

# IT-Integration mit semantischer Vernetzung für durchgängiges Digital Engineering

Joachim Caspar

CONWEAVER GmbH, Friedensplatz 12, 64283 Darmstadt,  
[joachim.caspar@conweaver.com](mailto:joachim.caspar@conweaver.com)

**Zusammenfassung:** Im Digital Engineering werden verschiedenste Tools zur Entwicklung komplexer technischer Produkte eingesetzt. Dabei entstehen Modelle, Spezifikationen, Dokumente entlang des gesamten Produktentwicklungsprozesses. Wie können die Tools und Entwicklungsartefakte durchgängig vernetzt werden (Traceability), um Experten bei fachübergreifenden Fragestellungen wie Change Impact Analysis zu unterstützen? Dazu beschreibt dieser Beitrag ein praxiserprobtes Verfahren zur IT-Integration, erklärt den Nutzen der Kombination von Information Retrieval und semantischen Netzen und erläutert anhand exemplarischer Artefakte zur Entwicklung eines elektrischen Bremssystems das Potenzial von Traceability-Systemen.

## 1 Einleitung

Die Entwicklung technischer Systeme wird immer komplexer. Um beispielsweise ein Bremssystem für PKW zu entwickeln, müssen viele fachübergreifende Aspekte und Wechselwirkungen zwischen einzelnen Teilsystemen berücksichtigt werden: Elektrik, Elektronik, Mechanik, Physik, Controller-Programmierung, etc.

Digital Engineering nutzt Modelle um frühzeitig testen, simulieren, ändern zu können. In der Praxis werden fachspezifische IT-Tools und Formate verwendet. Wenn nun ein Bauteil oder eine Baugruppe geändert werden soll, welche Auswirkungen hat das auf andere Teilsysteme? Welche Tests müssen durchgeführt werden, um beispielsweise die Erfüllung einer Anforderung zu überprüfen? Damit Experten bei der Beantwortung solcher Fragestellungen zumindest teilautomatisiert unterstützt werden können, müssen alle relevanten Dokumenteninhalte, digitale Modelle, Entwicklungsobjekte durchgängig vernetzt werden (Traceability). Wie könnte Traceability in der Praxis umgesetzt werden?

Verorten wir die Praxisfrage zuerst im zugrundeliegenden Themenfeld: *Verarbeitung, Prozesse, Management von Informationen und Daten*; und zwar aus Sicht des (Model Based) Systems Engineering.

## 2 Informations- und Datenmanagement

Die Begriffe *Informationen* und *Daten* meinen verschiedenes, werden aber oft synonym verwendet. Für unsere Zwecke reicht eine unscharfe Abgrenzung aus: Daten sind

digitale Repräsentationen von Informationen. Menschen interessieren sich unter anderem für Informationen. Computer arbeiten auf Daten, also auf Repräsentationen von Informationen. Der Mensch nutzt Computer als *Hilfsmittel* zur Informationsverarbeitung.

## 2.1 ...aus der Sicht des Systems Engineering

Im computerisierten Unternehmen sind Daten ein zentrales Asset. [Mi14] nutzt für sie sogar die starke Metapher des Blutes: „Data is the 'life blood' of an organization, for as it flows between systems, databases, processes, and departments, it carries with it the ability to make the organization smarter and more effective“. Schlüsseltechnologien liegen daher in den Bereichen der Verarbeitung, Integration und Analyse von Daten (bis hin zu Big Data Analytics).

Das Managen von Informationsflüssen ist die andere Seite der Medaille. Nach [Ad15, S. 461ff] kann ein Informationsmanagement (IM)-Prozess wie folgt charakterisiert werden: Informationen werden aus ein oder mehreren Quellen gesammelt, verarbeitet und zu ein oder mehreren Empfängern verbreitet. Sinn und Zweck ist dabei, Menschen beim Lösen von Problemen zu unterstützen. Der Mensch benötigt dazu relevante, zeitliche, vollständige und valide Informationen. Auch Informationssysteme, Architekturen, Services und Hardwareplattformen können zum IM-Prozess gezählt werden.

Mit dem Computer als Werkzeug zur Informationsverarbeitung ergibt sich dann in Summe: „A computer-based IM system is an organized combination of people, hardware, software, communication networks, and the data resources that collect, transform, and disseminate information in an organization.“ [ebd.]

