
Name der AiF-Forschungsvereinigung (FV)

IGF-Antrags-Nr:
(wird von der AiF eingesetzt)

Aktenzeichen der FV

Ausführliche Beschreibung zum Forschungsantrag

1. Forschungsthema

Wissensbasierte Aufbereitung von Prozessdaten

2. Wissenschaftlich-technische und wirtschaftliche Problemstellung

Durch Fortschritte in der Mess- und Informationstechnik ist die Menge verfügbarer Daten aus verfahrenstechnischen Produktionsprozessen, im Folgenden Prozessdaten genannt, in den letzten Jahren stark angewachsen. Neben den aktuellen Messwerten werden immer mehr auch historische Prozessdaten vorgehalten. Mit Hilfe solcher Daten und deren Analyse lassen sich eine Reihe ökonomischer Vorteile erzielen und die Wettbewerbsfähigkeit der Prozesse verbessern (Alsmeyer, 2007). In diesem Geschäftsumfeld sind insbesondere auch kleine und mittelständische Unternehmen (KMU) tätig, sowohl produzierende als auch dienstleistende, die z.B. spezielle Software-Werkzeuge zur Überwachung (Monitoring), Modellierung, Steuerung, Regelung und Optimierung eines Prozesses entwickeln bzw. Projekt- und Beratungsarbeit in diesen Bereichen anbieten.

Mit der Prozessdatenanalyse lassen sich auf zwei Wegen Vorteile erzielen. Erstens können die historischen Prozessdaten *analysiert* und so z.B. das Potential eines Prozesses ausgelotet werden. Insbesondere können Prozesszustände aus der Vergangenheit (im Sinne eines „golden batch“) als Referenz für den aktuellen Betrieb dienen (Dünnebier und v. Felde, 2003). Zweitens können sie zur Erstellung von *Vorhersagemodellen* für verschiedenste Aufgabenstellungen herangezogen werden. Beispiele sind die Prozessoptimierung bezüglich Produktionskosten und Produktqualität (z.B. Häck und Köhne, 1998; Estler und Soravia, 1998; Nohr und Boll, 2003), die Fehlererkennung bzw. Fehlersuche bei Produktionsproblemen (z.B. Kourti, 2002; Venkatasubramanian, 2003) oder die Nutzung bei präventiven Instandhaltungsmaßnahmen.

In der Praxis werden diese Möglichkeiten aber unzureichend genutzt, weil die *Aufbereitung* der Prozessdaten vor der eigentlichen *Analyse* mit herkömmlichen Methoden und Werkzeugen kaum handhabbar ist und enorme Kosten verursacht. Wesentliche Gründe neben der schieren *Datenmenge* sind die *Heterogenität der Datenquellen* sowie die oft schlechte *Qualität* der Messdaten. Als Beispiele sind fehlende oder fehlerhafte Werte, unberücksichtigte Einflüsse bei der Kalibrierung der Messinstrumente, aber auch Totzeiten oder korrelierte Messgrößen zu nennen.

Obwohl selbst grundlegende Fragestellungen der *Datenaufbereitung*, wie z.B. die Ausreißerererkennung, trotz der erzielten Fortschritte immer noch Gegenstand aktueller

Forschungsarbeiten sind (Chiang, 2003), sind in der Literatur für viele der skizzierten Probleme Lösungsmethoden verfügbar. Eine Reihe dieser Methoden sind in einer Vielzahl wenig kompatibler Software-Werkzeuge implementiert. Sie werden oft parallel genutzt und müssen für jeden Einzelfall neu zusammengestellt und geeignet parametrisiert werden. Die Konsistenz der einzelnen Schritte der Datenaufbereitung über Werkzeuggrenzen hinweg muss der Benutzer selbst sicherstellen. Mangels einheitlicher Schnittstellen kommt erschwerend der Zwang zur Konvertierung zwischen Datenformaten hinzu, insbesondere wenn bei iterativen Vorgängen mehrere Werkzeuge involviert sind, so dass die manuelle Datenkonvertierung mehrfach wiederholt werden muss.

Ein zusätzliches Problem besteht darin, dass das *Wissen* über verfügbare Algorithmen und Werkzeuge und geeignete Vorgehensweisen nur implizit bei den Bearbeitern vorhanden ist und nur selten explizit dokumentiert wird. Wertvolle Erfahrungen gehen so mit der Zeit oder durch den Wechsel von Bearbeitern verloren. Manche Analysen lassen sich später nicht mehr *nachvollziehen*, Algorithmen werden nicht oder nicht richtig *wieder verwendet*; viel Zeit und Geld wird dafür verwendet, Wissen wieder zu erarbeiten, welches früher schon vorhanden war.

Es verwundert nicht, dass Ingenieure in der Praxis den Aufwand für die Datenaufbereitung mit bis zu 80% der Gesamtanalyse angeben (z.B. Jaekle und Krell, 2003). Dies führt zu prohibitiven Kosten, insbesondere für die KMU, die nicht über entsprechende Fachabteilungen verfügen und somit gezwungen sind, die begrenzten Personalressourcen besonders effizient einzusetzen. In diesem Forschungsprojekt wird daher eine deutliche Verringerung des Aufwands zur Datenaufbereitung angestrebt.

2.1. Stand der Forschung und Technik

Da es sich bei Prozessdaten im Wesentlichen um Zeitreihen numerischer Daten handelt, gibt es auf den ersten Blick eine Vielzahl von einsetzbaren Software-Werkzeugen zur Datenaufbereitung. Allerdings weisen Prozessdaten bei genauerer Betrachtung eine Reihe besonderer Eigenschaften auf, die eine Nutzung der vorhandenen Werkzeuge zumindest erschweren. Erstens sind viele dieser Software-Werkzeuge nicht auf die Verarbeitung von großen Datenmengen zugeschnitten, wie sie bei industriellen Prozessdatenanalysen meist auftreten. Das gilt z.B. für die Visualisierung von Ergebnissen; manche Werkzeuge beschränken auch die Zahl der Datenpunkte. Zweitens liefern viele Datenquellen zu jedem Messpunkt zusätzlich einen Zeitstempel und ein Qualitätsmaß, wie sie etwa im OPC-Standard (OPC Foundation, 2003) definiert sind. Diese Informationen gehen bei den in der Praxis eingesetzten Aufbereitungsmethoden meist verloren. Drittens werden fehlende Werte, eines der häufigsten Probleme in industriellen Prozessdaten, von vielen Applikationen nicht korrekt behandelt, obwohl der Standard IEC-60559 zur Darstellung von Gleitkommazahlen diese Möglichkeit vorsieht („Not a number“ – NaN). Schließlich treten oft unbekannte Totzeiten oder große Zeitkonstanten im Prozess auf, die zu einem zeitlichen Versatz der Zeitreihen unterschiedlicher Messgrößen führen. Bei absatzweise betriebenen Produktionsprozessen (Batchprozessen) sind unterschiedliche Kampagnendauern für das gleiche Produkt oft nicht vermeidbar (Wold et al., 1998). Beides erschwert die Prozessdatenanalyse erheblich. Ein großer Teil der Prozessdaten besitzt nur einen geringen Informationsgehalt; viele Messgrößen sind stark miteinander korreliert, weil die Daten in den historischen Archiven nicht in geplanten und gut definierten Versuchen generiert, sondern während der Produktion mit unvollständig dokumentierten Randbedingungen aufgezeichnet werden.

Dieser Abschnitt gibt einen Überblick über verschiedene Klassen von Methoden und Werkzeugen, die für dieses Projekt relevant sind.

2.1.1. Datenhaltung und Datenzugriff

Die Datenhaltung ist auch heute noch uneinheitlich. In manchen Fällen werden die Daten zum laufenden Produktionsprozess nur handschriftlich abgelegt. Aber auch bei der Dokumentation der Prozessdaten mit rechnergestützten Methoden bleiben die Datenbasen heterogen, da die meisten Software-Hersteller auf proprietäre Datenformate setzen. Viele den Produktionsprozess betreffende Daten, die den Kontext der Prozessdaten festlegen, wie etwa die Protokolle eines Batchprozesses, werden in Dateien vorgehalten, die unabhängig von den Prozessdaten auf verschiedenen lokalen Rechnern verfügbar sind und zur Prozessdatenanalyse mit großem Aufwand zusammengetragen werden müssen. Eine Verbesserung dieser Situation verspricht der Einsatz spezieller Datenbanken. Dazu zählen so genannte „Process Historians“ oder Prozessinformationsmanagementsysteme (PIMS) wie PI (OSISoft, Inc., 2003) oder InfoPlus.21 (AspenTech, 2002) zur langfristigen Archivierung von Prozessdaten und Laborinformationsmanagementsysteme (LIMS) wie Labvantage (Labvantage Solutions Inc., 2010) oder Nautilus (Thermo Electron Corp., 2010) zur zentralen Verwaltung der Ergebnisse von Laboranalysen. Wegen der hohen Kosten für Implementierung und Betrieb finden solche Systeme aber gerade in den KMU nur eine geringe Verbreitung.

2.1.2. Methoden und Werkzeuge zur Datenaufbereitung

Die nächstliegende Lösung ist es, die Datenaufbereitung mit dem Werkzeug zu erledigen, mit dem auch die Datenanalyse oder die datengetriebene Modellierung erfolgt. Allerdings wird gerade in *spezialisierter Software* zur Prozessdatenverarbeitung die Datenaufbereitung möglicherweise unvollständiger Datensätze nur unzulänglich unterstützt, weil die Hersteller sich auf die Kernfunktionalität ihrer Software konzentrieren. Als Beispiele können Software-Werkzeuge zur Erstellung multivariater linearer Regressionsmodelle (z.B. The Unscrambler (CAMO Inc., 2010) und SIMCA (Umetrics Inc., 2010)) oder neuronaler Netze wie Neurointelligence (Alyuda Research Inc., 2010) genannt werden. Statistikpakete wie Statistica (Statsoft, 2010), PASW Statistics (SPSS Inc., 2010a) oder JMP (SAS Institute Inc., 2010) weisen ähnliche Unzulänglichkeiten auf.

