicher Support sichergestellt, der die Benutzung kostenpflichtiger Hotlines (z.B. für Linux oder MySQL) erspart. Die Erfahrungen mit dem SFB-IS unterstreichen die Bedeutung eines modularen Aufbaus und der strikten Einhaltung von Normen und Standards; diese garantieren erst eine weitgehende Allgemeingültigkeit, Wiederverwendbarkeit, Austauschmöglichkeit, Erweiterbarkeit sowie einfache Nutzung eines gesamten IS oder ausgewählter Module. Erst die Normen und Standards erlauben es, einzelne Module ohne größeren Aufwand zu verknüpfen, auch, wenn sie völlig unabhängig voneinander entwickelt wurden. Das eröffnet ein größeres Spektrum an Kombinationsmöglichkeiten. Folgende sind einige relevante Internationale Standards: ISO 11179, Metadata Interchange Specification, Dublin Core (überwiegend Dokumente) und FGDC (für Geodaten).

3 Einbindung des Datenmanagements in die Projektplanung

Große, interdisziplinäre Forschungsprojekte leiden oft unter den eingefahrenen, aus der klassischen einzeldisziplinären Erfahrung stammenden Vorbehalten gegen „allzu viel“ Datenaustausch. Unsicherheiten bezüglich der Rechte an den Daten und ihres unautorierten Gebrauchs durch andere behindern den Informations- und Datenfluss und somit den Aufbau einer zentralen Datenhaltung. Klare Vorgaben und Regelungen seitens der Projektleitung bzgl. der Bereitstellung und Nutzung der in den Teilprojekten erhobenen Daten sind für die schnelle Überwindung solch einer Zurückhaltung in Sachen Datenweitergabe unerlässlich. Darüber hinaus muss schon bei der konzeptionellen Planung des Projekts dem Datenmanagement genügend Beachtung geschenkt werden, um die gewünschten synergistischen Effekte zu erzielen. Die erfolgreiche Implementierung eines zentralen Datenmanagements bedarf einer nicht nachlassenden Unterstützung durch die Projektleitung.

Die frühe Einbindung eines IS in das Projekt hat den Vorteil, dass sich schon von Beginn an ein Gewöhnungseffekt einstellt, der ebenfalls zur Überwindung der beschriebenen Zurückhaltung beiträgt indem die positiven Erfahrungen beim Gebrauch der angebotenen Dienste eine Eigendynamik entwickeln und die generelle Akzeptanz des zentralen Datenmanagements erhöhen. So können frühzeitig die Vorteile einer zentralen Datenhaltung konkret erfahren werden, beispielsweise die Möglichkeiten der Integration heterogener Datensätze, die aus verschiedenen Disziplinen oder aus früheren Projekten stammen.

Zu Beginn eines SFB können zunächst nur mögliche Daten- und Informationsflüsse antizipiert werden, aufgrund derer die Entwickler Prototypen eines IS erarbeiten. Erst im Projektverlauf entstehen und konkretisieren sich die Anforderungen und Bedürfnisse. Im direkten Austausch und in ständiger Diskussion mit den Nutzern werden diese Prototypen weiterentwickelt und zur Projektreife gebracht. Die Nutzer stellen daher eine wichtige Gruppe beim Aufbau eines IS dar. Das wird oft verkannt, mit dem Ergebnis, dass Systeme aufgesetzt werden, die nicht so funktionieren, wie es für die Anwender erforderlich wäre und die in der Folge nicht (effizient) genutzt werden. Die frühzeitige Integration der Anwender in den Entwicklungsprozess ermöglicht daher die mit Abstand beste Qualitätssicherung für den Betrieb von IS.