## 2.2 ...aus der Sicht des Digital Engineering

Was wird entwickelt bzw. was ist das System-of-Interest? Das fragliche System ist ein materielles, technisches Zwischen- oder Endprodukt, im Gegensatz zu z. B. Dienstleistungen. In der Sprache des Systems Engineering wäre das Thema also *Product Systems Engineering*. Welche Informationen und Daten sind nun interessant?

Entlang des Engineering-Lebenszyklus entsteht eine ganze Reihe von Artefakten (Anforderungen, Spezifikationen, Konstruktionen, Testpläne etc.), die sich auf die in der Entwicklung befindlichen Teilsysteme beziehen. Je nach Lebenszyklus-Phase und je nach Detaillierungsgrad überschneiden sich dabei die Inhalte der fachspezifischen Repräsentationen mehr oder weniger.

Model Based Systems Engineering (MBSE) betont die Rolle von Modellen: „In a nutshell [...MBSE...] is about elevating models in the engineering process to a central and governing role in the specification, design, integration, validation, and operation of a system. For many organizations, this is a paradigm shift from traditional document-based [approaches] and acquisition lifecycle model approaches, many of which follow a 'pure' waterfall model of system definition, system design, and design qualification.“ [Es07]

Digitale Modelle sind für das Informationsmanagement auf zweifache Weise interessant. Erstens sind die Modelle digitale Repräsentationen (=Daten) von (Teil-)Systemen, bestehend aus *Artefakten* oder *Objekten*, die während der Entwicklung dieser Systeme entstehen. Diese Repräsentationen sind, im Gegensatz zum dokumentenbasierten Entwurf, direkt maschinenlesbar. Zweitens spezifizieren Modelle und Modellsprachen oft auch eine Art maschinenlesbare *Semantik*: die Typen der modellierbaren Objekte und die Typen der modellierbaren Verbindungen. Diese Zusatzinformationen erleichtern die digitale Vernetzung über einzelne Objekte hinaus.

### 2.3 Traceability-Systeme für das Digital Engineering

Alle Tools, Modelle, Formate etc. sowie alle am Entwicklungs- und Informationsprozess interessierten Parteien müssen im IM-Prozess berücksichtigt bzw. als holistisches IM-System betrachtet werden. Wie kann eine Computerunterstützung dazu aussehen?

Da der Modelleinsatz in der Praxis noch nicht 100% beträgt, müssen auch dokumentenbasierte Entwicklungsartefakte berücksichtigt werden. Gefordert ist also eine durchgängige Vernetzung von modell- und dokumentenbasierten Entwicklungsartefakten entlang des Produktentwicklungsprozesses, unter Einbezug der Experten und über domänenspezifische IT-Tools hinweg. Nennen wir ein System das diese Art von Traceability ermöglichen soll im Folgenden kurz *TR-System*.

Wie sollte bei der Planung und Umsetzung eines solchen TR-Systems vorgegangen werden und welche technischen Verfahren könnten innerhalb eines solchen Systems eingesetzt werden? Der Autor plädiert im Folgenden für eine iterative IT- und Datenintegration unter Zuhilfenahme von Verfahren aus den Bereichen Information Retrieval und (semantischen) Wissensnetzen. Ein TR-System entsteht dabei schrittweise durch planmäßige, unternehmens- und anwendungsfallspezifische IT-Integration.

## 3 Ein Vorgehensmodell zur IT-Integration

Für kleinere Firmen stellt sich die Frage zur IT-Integration weniger. Große Unternehmen hingegen haben über Jahrzehnte gewachsene, heterogene IT-Landschaften mit sehr vielen *Systembrüchen* bzw. *Datensilos*. Wie können die wichtigsten Systembrüche und mögliche Lösungen identifiziert werden? Wie können die Endanwender, die Experten möglichst früh einbezogen werden?

Bild 1 zeigt dazu unser Vorgehensmodell aus [Ca15], bestehend aus drei Phasen, die in jeweils drei Schritte unterteilt sind. Die meisten Schritte sehen einen Rücksprung vor, z. B. wenn sich bei der Anbindung einer bestimmten Datenquelle herausstellt, dass die bestehende Hardware für das TR-System nicht ausreicht und aufgestockt werden muss.

