

Investigations of concept development to improve data quality in research information systems

Otmane Azeroual
Deutsches Zentrum für Hochschul-
und Wissenschaftsforschung GmbH
Schützenstraße 6a, D-10117 Berlin
Azeroual@dzhw.eu

Gunter Saake
Otto von Guericke Universität
Magdeburg, Fakultät Informatik
P.O. Box 4120,
D-39016 Magdeburg
Saake@ovgu.de

Mohammad Abuosba
Hochschule für Technik und
Wirtschaft Berlin
Wilhelminenhofstraße 75a, D-12459
Berlin
Mohammad.Abuosba@HTW-
Berlin.de

ABSTRACT

The implementation of research information systems at German universities and research institutions is currently a topical subject. With their help, the documentation and reporting of the research activities of the respective institution can be supported and a significant part of the data incurring there can be managed. As there are usually many data sources available and the collection, transmission, and integration of research information in different research information systems can lead to different data errors which can have various negative effects on data quality. It is necessary to recognize these errors early and to handle them efficiently, so that users can get better results. For this reason, this paper examines data quality in research information systems and introduces measurement and enhancement methods that enable organizations to secure their quality of data.

Untersuchungen zur Konzeptentwicklung für eine Verbesserung der Datenqualität in Forschungsinformationssystemen

Otmane Azeroual

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Schützenstraße 6a, D-10117 Berlin

Azeroual@dzhw.eu

Gunter Saake

Otto von Guericke Universität
Magdeburg, Fakultät Informatik
P.O. Box 4120,
D-39016 Magdeburg

Saake@ovgu.de

Mohammad Abuosba

Hochschule für Technik und
Wirtschaft Berlin
Wilhelminenhofstraße 75a, D-12459
Berlin

Mohammad.Abuosba@HTW-
Berlin.de

ABSTRACT

Die Implementierung von Forschungsinformationssystemen an deutschen Hochschulen und Forschungseinrichtungen ist derzeit ein aktuelles Thema. Mit ihrer Hilfe können die Dokumentation und Berichterstattung der Forschungsaktivitäten der jeweiligen Einrichtung unterstützt und ein bedeutender Teil der dort anfallenden Daten verwaltet werden. Da hierfür meist sehr viele Datenquellen zur Verfügung stehen und bei der Erfassung, der Übertragung sowie der Integration von Forschungsinformationen in unterschiedlichen Forschungsinformationssystemen verschiedene Datenfehler entstehen können, welche vielfältige negative Wirkungen auf die Datenqualität haben können, ist es notwendig, diese Fehler frühzeitig zu erkennen und effizient zu behandeln, sodass die Nutzer bessere Ergebnisse erhalten können. Aus diesem Grund wird in diesem Beitrag die Datenqualität in Forschungsinformationssystemen untersucht sowie Methoden zur Messung und Verbesserung vorgestellt, mit denen Einrichtungen ihre Qualität der Daten sichern können.

Keywords

Current Research Information Systems (CRIS);
Forschungsinformationssystem (FIS); Datenmanagement;
Forschungsinformationen; Datenqualität; Nutzerakzeptanz;
Qualitätsmessung; Qualitätsverbesserung; Qualitätskontrolle

1. EINLEITUNG

Das Thema „Forschungsdatenbank und Forschungsinformationssysteme“ ist keinesfalls neu. Heutzutage hat sich sowohl in der Praxis als auch in der Forschung der Begriff FIS etabliert. Im internationalen Raum wird für Forschungsinformationssysteme der Begriff des „Current Research Information Systems“ (CRIS) verwendet. Unter einem Forschungsinformationssystem wird eine spezialisierte Datenbank oder ein „spezielles föderiertes Informationssystem“ [6] verstanden, mit dessen Hilfe Informationen zu Forschungsaktivitäten und Forschungsergebnissen erhoben, verwaltet und bereitgestellt werden. Die hierbei betrachteten Informationen stellen Metadaten über Forschungsaktivitäten wie z.B. Projekte, Drittmittel, Patente, Partner, Preise, Publikationen, Promotionen und Habilitationen etc. dar und werden als Forschungsinformationen bzw. Forschungsdaten bezeichnet.