Auch die Beratung und der Support für die Nutzer eines IS, spielen neben der eigentlichen Entwicklungsarbeit eine nicht zu unterschätzende Rolle. Hierfür müssen mit Projektbeginn ausreichend personelle Freiräume geschaffen werden, um zu gewährleisten, dass sich die Mitarbeiter diesen Problemfeldern mit einer notwendigen Mindestintensität widmen können. Dieser Punkt ist vor allem auch in Hinblick auf die Einarbeitung neuer Mitarbeiter, z.B. zu Beginn neuer Projektphasen, besonders wichtig. Eine grundlegende Herausforderung ist hierbei auch die Kommunikation zwischen den verschiedenen Fachdisziplinen, da jede Disziplin durch ihre fachspezifische Perspektive eine eigene Sicht auf das Gesamtprojekt hat und diese meist nur in ihrer spezifischen Fachsprache beschreiben kann. Das Zusammenführen dieses Wissens für interdisziplinäre Analysen kann durch ein integrierendes Datenmanagement unterstützt werden. Eine einfache, intuitive, nach Möglichkeit bereits vertraute Benutzerschnittstelle ist hierbei äußerst hilfreich. Zur Antizipation der Daten- und Informationsflüsse sind gute fachliche Kenntnisse des Gesamtprojekts notwendig. Die hierzu benötigte Einarbeitungszeit ebenso wie die eigentliche Entwicklungszeit für die Prototypen verlangen eine frühestmögliche Einbindung des IS-Teams in die Projektplanung.

4 Schlussfolgerungen und Ausblick

Im Rahmen von Umwelt-Großforschungsprojekten sollte jeder Wissenschaftler problemlos auf die von ihm benötigten Datensätze zugreifen können; die informationstechnische Realisierung mittels eines Informationssystems ist seit einigen Jahren möglich. Die Implementierung eines solchen Systems als feste Vorgabe in Umwelt-Großforschungsprojekten könnte nicht nur der initiale Schritt für eine fortschrittlichere und effizientere Zusammenarbeit in den Umweltwissenschaften sein, sondern auch neue methodische Ansätze bei der Interpretation von Daten erlauben, indem Datensätze über inhaltliche, räumliche, und zeitliche Skalenebenen hinweg verknüpft werden. Die Konzeption und der Aufbau eines Datenmanagementsystems zur Unterstützung interdisziplinär arbeitender Arbeitsgruppen ist ein langwieriger, kontinuierlicher Prozess. Hierbei sollte die inhaltliche Weiterentwicklung des Systems während des laufenden Betriebes stetig und im Austausch mit den Anwendern/ Nutzern erfolgen. Dazu ist eine fortwährende Interaktion mit den Nutzern notwendig, was zudem gleichzeitig die Akzeptanz und das Interesse am Datenmanagement stärkt. Das System muss nicht nur inhaltlich aktuell bleiben, sondern auch das Design und die technischen Funktionalitäten sollten mit der rasanten IT-Entwicklung Schritt halten.

Literaturverzeichnis

- [E98] Eisele, F. L.: Strategische Anforderungen an forstliche Informationssysteme. AFZ/ Der Wald 20, 1998; S. 1244-1247.
- [Mü04] Mückschel, C.; Nieschulze, J.; Schachtel, G.A.; Li, S.; Sloboda, B.; Köhler, W.: Web-basierte Informationssysteme in interdisziplinären Umwelt-Forschungsprojekten – am Beispiel der beiden DFG-Sonderforschungsbereiche 299 (Gießen) und 552 (Göttingen/Kassel). - Zeitschrift für Agrarinformatik 3, 2004; S. 46-55.
- [SR94] Stahl, R.; Roggendorf, W.: 10 Jahre Kommunale Umweltinformationssysteme – Bilanz und Konsequenzen. Salzburger Geographische Materialien. 1994; Heft 21.
- [WEM04] Wehrum, A.; Eulberg, P.; Mückschel, C.: Paradigmen des Open-Source-Konzeptes am Beispiel der Software Typo3. - Zeitschrift für Agrarinformatik 4, 2004; S. 104-111.