Um diese Mängel zu beheben, werden zusätzlich universelle Software-Werkzeuge verwendet. Häufig werden *Tabellenkalkulationsprogramme* (z.B. Microsoft Excel (Microsoft, 2002)) benutzt, da sie standardmäßig auf Rechnern installiert sind und leicht erweitert werden können. Da viele grundlegende Funktionen der Datenaufbereitung wie etwa die Interpolation fehlender Werte nicht verfügbar sind, müssen diese eigens entwickelt und mit Hilfe einer Makrosprache programmiert werden. Oft ist die Kompetenz in den KMU dafür nicht im nötigen Umfang verfügbar, so dass die Qualität der Lösung nicht sichergestellt werden kann. Mehr einschlägige Funktionalität bieten *Problemlösungsumgebungen* wie Matlab (MathWorks, 2009), Mathematica (Wolfram Research Inc., 2010) oder Maple (Maplesoft, 2010). Die Stärke dieser Lösungen ist die kommandozeilenbasierte Bearbeitung bzw. Erstellung von Programmen (Chonacky, 2005). Die kompetente Anwendung erfordert aber einen hohen Einarbeitungsaufwand, der in den KMU im laufenden Tagesgeschäft von den Projektingenieuren kaum zu leisten ist. Die Nutzung *höherer Programmiersprachen* wie FORTRAN, C oder Java scheidet wegen der noch geringeren spezifischen Unterstützung aus.

Eine weitere Klasse einschlägiger Software-Werkzeuge sind *Data Mining-Werkzeuge*, die entweder als Ergänzung zu Statistikpaketen (z.B. IBM SPSS Modeler (SPSS, 2010b)) oder in eigenständiger Form (z.B. Think Enterprise Data Miner (ThinkAnalytics, 2010) oder IBM Intelligent Miner (IBM, 2010)) kommerziell verfügbar sind. Diese Werkzeuge sind eher auf Geschäftsprozess-Analysen zugeschnitten, wo die verarbeiteten Informationsquellen primär aus Kundendatenbanken oder Webzugriffsstatistiken und weniger aus Zeitreihen aus dem

produktionstechnischen Umfeld bestehen. Interessante Vertreter sind das Open-Source-Werkzeug KNIME (Berthold, 2007) aus der Bioinformatik und das proprietäre Werkzeug BayWiz, das als Erweiterung des Think Enterprise Data Miner von Bayer Technology Services für das Data Mining bei Prozessdaten eingesetzt wird.

Visualisierungswerkzeuge (Trending Software) setzen direkt an der Datenquelle an. Sie sind oft fest in die unter Abschnitt 2.1.1 genannten Datenbanksysteme oder in die Prozessleitsysteme integriert und setzen damit deren Einsatz voraus. Beispielhaft können hier etwa Labview/DIAdem (National Instruments, 1998 u. 2010) im Bereich der Laborautomatisierung und Messdatenerfassung sowie SIMATIC WinCC (Siemens, 2010) im produktionstechnischen Umfeld genannt werden. Das Einbeziehen externer Daten ist meist nur unter zusätzlichem Aufwand realisierbar. Zudem liegt der Fokus dieser Lösungen primär auf der Visualisierung und weniger auf fortgeschrittenen Algorithmen zur Datenaufbereitung. Einen aktuellen Überblick über Forschungsprojekte im Bereich der Datenvisualisierung gibt Lawton (2009).

Unter dem Aspekt des Qualitätsmanagements werden kommerzielle *Softwarelösungen unter dem Oberbegriff Computer Aided Quality (CAQ)* angeboten und besonders in der Produktions- und Fertigungstechnik eingesetzt. Sie umfassen Datenhaltung und -analyse, oft basierend auf den klassischen Ansätzen der statistischen Prozesskontrolle (SPC; Shewart, 1986). Sie setzen genau wie die Software-Werkzeuge in der Verfahrenstechnik intakte Daten voraus. Die in diesem Antrag adressierten Probleme der Datenaufbereitung und der effizienten Nutzung zusätzlicher Werkzeuge der Prozessdatenanalyse sind in dieser Fachdisziplin ebenfalls ungelöst.

Insgesamt gibt es kein einzelnes Werkzeug, das per se den Anforderungen an die Aufbereitung von Prozessdaten vor einer weitergehenden Analyse in vollem Umfang gerecht würde. Immerhin erlauben viele der genannten Applikationen ein Nachrüsten fehlender Funktionalität z.B. durch Makrosprachen oder über Programmierschnittstellen. Dies wird in der Praxis auch genutzt. Allerdings wird wegen der Vielfalt der spezifischen Eigenschaften der Rohdaten oft eine *ad hoc*-Lösung programmiert, ohne jedoch die resultierenden Programme eingehenden Tests – z.B. über die untersuchten Daten hinaus – zu unterziehen. Solche Programme werden dann entweder nicht wieder verwendet, so dass häufige Neuimplementierungen immer ähnlicher Verarbeitungsschritte nötig sind, oder aber die Wiederverwendung von *ad hoc*-Lösungen führt zu Fehlern, weil später untersuchte Daten bestimmte Voraussetzungen nicht erfüllen.

2.1.3. Methoden und Werkzeuge zur Unterstützung von Arbeitsprozessen

Die Aufbereitung von Prozessdaten lässt sich als *Arbeitsprozess* (Fischer, 2005) auffassen, bei dem Eingangsdaten schrittweise, oft unter Nutzung mehrerer Methoden und Werkzeuge, verarbeitet und modifiziert werden. Diese Arbeitsprozesse sind allenfalls durch implizites Erfahrungswissen bei den jeweiligen Bearbeitern manifestiert, jedoch selten als explizit notierte Vorgehensrichtlinien dokumentiert. Das erschwert die *Nachvollziehbarkeit* und damit auch die *Wiederverwendung* erfolgreich angewandter Lösungen.

Erfassung von Arbeitsprozessen

Graphische Notationen und unterstützende Software-Werkzeuge zur Repräsentation von Arbeitsprozessen sind in verschiedenen Bereichen durchaus dem Stand der Technik zuzurechnen. Prominente Beispiele sind die Aktivitätsdiagramme in der UML (Unified Modeling Language) aus dem Bereich Softwaretechnik (Booch et al., 1999; Oestereich, 2004) oder Petrinetze, die z.B. in der Steuerungstechnik (Baumgarten, 1996) und der Modellierung von Geschäftsprozessen (Aalst v.d. und van Hee, 2002) eingesetzt werden. Diese Werkzeuge sind jedoch in aller Regel nicht mit den Werkzeugen zur Datenverarbeitung integriert, so dass lediglich eine nachträgliche Dokumentation bei hohem manuellen Aufwand möglich ist.

Die an der RWTH Aachen entwickelte Methodik K3 (Foltz et al., 2001; Eggersmann et al., 2008) und das dazugehörige *Modellierungswerkzeug* WOMS (Theißen et al., 2008a) hingegen sind stärker auf Arbeitsprozesse des Ingenieurs zugeschnitten und betonen stärker den Modellierungsaspekt. Die Methodik wurde erfolgreich zur Erfassung und Analyse industrieller Arbeitsprozesse in verschiedenen Phasen des Planungsprozesses verfahrenstechnischer Anlagen eingesetzt (Theißen et al., 2008b). Das Werkzeug WOMS plus, eine Weiterentwicklung von WOMS (Theißen et al., 2008c und 2010), kann mittels der Web Ontology Language (OWL), einer Sprache zur Wissensrepräsentation, leicht an spezielle Anforderungen des Nutzers angepasst werden. In dem hier beantragten Projekt können etwa Prozessschritte inklusive der Vorbedingungen mit WOMS plus erfasst werden, beispielsweise die Bestimmung der „Varianz“ für den Schritt „Entrauschung“.

Oft werden *Dokumentenmanagementsysteme* zur indirekten Dokumentation von Arbeitsprozessen verwendet. Sie können allerdings lediglich die Änderungen einzelner Dokumente im Sinne einer Versionshistorie nachvollziehbar machen. Der Weg der Dateninstanzen durch verschiedene Dateien und Werkzeuge kann damit aber nicht nachgezeichnet werden, so dass die zugehörigen Arbeitsprozesse nicht wirklich aufgezeichnet werden können.

Unterstützung von Arbeitsprozessen

Innerhalb der in Abschnitt 2.1.2 genannten Werkzeuge lässt sich die Wiederverwendung von Arbeitsprozessen und Algorithmen oft durch Verwendung von *Makrosprachen* erreichen, entweder durch Aufzeichnung interaktiver Schritte oder durch explizite Programmierung. Einige Werkzeuge bieten eine grafische Programmierung, indem automatisierbare Algorithmen aus elementaren Transformationen zusammengesetzt werden können. Die Makrosprachen sind jedoch untereinander inkompatibel, so dass Implementierungen zunächst an bestimmte Werkzeuge gebunden sind. Liegen Daten im Format eines anderen Werkzeugs vor, so ist entweder eine Portierung der Implementierung oder eine weitere Transformation der Daten erforderlich.

Ohne den Aufwand für eine explizite programmiertechnische Anbindung kommen *Makrorekorder-Programme* aus, die mit den anzubindenden Programmen nur auf Betriebssystemebene (Mausklicks, Tastendrucke) kommunizieren. Eine Wiederverwendung auch nach längeren Zeiträumen erscheint aber aufgrund vieler Abhängigkeiten von Systemkonfiguration und Programmversionen nicht realistisch; es fehlen insbesondere auch Kontrollmöglichkeiten, ob ein früher aufgezeichneter Arbeitsprozess weiterhin verwendbar ist.

Für hochgradig standardisierte Geschäftsprozesse (z.B. Abwicklung eines Schadenfalls in einem Versicherungsunternehmen) existieren kommerzielle *Groupware- und Unternehmenssoftware-Lösungen*, die Arbeitsprozesse mit mehreren Bearbeitern unterstützen. Sie sind aber an die Werkzeuge des jeweiligen Herstellers gebunden. Im Rahmen der Workflow Management Coalition (WfMC, 2010) wurde ein Referenzmodell für den Datenaustausch zwischen Workflow-basierten Werkzeugen geschaffen. Eine Reihe von *Workflow-Managementsystemen*, die teilweise quelltextoffen verfügbar sind (z.B. Enhydra Shark (ObjectWeb, 2010)) unterstützen diese Standards (für einen Überblick zum noch breiter gefassten Thema *Provenance-Management* und zu verschiedenen Workflow-basierten Systemen sei auf Freire et al. (2008) verwiesen).

Ingenieurwissenschaftliche Arbeitsprozesse sind stärker datenzentriert als Geschäftsprozesse, bei denen es vorwiegend um die Zuordnung von Aufgaben unter Berücksichtigung von Zuständigkeiten geht. Aufgrund dieser Unterschiede gibt es spezielle *Forschungsinitiativen mit eigenen, prototypischen Werkzeugen*, etwa das Kepler-Projekt (Altintas et al., 2004), myGrid

(Stevens et al., 2003), Taverna (De Roure, 2009) und das Triana-Projekt (Majithia et al., 2004). Deren Ziel ist die automatisierte Abarbeitung vorgegebener Arbeitsprozesse; dazu müssen die externen Werkzeuge über definierte Schnittstellen explizit an das Werkzeug zur Arbeitsprozessunterstützung angebunden werden.