Forschungsinformationssysteme ermöglichen eine ganzheitliche Darstellung der Forschungsaktivitäten und -ergebnisse an einer wissenschaftlichen Einrichtung. Sie bilden die Forschungsaktivitäten nicht nur für Einrichtungen, sondern auch für Forscher aktuell, zentral und übersichtlich ab. Durch die zentrale Abbildung im System ist eine Arbeitserleichterung für die Forschenden möglich. Daten werden mit dem FIS einmal eingegeben und können mehrfach verwendet werden, z.B. auf Websites, für Projektanträge oder Berichtsprozesse. Eine doppelte Datenhaltung und damit eine Mehrarbeit für die Anwender soll vermieden werden.

Die Bausteine einer FIS-Architektur kann als dreistufiges Gebilde gesehen werden. Diese werden mit folgenden Komponenten unterschieden:

- Datenzugriffsschicht
- Anwendungsschicht
- Präsentationsschicht

Die folgende Abbildung (vgl. Abb.1) gibt einen Überblick über die einzelnen Bausteine und zeigt, welche Komponenten zu welchem Prozessschritt gehören.

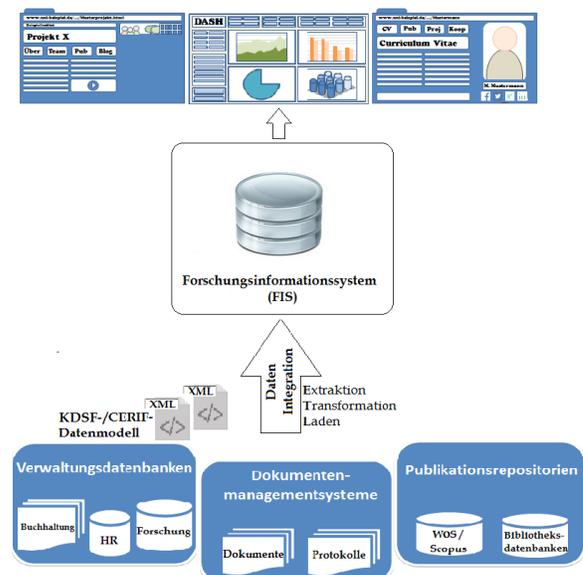


Abbildung 1: FIS-Architektur (eigene Darstellung)

In der *Datenzugriffsschicht* befinden sich die internen und externen Datenquellen (operative Systeme). Diese Ebene enthält beispielsweise Datenbestände aus der Verwaltung oder Publikationsrepositorien von Bibliotheken, Identifier wie z.B. ORCID oder bibliographische Daten aus dem Web of Science bzw. Scopus etc.

Angebote für die standardisierte Erhebung, Vorhaltung und den Austausch von Forschungsinformationen in FIS sind das Kerndatensatz-Forschung (KDSF)-Datenmodell und das Common European Research Information Format (CERIF)-Datenmodell [12]. In diesen beiden Modellen werden die Entitäten sowie deren Beziehung zueinander beschrieben.

Eine Befüllung des FIS erfolgt über einen klassischen ETL-Prozess.

Die *Anwendungsschicht* enthält das Forschungsinformationssystem und dessen Anwendungen, die die auf der zugrundeliegenden Ebene gehaltenen Daten zusammenführen, verwalten und analysieren. In der *Präsentationsschicht* sind die zielgruppenspezifischen Aufbereitungen und Darstellungen der Analyseergebnisse für den Anwender abgebildet. Diese werden mithilfe von den Business-Intelligence-Werkzeugen in Form von Berichten verfügbar gemacht. Neben diversen Möglichkeiten des Reportings lassen sich hier ebenfalls Portale und Webseiten der Einrichtungen befüllen.

Um ein Forschungsinformationssystem einzuführen, bedeutet das für wissenschaftliche Einrichtungen, ihre benötigten Informationen über Forschungsaktivitäten und Forschungsergebnisse in gesicherter Qualität zur Verfügung zu stellen. Eine einmalige Bereinigung ist dabei nicht ausreichend, Daten müssen fortlaufend gepflegt werden. Die wachsenden Datenmengen und die zunehmende Anzahl an Quellsystemen wird für Einrichtungen immer mehr zu einer Herausforderung. Zum Beispiel können sich manche Datenfehler wie z.B. fehlende Werte, Dubletten, Schreibfehler, fehlerhafte Formatierung und Widersprüche etc. sich über verschiedene Bereiche erstrecken und schwer auffindbar sein [2]. Wenn die Nutzer nicht in der Lage sind, auf die am dringendsten benötigten Informationen zuzugreifen und schnelle Entscheidungen zu treffen, sinkt der Wert der verwendeten Daten und das Vertrauen in das FIS und dessen Akzeptanz.