2.1.4. Integrationslösungen in der Verfahrenstechnik

Verschiedene Forschungsprojekte haben sich in der Vergangenheit um Lösungen für die nachträgliche Integration existierender Werkzeuge bemüht. Als Integrationslösungen zur Prozessdatenverarbeitung sind hier die von der EU geförderten Projekte CHEM (Cauvin et al., 2002) und MAGIC (Köppen-Seliger et al., 2003) zu nennen, die auf die Analyse von Online-Messdaten und eine darauf aufbauende Unterstützung des Bedienpersonals abzielen.

Noch breiter gefasst waren die Arbeiten im von der DFG geförderten Sonderforschungsbereich SFB 476 und dem nachlaufenden Transferbereich 61, die inzwischen abgeschlossen sind und sich auf die informatische Unterstützung verfahrenstechnischer Entwicklungsprozesse konzentrierten (Nagl und Marquardt, 2008). Der Fokus lag bei den vorbetrieblichen Phasen des Anlagenlebenszyklus, insbesondere der konzeptuellen Prozessentwicklung. Die Betrachtung betrieblicher Aspekte wie beispielsweise die Erfassung und Analyse von Prozessdaten während der Produktion erfolgte dort nicht.

In CHEM und MAGIC wird auf die *vollständige Integration* der Werkzeuge gesetzt, um sie dann innerhalb eines homogenisierten Frameworks nutzen zu können. Dieser Ansatz soll in dem hier beantragten Projekt nicht verfolgt werden. Vielmehr soll eine Plattform entwickelt werden, die bestimmte Grundfunktionen wie die Dateninterpolation bereitstellt. Mit dieser Plattform sollen externe Werkzeuge *partiell integriert* werden können, um deren interaktive Nutzung mit Arbeitsprozessunterstützung zu ermöglichen und nachhaltig zu dokumentieren. Gerade solche Ansätze zur partiellen Integration sind intensiv und erfolgreich im SFB 476 erforscht worden. Die dort erzielten methodischen Resultate insbesondere im Bereich der Konzeption und Architektur solcher Systeme und die entstandenen Datenmodelle können als Grundlage für die Integrationsplattform im hier beantragten Projekt genutzt werden. Dazu sind Anpassungen erforderlich, die dem Charakter betrieblicher Prozessdaten Rechnung tragen, insbesondere den im Vergleich zur Entwurfsphase erheblichen Datenmengen, die über einen langen Zeitraum verteilt anfallen. Außerdem bietet dieses Projekt die Möglichkeit, die konzeptuellen Ergebnisse des SFB 476 im industriellen Umfeld mit Fokus auf die Aufbereitung von Prozessdaten zu validieren.

2.1.5. Datenmodelle und Standards

Datenmodelle und Standards sind die Grundlage für Integrationslösungen. Im Bereich der Verarbeitung von Prozessdaten hat sich *OLE for Process Control* (OPC) als Industriestandard etabliert. Hier ist im Besonderen die Spezifikation zum Zugriff auf historische Daten (OPC-HDA, Historical Data Access, OPC Foundation (2003)) für das Projekt relevant. Eine verwandte Initiative mit dem Fokus auf die Messdatenerfassung im Automobilbereich ist ASAM (ASAM, 2010). Um den Datenzugriff kümmern sich auch Data-Mining-Standards wie das Common Warehouse Metamodel CWM (Object Management Group, 2010).

Ebenfalls aus dem Data Mining kommt die Predictive Model Markup Language (PMML) für statistische und Data Mining-Modelle, die auch die Repräsentation von Aufbereitungsschritten wie z.B. stückweise lineare Interpolation oder Transformationen zur Normierung unterstützt (Data Mining Group, 2010).

Für die Arbeitsprozesse sind Spezifikationen der Workflow Management Coalition (WfMC, 2010) relevant, u.a. die XML Process Definition Language XPDL.

Im Bereich von Verfahrenstechnik und Anlagenbau wurde eine Reihe von Datenmodellen vorgeschlagen (für einen Überblick siehe Bayer, 2003). Davon unterstützen aber nur wenige auch Messdaten aus Produktionsprozessen. Das sind beispielsweise das Application Protocol 231 des ISO-Standards STEP (ISO 10303, 1998) sowie das Datenmodell MDOOM (Batres et al., 1999).

2.2. Praktisches Fallbeispiel

Abschließend soll hier an einem konkreten Anwendungsbeispiel aus der industriellen Praxis illustriert werden, warum der Aufwand für die Datenaufbereitung bei alltäglichen Analysenaufgaben sehr hoch sein kann.

Für einen Batchprozess sollte der Endpunkt jeder Charge mit Hilfe der online verfügbaren Messgrößen bestimmt werden, um die aufwändige Probennahme und die totzeitbehaftete Laboranalytik zu vermeiden. Dazu sollte ein multivariates Regressionsmodell (PLS, Martens und Naes, 1989) zum Einsatz kommen. Zur Modellerstellung standen Daten für ca. 20 Messgrößen aus etwa 20 Chargen zur Verfügung, was in der Summe ca. 800.000 Datenpunkte ergab. Die Zeitreihen waren zuvor bereits von einem Kollegen analysiert worden. Um unterschiedliche Zeiten zur Prozessierung einer Charge auszugleichen, hatte der Kollege sie mittels eines Excel-Makros über einer Indikatorvariablen statt über der Zeit aufgetragen, was eine Interpolation der Messdaten erforderte. Für jede Variable existierte eine separate Excel-Datei mit einer Spalte pro Charge.

Für den Import der Daten in ein Software-Werkzeug zur multivariaten Regression (SMM) mussten die Daten in einer einzigen Tabelle vorliegen. Im Gegensatz zu den vorliegenden Tabellen musste dazu jeder Spalte eine Messgröße zugeordnet sein, die Chargen dagegen sollten einfach untereinander stehen. Es mussten also 400 Zeitreihen aus 20 Tabellen in eine einzelne Tabelle umsortiert werden. Beim Import stellte sich dann heraus, dass die Zahl der importierbaren Zeilen in SMM beschränkt ist. Die Chargen mussten daher nochmals auf zwei Tabellen aufgeteilt werden. Außerdem waren einzelne Zeilen in den Tabellen leer, weil im Interpolationsmakro einige Sonderfälle in den Daten nicht korrekt behandelt wurden. Nach Korrektur dieses Fehlers mussten sämtliche Zeitreihen erneut interpoliert, umsortiert und in SMM importiert werden.

Im Verlauf der multivariaten Analyse erwies es sich dann als sinnvoll, die Regression auf die letzte Produktionsphase des Batchprozesses zu beschränken, deren Beginn durch einen steilen Temperaturanstieg gekennzeichnet war. Zudem sollte der Füllstand auf den Beginn dieser Phase normiert werden. SMM unterstützte diese nachträglichen Manipulationen nicht. In Ermangelung automatischer Verfahren wurden Startzeitpunkt und Anfangsfüllstand manuell bestimmt. Danach musste nicht nur die gesamte Datenaufbereitung ab der Interpolation, sondern auch die gesamte Analyse mit SMM wiederholt werden.

Der Zeitaufwand für die Datenaufbereitung lag in diesem Beispiel vermutlich sogar über 80%, obwohl es sich nur um grundlegende Aufgaben handelte, nämlich um Konsistenzprüfung, Interpolation, Normierung, Auswahl einer Teilmenge und Umorganisation von Daten. Der Aufwand ergab sich durch mangelnde Nachvollziehbarkeit der vorangegangenen Arbeiten, fehlendes Wissen zu evtl. schon bei Kollegen verfügbaren Werkzeugen für die skizzierten Aufgaben und Unklarheiten über die Vorbedingungen für einen Analysenschritt (Import). Der Grund dafür ist eine mangelnde Unterstützung des Arbeitsprozesses in den benutzten Werkzeugen und über Werkzeuggrenzen hinweg (iteratives Vorgehen).

2.3. Zusammenfassung

Die beschriebenen Schwierigkeiten und die damit verbundenen Kosten sind der Hauptgrund dafür, dass ökonomische Vorteile aus Prozessdatenanalysen nicht in dem Maße realisiert werden, wie man es aufgrund der Datenverfügbarkeit erwarten müsste. Für die KMU entsteht hier ein besonderer wirtschaftlicher Nachteil, denn sie können sich keine Fachabteilungen mit einer starken Personaldecke leisten, in denen angepasste Inhouse-Werkzeuge qualitätsgesichert entwickelt und betreut werden können. Um gerade die Bedürfnisse der KMU zu adressieren, soll dieses Projekt einen neuen methodischen Ansatz und eine entsprechende informationstechnische Unterstützung untersuchen, um eine effektive Aufbereitung von Prozessdaten vor dem Hintergrund der folgenden Aufgabenstellung zu ermöglichen:

1. Die *Konsistenz* von Aufbereitungsschritten über Werkzeuggrenzen hinweg wird sichergestellt, um Fehler beim Datentransfer zu vermeiden.
2. Die bei der Aufbereitung der Prozessdaten durchgeführten Arbeitsschritte werden *nachvollziehbar* gemacht, um die Verwendung der Ergebnisse im Nachhinein besser beurteilen zu können.
3. Die *Wiederverwendbarkeit* von Algorithmen und Vorgehensweisen zur Datenaufbereitung wird unterstützt.
4. Mögliche *Fehler* in den Prozessdaten werden erkannt. Fehler in der Vorgehensweise werden, wenn möglich, *vermieden*, aber zumindest frühzeitig erkannt.

3. Forschungsziel / Ergebnisse / Lösungsweg

3.1. Forschungsziel

Dieses Forschungsprojekt strebt eine drastische Verringerung des Aufwandes für Prozessdatenanalysen an, indem eine wissenschaftliche Methodik zur effizienten Aufbereitung der Daten entwickelt wird. Hiervon profitieren in besonderem Maße die KMU, da vermiedene Fehler, eine erhöhte Wiederverwendbarkeit und Nachvollziehbarkeit der durchgeführten Arbeitsschritte eine wesentlich *höhere Planungssicherheit* (insbesondere für dienstleistende KMU) bieten und somit einen *deutlich effizienteren Personaleinsatz* ermöglichen. Um die Methodik praktisch nutzbar zu machen, wird eine *Wissensbasis* in Form eines *Methodenkatalogs* bereitgestellt, der bekannte Arbeitsschritte und etablierte Vorgehensweisen zur Datenaufbereitung strukturiert zusammenfasst. Desweiteren wird untersucht, wie diese Wissensbasis mit informationstechnischen Mitteln noch besser genutzt werden kann. Dazu soll eine Plattform (Wissensbasis, Architektur und Software) entwickelt werden, die mit Hilfe der Wissensbasis die Datenaufbereitung über Werkzeuggrenzen hinweg ermöglicht. Die Plattform erhebt dabei nicht den Anspruch, bereits existierende Werkzeuge zur Aufbereitung vollständig zu integrieren. Sie soll vielmehr eine gewisse Basisfunktionalität und insbesondere die Möglichkeit zur *manuellen Nutzung externer Werkzeuge mit Workflow-Support* bieten. Auf diese Weise werden existierende und bewährte Lösungen nicht ersetzt, sondern können systematisch und konsistent genutzt werden. Zur Entwicklung der Plattform werden Referenzdatensätze und -arbeitsprozesse aus der Industrie betrachtet.

3.1.1. Angestrebte Forschungsergebnisse

Das Projekt strebt drei wesentliche Ergebnisse an:

- einen Methodenkatalog, der bekanntes Wissen im Rahmen der Datenaufbereitung verfügbar macht,
- eine Architektur, die dieses Wissen zur Unterstützung der Datenaufbereitung einbezieht und insbesondere die konsistente Nutzung externer Werkzeuge unterstützt,
- sowie einen Software-Demonstrator, der die Architektur beispielhaft implementiert.

Die Komponenten der *Plattform zur Datenaufbereitung* werden in Bild 1 dargestellt. Die Plattform stellt die Wissensbasis (Methodenkatalog + Historie von Arbeitsprozessen) werkzeugübergreifend zur Verfügung und bietet dem Bearbeiter folgende Optionen:

- vorhandene Grundfunktionen nutzen, wie z.B. Dateninterpolation;
- externe Werkzeuge manuell, aber mit Arbeitsprozessunterstützung (Workflow-Support) verwenden, d.h. interaktive und nicht automatisierbare Schritte ausführen;
- externe Software-Werkzeuge vollständig integrieren.

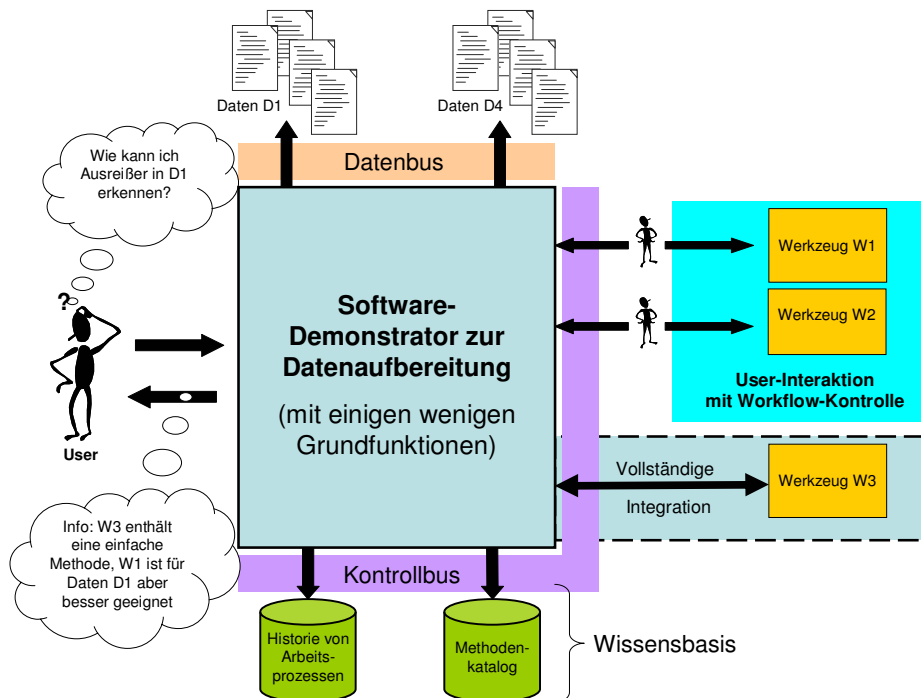


Bild 1: Grundstruktur der Plattform zur Datenaufbereitung

Der *Methodenkatalog zur Datenaufbereitung* soll im Sinne eines Handbuchs bekannte Arbeitsschritte (etwa „Entrauschung“) strukturiert auflisten, angereichert durch wichtige Eigenschaften wie zugehörige Ein-/Ausgangsparameter (z.B. „Varianz“) und Werkzeuge, die diesen Arbeitsschritt implementieren. Der Katalog wird um etablierte Vorgehensweisen ergänzt, um erfolgreiche Arbeitsprozesse im Sinne von „best practice“ festzuhalten und für eine Wiederverwendung verfügbar zu machen. Die Datenaufbereitung kann als eine Sequenz von Transformationen auf Prozessdaten betrachtet werden, die sich als verkettete Schritte eines Arbeitsprozesses abbilden und speichern lassen. Beispiele sind Umorganisation, Sortieren, Zusammenfügen, Interpolation oder Entrauschung der Zeitreihen. Zur Erstellung des Methodenkatalogs wird Wissen über die Datenaufbereitung aus der einschlägigen Literatur und durch Analyse durchgeführter industrieller Projekte erhoben. Mittels innovativer Technologien

zur Wissensrepräsentation (z.B. objektorientierte Modellierung, Produktionsregeln, Ontologien; siehe Brachman und Levesque, 2004) wird dieses Wissen in einer Wissensbasis formalisiert und einer Rechnerverarbeitung zugänglich gemacht.

Die Entwicklung der *Architektur* wird dabei von zwei Kernideen geleitet: Erstens soll der Zugriff auf Mess- und Analysedaten durch einen *Datenbus* transparent gestaltet werden, so dass ein Zugriff auf die Datensätze unabhängig von deren Speicherort (z.B. eine oder mehrere Dateien, Datenbanken, etc.) und deren Speicherformat möglich ist. Zweitens soll ein *Kontrollbus* auf Datensätzen durchgeführte Arbeitsschritte protokollieren und an Hand der Wissensbasis mögliche, weitere Schritte vorschlagen. Die geplante Software-Unterstützung erlaubt es dann, für jeden Datensatz zu jedem Zeitpunkt einer Datenaufbereitung übergeordnete Informationen (Metadaten) bereitzuhalten, die etwa die zuvor erfolgten Transformationen oder die getroffenen Annahmen betreffen. Damit können durchgeführte Aufbereitungsschritte nachvollzogen, einmal erstellte Lösungen effizienter wiederverwertet und potenzielle Fehler frühzeitig erkannt werden. Der Ansatz unterstützt so die Entwicklung von Referenzarbeitsprozessen, die im Sinne einer Know-How-Sicherung insbesondere im KMU-Umfeld wünschenswert sind.

Die Umsetzung der Architektur in einem *Software-Demonstrator zur Datenaufbereitung*, der die Architektur für einen Funktionsnachweis im industriellen Umfeld beispielhaft implementiert, erfolgt schrittweise mit zunehmender Komplexität. Auf der einfachsten Ebene werden die verwendeten Daten und die einzelnen Aufbereitungsschritte lediglich mitprotokolliert. Dies erlaubt es dem Bearbeiter, den Weg der Daten einer bereits durchgeführten Datenaufbereitung *nachzuvollziehen*, etwa um auf die verwendeten Ursprungsdaten rückzuschließen oder um im Verlauf der Aufbereitung getroffene Annahmen zu überprüfen. Auch ergibt sich dadurch ein direkter Nutzen für die Praxis, weil durchgeführte Arbeiten etwa dem Kunden oder Kollegen gegenüber einfach *dokumentiert* werden können. Ein weiterer wichtiger Aspekt ist die *Versionsverwaltung* der aufbereiteten Daten. Der Bearbeiter soll auf dieser Ebene mit möglichst geringem manuellen Aufwand belastet werden.

Auf der nächsten Komplexitätsebene steht die *Wiederverwendung* im Vordergrund. Der Bearbeiter wird darin unterstützt, eine bereits gemachte Aufbereitung mit neuen Rohdaten oder unter neuen Annahmen, etwa veränderten Parameterwerten oder mit alternativen Algorithmen, zu wiederholen. Sinnvoll auf dieser Ebene ist außerdem die Spezifikation von Vor- oder Nachbedingungen eines Transformationsschrittes. Solche Bedingungen ergeben sich oft erst im Nachhinein, etwa wenn bei der Verarbeitung neuer Rohdaten Fehler auftreten. Durch Nachpflegen dieser Bedingungen können potenzielle Fehler bei der Durchführung weiterer Datenaufbereitungen vermieden werden.

Auf der obersten Komplexitätsebene schließlich wird angestrebt, den Aufbereitungsprozess auch bei neuartigen Problemstellungen zu unterstützen. Mit Hilfe der Metainformationen über die Daten können die Vorbedingungen für einen Arbeitsschritt (Transformation) vorher geprüft und *Fehler vermieden* werden. Zum anderen können dem Bearbeiter auf Grund der Datenlage sinnvolle nächste Arbeitsschritte oder ein Referenzarbeitsprozess aktiv vorgeschlagen werden, indem auf die vorangegangenen Datenaufbereitungen als Wissensbasis zurückgegriffen wird.

Zur Sicherung der Praxistauglichkeit wird der Software-Demonstrator zur Datenaufbereitung in seinen verschiedenen Ausbaustufen schon während der Projektlaufzeit von den industriellen Projektpartnern getestet.

3.1.2. Innovativer Beitrag der angestrebten Forschungsergebnisse

Mit dem in diesem Projekt entwickelten Ansatz wird es erstmals möglich sein, die ingenieurmäßige Aufbereitung von Prozessdaten *unabhängig von den eingesetzten Methoden und*

spezifischen Software-Werkzeugen nachvollziehbar zu machen, dadurch Wiederverwendung zu fördern und Fehler zu vermeiden.

Die Erfassung und Analyse von Referenzarbeitsprozessen ist im Bereich der Datenaufbereitung nach Wissen des Antragstellers bisher noch nicht systematisch untersucht worden. Hier besteht eine Verwandtschaft zu den Methoden der Arbeitsprozessunterstützung, die bei Geschäftsprozessen Stand der Technik sind, im Ingenieursbereich insbesondere wegen ihrer nicht-deterministischen Eigenschaften und der enthaltenen kreativen Arbeitsschritte trotz der Fortschritte der letzten Jahre (u.a. im SFB 476 und im TB 61) noch Gegenstand der Forschung sind (Nagl und Marquardt, 2008).

Eine strukturierte Zusammenstellung bekannter Datenaufbereitungsschritte in einem Methodenkatalog, einschließlich der zugehörigen Eigenschaften und Werkzeuge in einer formalisierten Form, ist dem Antragsteller nicht bekannt. Eine derartige Wissensbasis würde es erlauben, durchgeführte Arbeitsprozesse im Sinne von „best practice“-Vorgehensweisen zu katalogisieren und zu einer (durchaus auch firmeninternen) Richtlinie zusammenzufassen.