Vor diesem Hintergrund sollte das Thema Datenqualität mit hoher Priorität behandelt werden. Hierzu werden in diesem Beitrag Methoden zur Messung verschiedener Dimensionen der Datenqualität und zu deren Verbesserung betrachtet, sodass diese dann an nutzenden Einrichtungen vorgestellt werden können.

2. DATENQUALITÄTSPROBLEME IN FORSCHUNGSSYSTEMEN

Daten in einem Datenbanksystem zu erfassen, zu integrieren, zu speichern und zu analysieren, ist an sich ein normaler Vorgang. In jeder Einrichtung werden Personaldaten, Informationen zu ihren wissenschaftlichen Aktivitäten, Projekten und Publikationen eingegeben und erfasst. Die Verarbeitung und Verwaltung dieser Daten müssen in der Regel in guter Qualität sein, damit die Nutzer qualitative Ergebnisse erhalten können.

Im Kontext der FIS wird die Datenqualität als die Eignung dieser Daten zur Verwendung bei bestimmten geforderten Verwendungszielen definiert. Diese müssen fehlerfrei, vollständig, korrekt, aktuell und konsistent sein. Anforderungen können dabei von unterschiedlichen Beteiligten aufgestellt

werden, z.B. insbesondere von Nutzern eines FIS, aber auch vom FIS-Administrator. Daten von schlechter Qualität enthalten Fehler, welche im FIS analysiert und anschließend durch eine Datenbereinigung behoben werden müssen. Im Folgenden wurde durch eine Umfrage, die europaweit durchgeführt wurde, die typischen Qualitätsprobleme von Daten zum FIS ermittelt (siehe Abb. 2).



Abbildung 2: Umfrage zu den Datenqualitätsproblemen in FIS (eigene Darstellung)

3. QUALITÄTSMESSUNG

Bevor eine aussagekräftige Messung der Datenqualität durchgeführt werden kann, müssen die Dimensionen klar definiert werden. Dabei müssen die Einrichtungen entscheiden, welche Datenqualitätsdimensionen wichtig sind und wie diese gemessen werden sollen.

Im Rahmen dieses Beitrags werden nur die vier Datenqualitätsdimensionen (*Vollständigkeit*, *Aktualität*, *Korrektheit/Fehlerfreiheit* und *Konsistenz*) mit deren einfachen Metriken im Kontext FIS betrachtet (wie in der Tabelle 1 dargestellt).

Tabelle 1: Dimensionen der Datenqualität im FIS-Kontext (In Anlehnung an [5][10][13][14][18])

Datenqualitätsdimensionen	Definition	Metriken
Vollständigkeit	Forschungsinformationen sind vollständig, wenn sie nicht fehlen und zu den festgelegten Zeitpunkten in den jeweiligen Prozess-Schritten zur Verfügung stehen.	$Q_{\text{Vollständigkeit}} = 1 - \frac{\text{(Anzahl unvollständiger Einheiten)}}{\text{(Anzahl überprüfter Einheiten)}}$ <p>Erreichungsgrad = 0 - 100%</p>
Korrektheit / Fehlerfreiheit	Forschungsinformationen sind korrekt und fehlerfrei, wenn sie mit der Realität übereinstimmen.	$Q_{\text{Korrektheit}} = 1 - \frac{\text{(Anzahl unkorrekter in Dateneinheiten)}}{\text{(Gesamtanzahl Dateneinheiten)}}$ <p>Erreichungsgrad = 0 - 100%</p>
Aktualität	Forschungsinformationen sind aktuell, wenn sie die tatsächliche Eigenschaft des beschriebenen Objektes zeitnah abbilden.	$Q_{\text{Aktualität}}(W,A) = e^{(-\text{decline}(A) \cdot \text{age}(W,A))}$ <p>Erreichungsgrad = 0 - 100%</p>
Konsistenz	Forschungsinformationen sind konsistent dargestellt, wenn sie fortlaufend auf dieselbe Art und Weise abgebildet werden.	$Q_{\text{Konsistenz}} = 1 - \frac{\text{(Anzahl inkonsistenter Einheiten)}}{\text{(Anzahl durchgeführter Konsistenzprüfungen)}}$ <p>Erfüllungsgrad = 0 - 100%</p>