Der *flexible Ansatz zur Integration existierender Werkzeuge* im Bereich der Datenaufbereitung ist neu und innovativ, da man sich nicht auf einzelne Werkzeuge festlegt, sondern der Tatsache Rechnung trägt, dass vom Ingenieur im unternehmerischen Alltag vielfältige Werkzeuge eingesetzt werden. Diese lassen sich nicht alle mit klassischen Integrationsansätzen anbinden (hoher Aufwand, proprietäre und schlecht dokumentierte Schnittstellen, etc.). Daher werden in diesem Projekt zwei Möglichkeiten geboten, auch externe Werkzeuge zu verwenden. Eine vollständige Integration wird durch die angestrebte Architektur zwar unterstützt, bringt aber die genannten Probleme mit sich, so dass sie nur für eine kleine Menge von Basiswerkzeugen realisierbar ist. Die Alternative besteht in einer *teilmanuellen (interaktiven) Anbindung externer Werkzeuge*, wobei der Weg der Daten durch diese Werkzeuge, die der User interaktiv aus dem Software-Demonstrator heraus benutzt, nachverfolgt wird. Es werden also in hohem Maße auch interaktive Schritte unterstützt, die gar nicht automatisch erfasst und wiederholt werden können.

Im Gegensatz dazu versuchen derzeitige kommerzielle Lösungen, den kompletten Problemlösungsprozess innerhalb eines *einzelnen Werkzeuges* zu realisieren. Die Nachteile wurden in Abschnitt 2.1 beschrieben. Herkömmliche *Integrationslösungen* wiederum zielen auf eine automatische Abarbeitung ab, die oft nur Teile des gesamten Arbeitsprozesses erfasst und damit nicht in der erforderlichen Stringenz möglich ist.

3.2. Lösungsweg zur Erreichung des Forschungsziels

Erarbeitung von Referenzarbeitsprozessen

Als erster Schritt zur Erreichung des Forschungsziels (vgl. Arbeitsdiagramm in Bild 2) werden in Zusammenarbeit mit industriellen Anwendern für einige relevante Anwendungsfälle Referenzarbeitsprozesse erarbeitet und Referenzdatensätze im Sinne von Benchmark-Problemen zusammengetragen. Besonders interessant sind in diesem Zusammenhang abgeschlossene Projekte zur Prozessdatenanalyse, bei denen der Aufwand für die Datenaufbereitung im Rückblick besonders hoch war. Das unter Abschnitt 2.2 beschriebene Fallbeispiel zur Endpunktsbestimmung in einem Batchprozess entstammt einem solchen industriellen Projekt und bietet sich als Einstieg an. Weitere Beispiele werden im Projektverlauf von m2p-labs, Umesoft, Bayer Technology Services und Evonik Degussa bereitgestellt und analysiert.

Zur Erfassung der Arbeitsprozesse kommt das Werkzeug WOMS plus zur Arbeitsprozessmodellierung (Theißen et al., 2008c und 2010) zum Einsatz, das sich bereits bei der Modellierung komplexer, kreativer Arbeitsprozesse im Rahmen DFG-geförderter Vorhaben

(Nagl und Marquardt, 2008, SFB 476, DFG-Transferprojekt) in enger Kooperation mit industriellen Partnern bewährt hat. Das Ziel des Forschungsprojektes ist es, die Referenzarbeitsprozesse mit Hilfe des Software-Demonstrators zur Datenaufbereitung durchgängig unterstützen zu können.

Anforderungsanalyse und Methodenkatalog

Die anschließende Analyse der erfassten Arbeitsprozesse erlaubt es, die benötigte Funktionalität für eine Werkzeugunterstützung näher zu spezifizieren. Von besonderer Wichtigkeit sind hierbei

- die Art der Datenquellen und -senken (Dateien unterschiedlicher Formate, Datenbanken) und des Zugriffs hierauf (lokal, Netzwerk- bzw. Web-basiert),
- die eingesetzten Software-Werkzeuge zur Bestimmung sinnvoller Grundfunktionen des Software-Demonstrators zur Datenaufbereitung,
- die Form der Datensätze und die in der Praxis benötigten Metainformationen hierüber,
- die Herausarbeitung von Gemeinsamkeiten in den Arbeitsprozessen (z.B. wiederkehrende Abfolgen von Einzelschritten im Sinne von „best practice“)
- typische Muster in der Bearbeitung und Gründe hierfür (z.B. sequentielle oder parallele Schritte, Rücksprünge).

Die Analyse der Referenzarbeitsprozesse bildet die Grundlage eines Datenmodells für die betrachteten Datensätze einerseits und für die bei der Datenaufbereitung durchzuführenden und zu unterstützenden Arbeitsabläufe andererseits. Bei der Entwicklung dieses Datenmodells werden existierende Datenmodelle und Spezifikationen im Bereich der Prozessdatenverarbeitung weitestgehend berücksichtigt (z.B. OPC-HDA (OPC Foundation, 2003) oder dessen Nachfolger OPC-UA Part 8 (OPC Foundation, 2009), das Application Protocol 231 des ISO-Standards STEP (ISO 10303, 1998) oder MDOOM (Batres et al., 1999)).

Architektur für eine werkzeugübergreifende Plattform

Neben der Erarbeitung und Bereitstellung des Methodenkatalogs ist die Entwicklung einer Architektur für eine werkzeugübergreifende Plattform zur Datenaufbereitung mit wissensbasierter Unterstützung ein Kernbestandteil der Forschungsarbeiten. Ausgangspunkt ist die in Bild 1 dargestellte Struktur. Die Architektur basiert auf einer Komponente für die Datenintegration (Datenbus), die den konsistenten Datentransfer zwischen den Werkzeugen sicherstellt, sowie einer Komponente für die Prozessintegration (Kontrollbus), die durchgeführte Arbeitsschritte protokolliert und somit Vorgehensweisen nachvollziehbar und wiederverwendbar macht.

Die Komponente für die *Datenintegration* stellt sicher, dass die jeweils in einem Werkzeug benötigten Daten im richtigen Format vorliegen. Diese Komponente muss daher Informationen über Struktur und Speicherort der Prozessdaten besitzen und muss darüber hinaus in der Lage sein, Konvertierungen zumindest in gängige Austauschformate vorzunehmen. Im Sinne einer offenen Architektur wird nicht die zentrale Speicherung der Daten in einer Datenbank angestrebt, sondern eine mit einem *Data Warehouse*-Ansatz beschreibbare Lösung, bei der lediglich gespeichert wird, welche Datensätze an welcher Stelle in welchen Dateien oder Datenbanken stehen (Metainformationen). Diese Informationen sind ausreichend zu formalisieren, so dass benötigte Konvertierungen bei Bedarf automatisch durchgeführt werden können.

Die Komponente für die *Prozessintegration* hingegen verwaltet ein Modell des durchgeführten Arbeitsablaufes, also einzelne Arbeitsschritte (oder Transformationen wie etwa „Entrauschen“) sowie dazu benötigte Parameter (z.B. „Varianz“) und vermerkt das dazu verwendete Werkzeug.

Basierend auf diesen Informationen werden die Nachvollziehbarkeit und Wiederverwendung von Vorgehensweisen erreicht. Darüber hinaus ergibt sich durch die zentrale Speicherung der Metainformationen über Daten und zuvor ausgeführte Aufbereitungsschritte eine Wissensbasis, auf deren Grundlage neue Aufbereitungen einfacher und mit einer verbesserten Qualität durchführbar werden.

Die Architektur soll modular und iterativ aufgebaut werden, so dass die Entwicklung des Software-Demonstrators zur Datenaufbereitung schrittweise und dessen Anwendung auf zunehmend komplexe Anwendungsfälle möglich sind. Die Entwicklung soll in den folgenden Schritten verlaufen, wobei die Bestandteile der Architektur (Datenbus, Kontrollbus) und der Software jeweils um die benötigten Fähigkeiten erweitert werden.

1. *Konsistenz von Aufbereitung und Daten.* Zunächst wird davon ausgegangen, dass eine formalisierte Repräsentation des Arbeitsprozesses (ein sog. Arbeitsprozessmodell) vorliegt, gemäß dessen ein Bearbeiter ein konkretes Datenaufbereitungsproblem angehen möchte. Die Kontrollkomponente soll bei diesem Arbeitspaket lediglich sicherstellen, dass die Aufbereitung gemäß dem Arbeitsprozessmodell erfolgt, indem die richtigen Werkzeuge in der korrekten Reihenfolge benutzt werden. Die Datenbus-Komponente stellt die Konsistenz der benutzten Daten sicher. Die Metainformationen über den Arbeitsprozess und über die tatsächlich erzeugten Daten werden in der Wissensbasis abgelegt. Ziel des Arbeitspaketes ist es, beliebige Werkzeuge ansprechen zu können.
2. *Nachvollziehbarkeit und Dokumentation.* Im zweiten Szenario wird die Mitprotokollierung eines Arbeitsprozesses zur Datenaufbereitung unterstützt. Zunächst genügt es, den Fluss der Daten durch die Werkzeuge nachzuhalten, wobei der Zusatzaufwand für den Bearbeiter minimal sein soll. Der Nutzen der Rückverfolgbarkeit auf dieser einfachen Ebene ist bereits erheblich, da in der Praxis z.B. häufig die Frage nach dem letzten gültigen Datenaufbereitungsschritt auftaucht. Um die neue Vorgehensweise in der Wissensbasis aber später auch für ähnlich gelagerte Problemstellungen nutzen zu können, muss sie abstrahiert werden. Eine denkbare Lösung ist hier die Einordnung der auf die Originaldaten angewandten Transformationen in ein (möglichst hierarchisch aufgebautes) Kategoriensystem für Arbeitsschritte, das z.B. Kategorien wie Entrauschen, Sortieren oder Umorganisieren enthält und diese durch eine formale Charakterisierung weiter spezialisiert (z.B. ein „Entrauschen“ mittels „Wavelets“ in einer „Orthogonalbasis“). Die Art der Abstraktion (Kategorisierung und weitere Spezialisierung) ist von den zuvor entwickelten Datenmodellen abhängig. Die Wahl einer Kategorie für einen Arbeitsschritt erfolgt durch Benutzerinteraktion und soll für beliebige Werkzeuge nutzbar gemacht werden. Die Formalisierung und Kategorisierung von Arbeitsschritten in den Ingenieurwissenschaften ist Gegenstand der aktuellen Forschung (Theißen et al., 2010); gute Lösungen sind auch für dieses Projekt Erfolgsvoraussetzung.
3. *Wiederverwendung und Versionierung.* Um die Wiederverwendung besser zu unterstützen, wird im dritten Arbeitspaket die Verwaltung von Versionen der Arbeitsprozesse für die Datenaufbereitung ermöglicht. Zudem werden Vor- oder Nachbedingungen für die Arbeitsschritte (Transformationen) eingeführt. Damit kann geprüft werden, ob eine bestimmte Transformation auf die vorliegenden Daten überhaupt anwendbar ist.
4. *Wiederverwendung und Fehlervermeidung bei neuen Problemstellungen.* In diesem Arbeitspaket soll die Kontrollkomponente mit Hilfe der Wissensbasis die Lösung neuer Problemstellungen unterstützen. Denkbar ist das Auffinden ähnlicher (Referenz-) Aufbereitungen, aber auch die Unterbreitung konkreter Vorschläge für den jeweils

nächsten Arbeitsschritt (Transformation). Dieses Szenario hat einen explorativen Charakter, ist aber Erfolg versprechend, weil die Datenaufbereitung zwar nicht immer einem identischen Schema folgt, sich aber dennoch Regeln etwa für die Reihenfolge von Aufbereitungsschritten (z.B. „Eliminierung von Ausreißern muss vor der Filterung erfolgen“) oder die Vorbedingungen für einen Aufbereitungsschritt (z.B. „Logarithmische Transformation erfordert strikt positive Werte“) formulieren lassen. Die Datenaufbereitung nimmt damit eine Mittelstellung ein zwischen standardisierten Arbeitsabläufen, wie sie z.B. im Bereich der Geschäftsprozesse zur Auftragsbearbeitung auftreten, deren werkzeugseitige Unterstützung Stand der Technik ist, und kreativen Tätigkeiten, wie etwa im Bereich des Entwurfs oder der Konstruktion, wo eine solche Arbeitsprozessunterstützung noch in den Kinderschuhen steckt.

Bei der Detaillierung der Architektur wird ein offener Ansatz basierend auf Standardschnittstellen, -formaten und -protokollen (OPC, XML, SQL) verfolgt. Außerdem sollen, soweit möglich, frei verfügbare Software-Komponenten zum Einsatz kommen. Angedacht ist z.B. der Einsatz des frei verfügbaren Versionsverwaltungssystems *Subversion*. Die Anforderungen an diese Komponenten ergeben sich im Verlauf des Projektes aus den Datenmodellen und werden im Rahmen der Architekturentwicklung überprüft.

Implementierung eines Software-Demonstrators und Test durch KMU-Nutzer

Die Module der entwickelten Architektur werden im Rahmen des Projektes als Teile eines Software-Demonstrators implementiert. Anhand von Tests mit den von den Partnern im Projektbegleitenden Ausschuss bereit gestellten Referenzarbeitsprozessen wird die Machbarkeit und die Nützlichkeit des wissensbasierten Ansatzes untersucht. Zur Implementierung werden Microsoft .NET und die objektorientierten Programmiersprachen C# oder C++ genutzt. Diese Technologien sind Bestandteil aktueller wie auch künftiger Windows-Betriebssysteme; damit ist die Nutzbarkeit auf Standard-Arbeitsplatzrechnern im industriellen Umfeld garantiert. Die Dokumentation erfolgt mit UML und den in gängigen Entwicklungswerkzeugen verfügbaren Funktionalitäten, um eine nachhaltige Pflege und spätere Weiterentwicklung des Demonstrators zu ermöglichen. Die fertig gestellten und in der Forschungsstelle getesteten Software-Module werden schon während des Projektverlaufs durch die KMU-Anwender im Projektbegleitenden Ausschuss erprobt. Ihre Erfahrungen können so umgehend in die Weiterentwicklung einfließen.

Insgesamt zielt das Forschungsvorhaben nicht auf eine umfassende Lösung mit möglichst vielen Funktionalitäten ab. Stattdessen werden Wissensbasis und Funktionalitäten im Wesentlichen auf die betrachteten Referenzfälle beschränkt, wobei eine spätere Erweiterbarkeit für andere Arbeitsprozesse zur Datenaufbereitung berücksichtigt wird. Der verfolgte Ansatz soll so ausgestaltet werden, dass er auch für weitergehende Aufgaben im Umfeld der Prozessdatenverarbeitung (z.B. Mustererkennung, Modellidentifikation) geeignet ist, auch wenn im Rahmen dieses Vorhabens nur der Aspekt der „Aufbereitung“ betrachtet wird.

3.3. Arbeitsdiagramm und Konzept zum Projektmanagement

Der geplante Personaleinsatz beträgt 2 Personenjahre für einen wissenschaftlichen Mitarbeiter, zuzüglich einer studentischen Hilfskraft mit 86h/Monat, die im Wesentlichen Unterstützung bei den Programmieraufgaben leistet. Der Arbeitsplan ist in Bild 2 dargestellt und folgt der Darstellung und Gliederung aus Abschnitt 3.2.

Arbeitspaket	Jahr 1												Jahr 2											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
1. Referenzarbeitsprozesse und -datensätze																								
2. Datenmodell und Methodenkatalog																								
3. Methodikentwicklung und Architektur																								
3.1 Konsistenzsicherung																								
3.2 Nachvollziehbarkeit und Dokumentation																								
3.3 Wiederverwendung und Versionierung																								
3.4 Wiederverwendung/Fehlervermeidung neue Probleme																								
4. Implementierung des Demonstrators und Tests																								
4.1 Implementierung																								
4.2 Anwendertests																								
5. Projekt-Dokumentation																								
6. Veröffentlichung und Transfermaßnahmen																								

Personal	
Wissenschaftlicher Mitarbeiter	1 1
	$\Sigma = 24$ MM
Studentische Hilfskraft (86h/Monat)	1 1
	$\Sigma = 24$ MM

Bild 2: Zeitplan und Arbeitseinsatz

4. Plan zum Ergebnistransfer in die Wirtschaft

	Zeitraum	Maßnahme	Ziel/Bemerkung
Während der Laufzeit (01/2011 – 12/2012)	halbjährlich	Regelmäßige Präsentation und Diskussion der aktuellen Forschungsergebnisse im Projektbegleitenden Ausschuss (PA) , der sich maßgeblich aus KMU aus dem Umfeld der Prozessindustrie zusammensetzt.	Fortschrittsbericht, Diskussion, Festlegungen, Abstimmung der nächsten Schritte
	4. Quartal 2011	Publikation des erarbeiteten Methodenkatalogs (Wissensbasis), unabhängig vom Entwicklungsstand der Architekturentwicklung und der Implementierung des Software-Demonstrators.	Nach Absprache mit dem PA kann dies z.B. in Form eines Wiki im Internet umgesetzt werden, um so ein möglichst breites Publikum zu erreichen und eine Erweiterung durch Beiträge von Nutzern zu ermöglichen.
	2011/2012	KMU-Anwendertests (evtl. in Form von Workshops) zur Erprobung des Methodenkatalogs und der verschiedenen Ausbaustufen des Software-Demonstrators.	Erfahrungen durch Anwendung der Ergebnisse auf vielfältige praktische Problemstellungen fließen in die Weiterentwicklung der methodischen Konzepte und der Software ein. So wird insbesondere die einfache Nutzbarkeit der Ergebnisse gewährleistet.
	2012	Vorstellung der Projektergebnisse auf thematisch passenden Ausschusssitzungen und Aussprachetagen von ProcessNET (z.B. Fachausschuss Prozess- und Anlagentechnik) und des VDI (z.B. GMA-FA 6.22) sowie auf der Achema 2012.	

	2012	Aufsätze in praxisnahen Zeitschriften wie CITplus oder Process sowie Bekanntmachung über fachlich einschlägige Internetportale wie etwa das der OPC Foundation.	
	2012	Direkte Ansprache / Beratung einschlägiger Institute und Lehrstühle , um die Plattform zur Datenaufbereitung im Hochschulbereich zu etablieren.	Kontakte zum Lehrstuhl für Prozesstechnik (AVT.PT, RWTH Aachen), zur Hochschule Niederrhein und zur Fachhochschule Frankfurt am Main existieren bereits.
	2012	Abschlussbericht	Zusammenstellung der Forschungsergebnisse und Software-Dokumentation.
Nach Abschluss des Vorhabens	2013	(Mit-)Veranstaltung von Seminaren zur Datenanalyse, etwa im Haus der Technik, Essen, oder im VDI Wissensforum.	
	ab 2013	Einsatz der Forschungsergebnisse in der akademischen Ausbildung , sowohl in der Lehre etwa im Rahmen von Praktika als auch in der Forschung zur Analyse von Messdaten (durch die Hochschulen, siehe Beratung von Instituten).	Über die Absolventen werden die Projektergebnisse nachhaltig in die mittelständische Wirtschaft weiter getragen.
	ab 2013	Längerfristige Pflege der Forschungsergebnisse und Moderation des Ergebnistransfers als Teil der Vereinsaktivitäten von AixCAPE e.V., dessen in der Satzung festgelegtes Ziel genau solche Forschungstransfermaßnahmen sind.	AixCAPE wird die Kontakte zu Firmen, die im Rahmen anderer Vereinsaktivitäten bereits bestehen oder künftig entstehen und die auch über Deutschland hinausgehen, zur Bekanntmachung der Projektergebnisse nutzen. Der Internetauftritt von AixCAPE soll genutzt werden, um den Methodenkatalog, die Software und die später entwickelten Erweiterungen zentral für die Allgemeinheit frei verfügbar zu machen.
	ab 2013	Einige Firmen aus dem PA sind beratend tätig, konzipieren informationstechnische Lösungen für Produktionsprozesse und setzen diese um. Es wird erwartet, dass durch diese Dienstleister als Multiplikatoren die Software eine zunehmende Verbreitung findet.	Die freie Verfügbarkeit der Software ist diesem Ziel dienlich und führt zu einer nachhaltigen, allgemeinen Bekanntheit der Methodik.

Tabelle 1: Tabellarische Übersicht zu den geplanten Transfermaßnahmen

Da der Methodenkatalog und der Software-Demonstrator Ergänzungen zu der in den Unternehmen bereits eingesetzten Software darstellen und bei der Software-Entwicklung ausschließlich frei oder als Teil des Windows-Betriebssystems verfügbare Software-

Komponenten verwendet werden sollen, fallen keine zusätzlichen Beschaffungskosten für Soft- oder Hardware an, was den Einsatz gerade in den KMU begünstigen wird.

5. Nutzen und wirtschaftliche Bedeutung der Forschungsergebnisse für KMU

5.1. Voraussichtliche Nutzung der angestrebten Forschungsergebnisse in KMU

Mit dem vorgestellten Ansatz sind insbesondere dann Vorteile erzielbar, wenn die analysierten Prozessdaten die Produktqualität oder die Produktionskosten direkt beeinflussen. Das betrifft alle verfahrenstechnischen Produktionsprozesse

- in den Fachgebieten: Chemie, Verfahrenstechnik, Lebensmitteltechnik, Umwelttechnik, Automatisierungstechnik, Informations- und Kommunikationstechnik;
- in den Wirtschaftszweigen: Ernährungsgewerbe, Papiergewerbe, Mineralölverarbeitung, Chemische Industrie, Maschinenbau, Energie- und Wasserversorgung, Erbringung von Dienstleistungen überwiegend für Unternehmen.