Diese vier Dimensionen wurden deshalb untersucht, weil sie zum einen in wissenschaftlichen Veröffentlichungen und in einer internationalen Umfrage, besonders intensiv diskutiert wurden (siehe Abb. 3) und zum anderen, hat sich bei diesen vier ausgewählten Metriken herausgestellt, dass diese außergewöhnlich einfach zu messen sind und eine besonders repräsentative Abbildung der Berichterstattung für die FIS-Nutzer abbilden bzw. zu einer verbesserten Entscheidungsgrundlage führen [5].

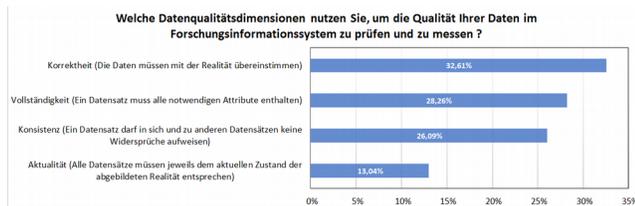


Abbildung 3: Umfrage der wichtigsten Datenqualitätsdimensionen zur Messung und Prüfung der Qualität der Daten in FIS (eigene Darstellung)

Bei den Messpunkten der Datenqualität wird auf die FIS-Architektur zurückgegriffen. Mittels dieser werden Messbereiche aufgezeigt, die für eine Messung und Überwachung der Datenqualität geeignet sind.

Die Messmöglichkeiten im FIS befinden sich in folgenden Bereichen:

- Interne und externe Datenquellen (oder auch Quellsysteme)
- Zentrales Forschungsinformationssystem
- FIS-Frontend

Das FIS sammelt Informationen über die zu Einrichtungen und ihren Wissenschaftlern affilierten Forschungsaktivitäten und Forschungsergebnisse durch eine automatisierte Synchronisierung der im Bestand vorhandenen Datenmengen mit verschiedenen externen Datenquellen [8]. Für einen automatisierten Datenimport aus bestehenden Systemen kann eine Anbindung von internen, sowie externen Anwendungssystemen realisiert werden [11]. Diese Anwendungssysteme, die zur Erfassung von Forschungsinformationen idealerweise einbezogen werden können, sind das Campusmanagement-System sowie das Identitätsmanagement-System als interne Systeme sowie öffentliche Publikations- und Projektdatenbanken [11]. Im Fall der Publikationen sind das beispielsweise Web of Science, Scopus oder PubMed, sowie das Finanzsystem der Drittmittelverwaltung für die Daten zu Drittmittelprojekten und das Personalverwaltungssystem für Informationen über das wissenschaftliche Personal.

Im Bereich der Quellsysteme ist eine Kontrolle sehr vorteilhaft, da eine Verbesserung der Datenqualität bereits in diesem Bereich dafür sorgen würde, dass die Daten für alle Systeme und Bereiche, die sie anschließend nutzen wollen, korrekt sind.

Abgesehen von der technischen Umsetzung wären im Bereich der Quellsysteme Kontrollen bei der Datenerfassung äußerst sinnvoll. Hierzu könnten Masken bzw. Formatvorlagen geschaffen werden, wie beispielsweise Datumseingaben (JJJMMTT) [1].

Einem zentralen FIS liegen die Daten der gesamten Forschungseinrichtungen vor. Aus diesem Grund eignen sich hier Messungen, die mehrere Bereiche einer Forschungseinrichtung betreffen und die miteinander verglichen werden können. Somit wäre hier die Möglichkeit gegeben, verschiedene Quellsysteme auf ihre Widerspruchsfreiheit zu prüfen [1].

Im letzten Schritt des FIS-Prozesses wird die Datenpräsentation (FIS-Frontend) dargestellt. Mittels von Portalen, Reporting und weiteren Front-End-Anwendungen, werden die aus dem System kommenden Informationen visualisiert. Hier werden dem Anwender die aufbereiteten Informationen und Analysen in übersichtlicher Form durch verschiedene Anwendungskomponenten zur Verfügung gestellt. Im Bereich FIS-Frontend können Messungen bzw. Prüfmechanismen durch Personen durchgeführt werden. Im einfachsten Fall werden Untersuchungen der Kennzahlen in Bereichen vorgenommen. Darüber hinaus können Kennzahlen der FIS-Anwendung mit den Quellsystemen verglichen werden. In diesem Fall ist eine Person mit fachlichen und technischen Kenntnissen nötig [1].