Auch wenn im Rahmen des Projektes primär die Prozessindustrie im Fokus steht, so sollte die Anwendbarkeit der Projektergebnisse auch über dieses (für sich bereits sehr große) industrielle Segment hinausgehen. Es soll im Rahmen des Ergebnistransfers durchaus auch versucht werden, Industriesegmente über die Prozessindustrie hinaus zu erreichen.

5.2. Voraussichtlicher Beitrag zur Steigerung der Wettbewerbsfähigkeit der KMU

Gerade die KMU tun sich heute noch schwer, ökonomische Vorteile aus vorhandenen Prozessdaten zu erzielen. Im Unterschied zu Großunternehmen verfügen sie nicht über spezialisierte Fachabteilungen, die Prozessdatenanalysen regelmäßig durchführen, oder gar eine eigene Infrastruktur für eine qualitätsgesicherte und effiziente Bearbeitung dieser Aufgabenstellungen. Deshalb scheuen die KMU den enormen initialen Aufwand (oft bis zu 80%), der bereits zu Beginn einer Prozessdatenanalyse bei der Datenaufbereitung anfällt. Sie lassen das in den Daten enthaltene Potential zur Prozessoptimierung damit weitestgehend ungenutzt.

Die Ergebnisse dieses Forschungsvorhabens stärken die Wettbewerbsfähigkeit von KMU in den o.g. Sektoren, weil der Zeitaufwand für die Aufbereitung drastisch reduziert und die Qualität der Arbeitsergebnisse deutlich erhöht wird. Dies führt u.a. auch zu einem effizienteren Personaleinsatz, einer höheren Planungssicherheit und erlaubt eine Fokussierung auf die eigentliche Optimierung der Produktionsprozesse, wie die Analyse, die Verbesserung und die Überwachung der Prozessgüte oder die Fehlererkennung und -diagnose. Prozessdatenanalysen werden auch nicht besonders geschulten Sachbearbeitern ermöglicht, für die solche Projekte nicht zum Tagesgeschäft gehören. Sie können sich einen schnelleren Überblick über die Datenlage verschaffen, sparen Einarbeitungs- und Bearbeitungszeit und vermeiden fehlerhafte Prozessdatenanalysen. Die entwickelte Methodik trägt zur nachhaltigen Know How-Sicherung, zur Qualitätssicherung und schließlich zur Effizienzsteigerung bei allen Tätigkeiten bei, die auf der Verarbeitung von Prozessdaten basieren.

Der Antragsteller rechnet damit, dass anfangs besonders *dienstleistende KMU* direkt von den Ergebnissen dieses Projekts profitieren werden. Beispielsweise zählen Dienstleistungs-

unternehmen, die Prozessdatenanalysen mit spezialisierten Methoden, etwa neuronalen Netzen, anbieten, vorwiegend zu den KMU. Durch eine Fokussierung ihres Aufwandes von der Aufbereitung weg hin zu ihren Kernkompetenzen werden sie im Wettbewerb gestärkt. Von Bedeutung sind auch die Verbesserungen im Berichtswesen: Eine detaillierte Aufstellung der durchgeführten Aufbereitungs- und Analyseschritte sowie der getroffenen Annahmen für den Kunden kann bislang nur mit hohem Aufwand erstellt werden. Insbesondere für Bearbeiter, die derartige Analysen nicht routinemäßig durchführen, ist die Akquisition einschlägiger Erfahrung mühsam und zeitaufwändig.

Der erfolgreiche Einsatz bei Dienstleistern im Auftrag ihrer Kunden wird indirekt auch bei den *produzierenden KMU* Nutzen erzielen und längerfristig zu einer Verbreitung in diesem Bereich führen. Die direkte Nutzung bei den produzierenden KMU wird neben der freien Verfügbarkeit des Methodenkatalogs und der entwickelten Software, die wie im Abschnitt 5.3 näher erläutert unter einer freien Lizenz verfügbar gemacht werden, auch durch die Weiterbildungsangebote von AixCAPE e.V. erleichtert.

In der Vergangenheit haben sich im Umfeld solcher freien Produkte neue Betätigungsmöglichkeiten gerade für *kleine und mittlere Softwareunternehmen* ergeben; prominente Beispiele sind das Betriebssystem Linux, der Browser Firefox, die Software-Entwicklungsplattform Eclipse oder das Content Management System für Internetauftritte Typo3. Aber auch im automatisierungstechnischen Bereich gibt es Beispiele wie etwa die ACPLT-Technologien (Münnemann, 2002). Konkret können die KMU hier als Dienstleister für das verarbeitende Gewerbe auftreten und sich neue Betätigungsfelder erschließen, indem sie für Einführung, Pflege und Wartung des Systems sorgen oder unternehmensspezifische Erweiterungen realisieren. Weitere Betätigungsfelder sind die Bereitstellung von Komponenten zur Datenaufbereitung, zur Anbindung weiterer Datenquellen oder sogar kompletter Arbeitsprozesse. Das Anwendungsfeld ist hierbei nicht auf die Datenaufbereitung beschränkt, sondern kann wegen der Offenheit des Konzeptes deutlich darüber hinausgehen und andere Anwendungsbereiche erschließen.

Schließlich profitieren auch *Anbieter von Spezialsoftware* zur Datenanalyse und Modellierung, die meist zu den KMU zählen, von den Forschungsergebnissen, wenn ihre Werkzeuge durch die verbesserte Anbindung an die bestehende informationstechnische Infrastruktur attraktiver werden und durch den reduzierten Aufwand verstärkt angewandt werden.

5.3. Aussagen zur voraussichtlichen industriellen Umsetzung der FuE-Ergebnisse nach Projektende

Das angestrebte Gesamtkonzept lässt eine hohe Verbreitung der Ergebnisse bei den KMU der bereits genannten Sparten erwarten, da eine Integration in laufende Arbeiten *leicht, schnell und mit geringen bis vernachlässigbaren finanziellen Kosten* verbunden ist. Der ökonomische Vorteil hingegen, den eine durchgehende Unterstützung von Schritten zur Datenaufbereitung bietet, ist enorm. Diese Einschätzung wird auch von den KMU im PA bekräftigt, die bereits erhebliches Interesse am Einsatz der Ergebnisse im eigenen Unternehmen oder auch bei Kunden deutlich gemacht haben.

Dabei bildet der *Methodenkatalog*, der der Allgemeinheit unabhängig von der entwickelten Software frei zur Verfügung gestellt wird, die Grundlage. Bereits ohne entsprechende Software kann dieser Katalog als eine Art *Orientierungshilfe zur Datenvorbereitung* eingesetzt werden. Er beschreibt etablierte Vorgehensweisen in strukturierter Form, bietet auch dem Nutzer ohne Expertenwissen einen Einstieg und Überblick und trägt somit nicht zuletzt auch zu einer größeren Akzeptanz der entwickelten Software bei. Der Installationsaufwand des Katalogs als Hilfe-Datei

auf dem Arbeitsplatzrechner ist vernachlässigbar und entfällt sogar, falls die Wissensbasis wie angedacht eine internetbasierte Erweiterung durch Beiträge von Nutzern möglich macht. Auf diese Weise wird außerdem eine kontinuierliche Aktualisierung der Wissensbasis ermöglicht.

Es ist geplant, die entwickelte *Software* unter einer freien Lizenz verfügbar zu machen. Die genaue Detaillierung der Lizenzbedingungen ist in Zusammenarbeit mit dem PA als Teil des Projektes vorzunehmen. Eine Open Source-Lizenz wie die GNU General Public License (Free Software Foundation, 2007), bei der auch der Quelltext verfügbar gemacht wird, hätte folgende Vorteile:

- Wegen der niedrigen laufenden Kosten rechnet sich der Einsatz bei den KMU gerade auch bei einer nur gelegentlichen Verwendung. Entscheidend ist allerdings, dass einige praxisrelevante Randbedingungen eingehalten werden, wie die Sicherstellung der Lauffähigkeit auf aktuell in der Industrie eingesetzter Hard- und Software (Standard-PC mit Windows) oder der Einsatz von Software-Komponenten, die frei oder als Teil des Windows-Betriebssystems verfügbar sind.
- Attraktivität für den Einsatz in Hochschulen und Forschungseinrichtungen wegen der geringen Nutzungskosten und der Anpass- und Erweiterbarkeit; über die Absolventen ergibt sich dadurch eine größere Verbreitung besonders auch bei den KMU.
- Erhöhte Bereitschaft von Software-Herstellern, mit ihren Produkten eine quelltextoffene Plattform zu unterstützen, da dann keine Abhängigkeit zu anderen Anbietern besteht und keine Lizenzgebühren anfallen.
- Zukunftsfähigkeit, da die Quelltexte offen sind und die Verfügbarkeit nicht vom wirtschaftlichen Erfolg eines einzelnen Anbieters abhängt.

Insgesamt erleichtert der verfolgte Ansatz zur interaktiven Integration mit Arbeitsprozessunterstützung den Einsatz der Software in einer gewohnten Arbeitsumgebung mit bereits existierenden Spezialwerkzeugen.

6. Durchführende Forschungsstelle

AixCAPE e.V.

Intzestr. 1

52072 Aachen

Tel.: 0241/80 94860, Fax.: 0241/80 92325, E-Mail: soemers@aixcape.org

6.1. Leitung der Forschungsstelle

Prof. Dr.-Ing. Wolfgang Marquardt, Vorsitzender des Vorstands von AixCAPE e.V.

6.2. Projektleitung

Dr. rer. nat. Marcus Soemers

7. Literaturverzeichnis

Referenzen zu Produkten sind unter Abschnitt 7.2 separat aufgeführt.

7.1. Literatur

- Aalst van der, W. und K. van Hee (2002): *Workflow Management - Models, Methods, and Systems*, MIT Press, Cambridge, Massachusetts
- Alsmeyer, F. und W. Marquardt (2004): Trendbasierte Aufbereitung von Prozessdaten – Anwendung auf praktische Problemstellungen. Vortrag anlässlich der *GVC/DECHEMA-Jahrestagungen 2004*, 12.-14. Oktober 2004, Karlsruhe
- Alsmeyer, F. (2007): Durchgängige Nutzung von Prozessdaten im Lebenszyklus verfahrenstechnischer Anlagen, *atp - Automatisierungstechnische Praxis* 7/2007
- Altintas, I., C. Berkley, E. Jaeger, M. Jones, B. Ludäscher und S. Mock (2004): Kepler: An Extensible System for Design and Execution of Scientific Workflows, system demonstration, 16th *Intl. Conf. on Scientific and Statistical Database Management (SSDBM'04)*, 21-23 Juni 2004, Santorini Island, Greece
- ASAM e.V. (2010): Association for Standardisation of Automation and Measuring Systems, <http://www.asam.net/> (Zugriff 11.1.2010)
- Batres, R., Y. Naka und M. L. Lu (1999): A multidimensional design framework and its implementation in an engineering design environment, *Concurrent Engineering: Research and Applications* 7 (1), 43-54
- Baumgarten, B. (1996): *Petri-Netze*, Spektrum Akademischer Verlag
- Berthold, M. et al. (2007): KNIME: The Konstanz Information Miner, *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007, Freiburg, Germany)*, Springer
- Booch, G., James Rumbaugh und Ivar Jacobson (1999): *Das UML-Benutzerhandbuch*. Addison-Wesley
- Brachman, R. und H. Levesque (2004): *Knowledge Representation and Reasoning*, Morgan Kaufmann, San Fransisco
- Cauvin, S. (2002): *CHEM: Advanced decision support system for chemical/petrochemical manufacturing processes*, von der EU unter G1RD-CT-2001-00466 gefördertes Projekt, 1998-2002
- Chiang, L.H., R.J. Pell, und M.B. Seasholtz (2003): Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control*, Vol.13 no.5
- Chonacky, N und D. Winch (2005): Maple, Mathematica, and Matlab: The 3M's without the Tape, *Computing in Science and Engineering*, vol. 07, no. 1, 8-16
- Data Mining Group (2010): <http://www.dmg.org/> (Zugriff 11.1.2010)
- De Roure, D. und Goble, C. (2009): Software Design for Empowering Scientists, *IEEE Software*, Januar 2009
- Dünnebier, G. und M. v. Felde (2003): Performance Monitoring - Ein entscheidender Beitrag zur Optimierung der Prozessführung, *Chemie Ingenieur Technik* 75 (5), 528-533
- Eggersmann, M., B. Kausch, H. Luczak, W. Marquardt, C. Schlick, N. Schneider, R. Schneider, and M. Theißen (2008): Work process models. In: Nagl, M., Marquardt, W. (Eds.): Collaborative and distributed chemical engineering : from understanding to substantial design process support; results of the IMPROVE project , Springer Berlin / Heidelberg, (2.4), 126-152

- Estler, M. und Soravia, S. (1998): Expedition ins Datenreich: Wie kann man Informationen in Produkt- und Prozessdaten umfassend erschließen?, *Chemie Ingenieur Technik* 70 (11), 1388-1392
- Fischer, L. (2005): *Workflow Handbook 2005*, Published by Layna Fischer in collaboration with the Workflow Management Coalition, Lighthouse Point, Florida
- Foltz, C., Killich, S., Wolf, M., Schmidt, L., Luczak, H. (2001): Task and Information Modeling for Cooperative Work, in: Smith, M.J., Salvendy, G. (Hrsg.) "Systems, Social and Internationalization Design Aspects of Human-Computer Interaction", *Proceedings of HCI International*, vol. 2, Lawrence Erlbaum Associates, Mahwah, S. 172-176
- Free Software Foundation (2007): GNU GENERAL PUBLIC LICENSE, Version 3, Juni 2007, Online verfügbar auf <http://www.fsf.org/licensing/licenses/gpl.html> (Zugriff 12.1.2010)
- Freire, J., D. Koop, E. Santos and C. T. Silva (2008): Provenance for Computational Tasks: A Survey, *Computing in Science and Engineering*, vol. 10, no. 3, 11-21
- Hai, R., M. Theißen und W. Marquardt (2009): An Integrated Ontology for Operational Processes. In: J. JeŚowski and J. Thullie (Editors): *European Symposium on Computer Aided Process Engineering - 19*, Elsevier, 1087-1091
- Häck, M. und M. Köhne (1998): Lokale und globale Validierung der Messsignale kontinuierlich arbeitender Prozessanalytoren, *Automatisierungstechnik* 46, 135-325
- ISO 10303 (1998): *Part 231, Process Engineering Data: Process Design and Process Specifications of Major Equipment*. Committee Draft ISO TC184/SC4/WG3 N740
- Jaekle, C. und M. Krell (2003): Data Mining with Process Data?, Vortrag auf der *ACHEMA* 22.05.2003, Kurzfassung online verfügbar auf http://www.umesoft.de/talks_multivar_anal.pdf (Zugriff 14.1.2005)
- Köppen-Seliger, B., T. Marcu, M. Capobianco, S. Gentil, M. Albert and S. Latzel (2003): MAGIC: Ein integriertes Konzept zum Diagnosemanagement und Operator Support, *Automatisierungstechnische Praxis*, Vol.5, No.5, pp. 79-83
- Kourti, T. (2002): Process Analysis and Abnormal Situation Detection: From Theory to Practice, *IEEE Control Systems Magazine*, October 2002, 10-25
- Lawton, G. (2009): Users Take a Close Look at Visual Analytics, *IEEE Computer*, Februar 2009
- Majithia, Shalil, Matthew S. Shields, Ian J. Taylor, und Ian Wang (2004): Triana: A Graphical Web Service Composition and Execution Toolkit. In: *Proceedings of the IEEE International Conference on Web Services (ICWS'04)*, S. 514-524ff
- Martens, H. und T. Naes (1989): *Multivariate Calibration*, Wiley, New York
- Münnemann A., Heeg M., Epple U. (2002): Stein auf Stein : ACPLT - Ein leittechnischer Baukasten. In: *MessTec & Automation*, 10 (2002) 1-2, GIT Verlag GmbH & Co. KG, Darmstadt, Germany, S. 54 - 56
- Nagl, M. und Marquardt, W. (2008): Collaborative and distributed chemical engineering : from understanding to substantial design process support; results of the IMPROVE project, Springer Berlin / Heidelberg
- Nohr, M. und M. Boll (2003): "Controller Performance Assessment" – Eine Möglichkeit zur Verbesserung der Prozessführung, *Chemie Ingenieur Technik* 75 (12), 1828-1831
- Object Management Group (2010): <http://www.omg.org/technology/cwm/> (Zugriff 11.1.2010)

Oestereich, B. (2004): *Objektorientierte Softwareentwicklung mit der UML 2.0*, Oldenbourg Verlag

OPC Foundation (2003): *OPC Historical Data Access Specification Version 1.20*, online verfügbar auf <http://www.opcfoundation.org> (Zugriff 12.1.2010)

OPC Foundation (2009): *OPC UA Part 8 - Data Access 1.01 Specification*, online verfügbar auf <http://www.opcfoundation.org> (Zugriff 12.1.2010)

Shewhart, W. A. (1986): *Statistical Method from the Viewpoint of Quality Control*. Dover Publications Inc., New York

Stevens, R., A. Robinson, und C.A. Goble (2003): myGrid: Personalised Bioinformatics on the Information Grid, *Bioinformatics* Vol. 19 Suppl. 1, i302-i304

Theißen, M., R. Hai, J. Morbach, R. Schneider und W. Marquardt (2008a): Scenario-based analysis of industrial work processes. In: Nagl, M., Marquardt, W. (Eds.): *Collaborative and distributed chemical engineering: from understanding to substantial design process support; results of the IMPROVE project*, Springer Berlin / Heidelberg, (5.1), 433-450

Theißen, M., R. Hai und W. Marquardt (2008b): Design process modeling in chemical engineering. *Journal of Computing and Information Science in Engineering*, 2008, 8(1), 011007 (9 Seiten)

Theißen, M., R. Hai und W. Marquardt (2008c): Computer-assisted work process modeling in chemical engineering. In: Nagl, M., Marquardt, W. (Eds.): *Collaborative and distributed chemical engineering: from understanding to substantial design process support; results of the IMPROVE project*, Springer Berlin / Heidelberg, (7.3), 656-665

Theißen, M., R. Hai und W. Marquardt (2010): A Framework for Work Process Modeling in the Chemical Industries. Eingereicht bei *Computers and Chemical Engineering*

Venkatasubramanian, V., R. Rengaswamy, S. N. Kavuri und K. Yin (2003): A review of process fault detection and diagnosis, Part III: Process history based methods, *Computers and Chemical Engineering* 27, 327-346

Wold S., N. Kettaneh, H. Friden und A. Holmberg (1998): Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemom. Intell. Lab. Syst.* 44 (1998), 331-340

Workflow Management Coalition (2010): Workflow Management Coalition Website auf <http://www.wfmc.org/> (Zugriff 11.1.2010)

7.2. Informationen zu relevanten Produkten

Alyuda Research Inc. (2010): <http://www.alyuda.com/neural-networks-software.htm> (Zugriff 11.1.2010)

AspenTech (2002): *InfoPlus.21™ 5.0 Database User's Manual*

CAMO Inc. (2010): <http://www.camo.com/> (Zugriff 11.1.2010)

IBM (2010): <http://www-01.ibm.com/software/data/iminer/> (Zugriff 11.1.2010)

Labvantage Solutions Inc. (2010): <http://www.labvantage.com/> (Zugriff 11.1.2010)

Maplesoft (2010): <http://www.maplesoft.com/> (Zugriff 11.1.2010)

MathWorks (2009): *Using Matlab*, Version 7, The MathWorks Inc., Natick, USA.

Microsoft (2007): *Excel 2007*, Microsoft Corp., Redmond, Washington, USA.

National Instruments (1998): *Labview Manual*, Austin, Texas, USA.

National Instruments (2010): <http://www.ni.com/diadem/> (Zugriff 11.1.2010)

OSISoft, Inc. (2003): *PI Server Reference Guide*, Version 3.4

ObjectWeb (2010): <http://shark.objectweb.org/> (Zugriff 11.1.2010)

SAS Institute Inc. (2010): <http://www.jmp.com/> (Zugriff 11.1.2010)

Siemens AG (2010): <http://www.siemens.de/simatic-wince> (Zugriff 25.3.2010)

SPSS Inc. (2010a): <http://www.spss.com/statistics/> (Zugriff 11.1.2010)

SPSS Inc. (2010b): <http://www.spss.com/software/modeling/modeler/> (Zugriff 11.1.2010)

StatSoft (2010): <http://www.statsoftinc.com/> (Zugriff 11.1.2010)

Thermo Electron Corp. (2010): <http://www.thermo.com/> (Zugriff 11.1.2010)

ThinkAnalytics (2010): <http://www.thinkanalytics.com/productServices/edm/index.htm> (Zugriff 11.1.2010)

Umetrics Inc. (2010): http://www.umetrics.com/default.asp/pagename/Software_OverView/c/1 (Zugriff 11.1.2010)

Wolfram Research Inc. (2010): <http://www.wolfram.com/> (Zugriff 11.1.2010)

Ort, Datum

Rechtsverbindliche Unterschrift des Leiters
und Stempelabdruck
der federführenden Forschungsstelle