4. QUALITÄTSVERBESSERUNG

Aufgrund der Sammlung, Integration und Speicherung von unterschiedlichen internen Datenquellen der Einrichtungen und von externen Quellen im FIS sind Probleme, wie in Kapitel 2 aufgeführt worden ist, zu bewältigen. Nun gilt es, in diesem Schritt die Ursachen zu bekämpfen und für die Verbesserung der Datenqualität im FIS zu sorgen.

Der Prozess der Identifikation und Berichtigung von Fehlern mit dem Ziel, die Qualität von vorgegebenen Datenquellen im FIS zu erhöhen wird als Datenbereinigung (engl. „Data Cleaning“ oder „Data Cleansing“) bezeichnet [9][15]. Data Cleansing erfasst alle nötigen Aktivitäten, um „verunreinigte“ Daten wie zum Beispiel, nicht vollständig, inkorrekt, nicht aktuell, inkonsistent oder redundant, zu bereinigen. Der Data Cleansing Prozess lässt sich grob wie folgt gliedern [9]:

1. Definieren und Bestimmen des eigentlichen Problems bzw. Messkriteriums
2. Suchen und Identifizieren von fehlerhaften Instanzen
3. Korrektur der gefundenen Fehler

Im Rahmen der Datenbereinigung werden spezielle Methoden und Technologien innerhalb des Datenbereinigungsprozesses eingesetzt. [16] unterteilen diese in folgende Phasen:

Syntaxanalyse

Das Parsing bildet die erste kritische Komponente der Datenbereinigung und hilft dem Anwender, die Attribute genauer zu verstehen und zu transformieren. Dieser Prozess lokalisiert, identifiziert und isoliert einzelne Datenelemente, wie zum Beispiel Namen, Adressen, PLZ und Ort.

Berichtigung / Standardisierung

Hier ist es notwendig, die geparsten Daten auf ihre Korrektheit zu überprüfen, zu korrigieren und anschließend zu standardisieren. Standardisierung bildet die Voraussetzung für ein erfolgreiches Matching und es führt kein Weg an der Verwendung einer zweiten verlässlichen Datenquelle vorbei. Für Adressdaten empfiehlt sich eine postalische Validierung.

Anreicherung

Datenanreicherung bezeichnet den Prozess, der vorhandene Daten mit Daten anderer Quellen erweitert. Hier werden zusätzliche Daten hinzugefügt um bestehende Informationslücken zu schließen. Typische Anreicherungswerte sind demografische, geografische oder Adressinformationen.

Abgleich

Hier gibt es verschiedene Typen von Matching: zum Reduplizieren, zum Abgleichen gegenüber verschiedenen Datenmengen, zum Konsolidieren oder zum Gruppieren. Die Anpassung ermöglicht das Erkennen von gleichen Daten. Zum Beispiel können Redundanzen erkannt und zu weiteren Informationen verdichtet werden.

Zusammenführung

Durch die Zusammenführung werden übereinstimmende Datenelemente mit Zusammenhängen erkannt und zusammengeführt (merge).

All diese Methoden sind wesentlich für die Erreichung und Aufrechterhaltung maximaler Datenqualität im FIS. Durch die Bereinigung werden Fehler bei der Erfassung, Integration und Speicherung mehrerer Datenquellen im FIS eliminiert.

Um die kontinuierliche Überwachung zu behalten, bedarf es zur Sicherung, Verbesserung und Steigerung der Datenqualität im FIS neben den Methoden des Data Cleansing, bestimmte Maßnahmen zu ergreifen. Hierfür werden drei Maßnahmen betrachtet und die Wahl des optimalen Vorgehens hängt von der Änderungshäufigkeit der Daten und ihrer Bedeutung für den Nutzer ab, wie in Abbildung 4 dargestellt.

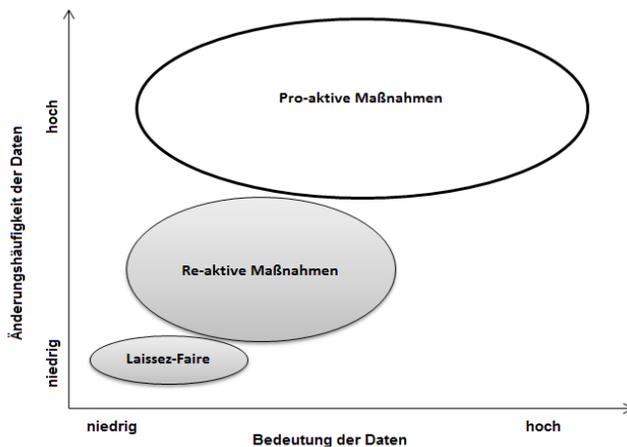


Abbildung 4: Maßnahmenportfolio (In Anlehnung an [3][17])

Laissez-Faire

Beim Laissez-Faire-Prinzip werden die auftretenden Fehler ohne Behandlung hingenommen. Das heißt sie werden schlichtweg ignoriert oder wenn, dann nur nebenbei behoben, um den Geschäftsprozess nicht zu stoppen. Dieses Prinzip gilt allerdings nur für wenige und sich selten ändernde Daten.

Re-aktive Maßnahmen

Für wichtige und sich nur selten ändernde Daten eignen sich die reaktiven Maßnahmen. Das reaktive Vorgehen beschreibt die Handlungen, die nötig werden, wenn es einmalig zu einem Datenqualitätsproblem kommt. Diese Bereinigung kann

maschinell und manuell erfolgen. Hier findet keine längere Überwachung der Datenqualität statt und die Maßnahmen werden stets nur akut und punktuell vorgenommen.

Pro-aktive Maßnahmen

Für wichtige und sich häufig ändernde Daten bieten sich dagegen proaktive Maßnahmen an. Hier werden vornehmlich Maßnahmen zur Beseitigung der Fehlerquellen und zur Verhinderung der Entstehung solcher Fehler vorgenommen. Es findet eine kontinuierliche Überwachung auf mögliche Fehler statt sowie kontinuierliche Ergreifung von Maßnahmen zu deren Beseitigung und Verhinderung.

Diese Bereinigungsmaßnahmen sind nur bedingt für die Verbesserung von unvollständigen, fehlenden, inkorrekten, inaktuellen oder inkonsistenten Daten einzusetzen. Die Entscheidung für den Einsatz einer bestimmten Maßnahme muss von einem Domänenexperten getroffen werden, welcher mit den Geschäftsprozessen seiner Organisation vertraut ist und bewerten kann, inwiefern Qualitätsmängel diese negativ beeinflussen würden [7] [17].

5. FAZIT

In der Forschung wird überwiegend von einem direkten positiven Wirkungszusammenhang zwischen Datenqualität und FIS diskutiert. Viele wissenschaftliche Einrichtungen erkannten die Wichtigkeit der Datenqualität für ihre gespeicherten Daten im FIS. Wo auch immer Daten entstehen oder verarbeitet werden, entstehen auch Datenfehler. Um diese Fehler zu entdecken und zu korrigieren, werden technische und organisatorische Methoden eingesetzt, die die Fehlerursachen bekämpfen, um für hohe Datenqualität zu sorgen.

Die Sicherung von Datenqualität im FIS lässt in diesem Beitrag darauf schließen, dass die zwei Phasen (Qualitätsdefinition/Qualitätsmessung und Qualitätsverbesserung) der geeignetste Weg zum Erfolg ist (wie in Abbildung 5 veranschaulicht).

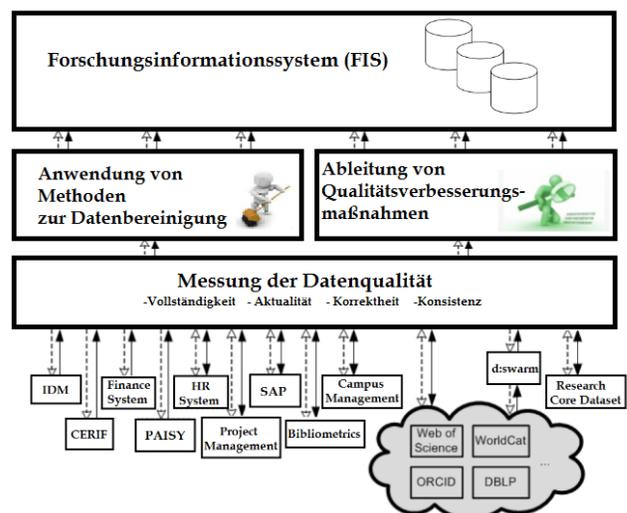


Abbildung 5: Sicherung von Datenqualität im FIS (eigene Darstellung)

Das dargestellte Modell wird hier analog bei der Datenerfassung in die Prozesse des FIS integriert und innerhalb des gesamten Zyklus, der Messung und Verbesserung, der Daten sichergestellt. Mit dessen Umsetzung kann ehemals lästiges

Qualitätsmanagement zur treibenden Kraft einer wissenschaftlichen Einrichtung werden.

6. LITERATUR

- [1] Apel, D., Behme, W., Eberlein, R. and Merighi, C. 2015. Datenqualität erfolgreich steuern. Praxislösungen für Business Intelligence-Projekte, 3., überarbeiteten und erweiterten Auflage, dpunkt.verlag, 2015.
- [2] Azeroual, O. and Abuosba, M. 2017. Improving the Data Quality in the Research Information Systems. *International Journal of Computer Science and Information Security*, 15(11): 82-86, November 2017.
- [3] Azeroual, O., Saake, G. and Abuosba, M. 2018. Data Quality Measures and Data Cleansing for Research Information Systems. *Journal of Digital Information Management*, 16(1): 12-21, February 2018.
- [4] Azeroual, O., Saake, G. and Schallehn, E. 2018. Analyzing data quality issues in research information systems via data profiling. *International Journal of Information Management*, volume 41, pages 50-56, April 2018. DOI: <https://doi.org/10.1016/j.ijinfomgt.2018.02.007>
- [5] Azeroual, O., Saake, G. and Wastl, J. 2018. Data measurement in research information systems: metrics for the evaluation of data quality. *Scientometrics*, volume 116, pages 1-20, April 2018. DOI: <https://doi.org/10.1007/s11192-018-2735-5>
- [6] Conrad, S. 1997. Föderierte Datenbanksysteme: Konzepte der Datenintegration. Berlin und Heidelberg 1997.
- [7] English, L. P. 1999. Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. New York, NY, USA: John Wiley & Sons, Inc., 1999.
- [8] Fondermann, P. and Köppen, D. 2013. Zahlen, Daten, Fakten – ein Forschungsinformationssystem als Grundlage des Qualitätsmanagements für die Forschung am Karlsruher Institut für Technologie (KIT), DOI 10.1515/bfp-2013-0026 – Bibliothek, Forschung und Praxis 37(2): 172-181, Juli 2013.
- [9] Helmig, S. and Hollmann, R. 2009. Webbasierte Datenintegration – Ansätze zur Messung und Sicherung der Informationsqualität in heterogenen Datenbeständen unter Verwendung eines vollständig webbasierten Werkzeuges. Vieweg+Teubner / GWV Fachverlage GmbH, Wiesbaden, 2009.
- [10] Heinrich, B. and Klier, M. 2009. Die Messung der Datenqualität im Controlling – Ein metrikbasierter Ansatz und seine Anwendung im Kundenwertcontrolling. Controlling & Management, Augsburg, 2009.
- [11] Herwig, S. and Schlattmann, S. 2016. Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn, 2016.
- [12] IFQ. 2015. Spezifikation des Kerndatensatz-Forschung, Berlin, 2015. DOI: http://kerndatensatz-forschung.de/version1/Spezifikation_KDSF_v1.pdf
- [13] Krcmar, H. 2015. Informationsmanagement. Springer Gabler, 2015.
- [14] Lee, Y. M., Pipino, L.L., Funk, J.D. and Wang, R.Y. 2006. Journey to Data Quality, MIT Press, Cambridge, Massachusetts, 2006.
- [15] Naumann, F. and Leser, F. 2007. Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen. Dpunkt Verlag, 1. Edition, Oktober 2007.
- [16] Rahm, E. and Do, H.H. 2000. Data cleaning: Problems and current approaches. IEEE Bulletin of the Technical Committee on Data Engineering, 23(4), 2000.
- [17] Redman, T. 1996. Data Quality for the Information Age, Norwood 1996.
- [18] Wang, R. and Strong, D. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers, in: Journal of Management Information Systems, 12. Jg., 4/1996, S. 5-33.