**Analyse der Datenlandschaft und Prozesse:** Ziel des ersten Schrittes ist das Verständnis der IT-Systeme, Datenarten und Zusammenhänge in den operationalen Prozessen. Der Kunde stellt verwendete Softwaresysteme vor, erläutert deren Bedeutung

für die einzelnen Fachabteilungen, veranschaulicht die intern bestehenden Informations- und Dokumentationsprozesse und stellt Auszüge relevanter Datenquellen zur Verfügung. Dies hilft bei der ersten Festlegung eines übergreifenden Vernetzungsmodells. Fehlende Verbindungen werden identifiziert und erste Möglichkeiten erarbeitet, wie die bestehenden Systembrüche überwunden werden können.

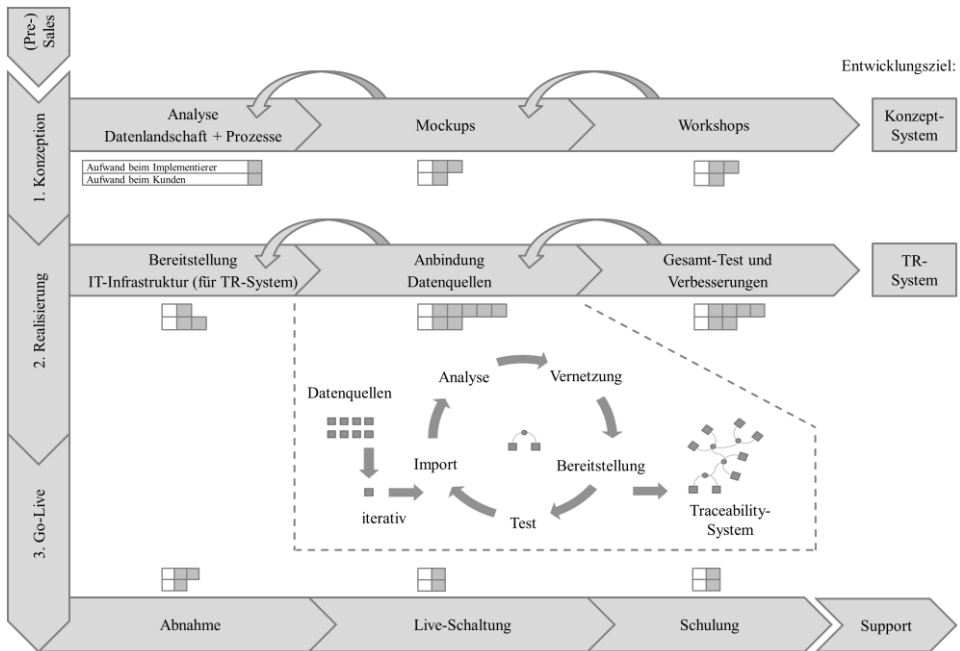


Bild 1: Vorgehensmodell zur IT-Integration

**Mockups:** Vorschläge für Benutzeroberflächen werden entworfen und helfen dabei, grundlegende Funktionalitäten und verschiedene Darstellungsarten zu diskutieren.

**Workshops:** Einer repräsentativen Gruppe von Endanwendern wird der jeweils aktuelle Stand des Konzeptsystems vorgestellt. Das Feedback der Anwender ist dann die Grundlage der nächsten Iteration. Der letzte Workshop ist erreicht, wenn das Konzeptsystem so weit ausgearbeitet ist, dass es durch die Entscheidungsträger abgenommen wird. Die Akzeptanz und Empfehlung durch die Endanwender ist dafür eine wichtige Voraussetzung.

**Bereitstellung von IT-Infrastruktur:** Der Kunde muss Rechner mit ausreichender Hardware bereitstellen, auf denen das geplante TR-System installiert, konfiguriert und in Betrieb genommen werden kann. Die Hardware ist abhängig von der Menge der zu verarbeitenden Daten und der Anzahl der geplanten (parallelen) Benutzeranfragen pro Sekunde. Auch müssen bei der Datenverarbeitung gesetzliche Vorschriften, Betriebsrats- und Datenschutzbelange berücksichtigt werden.

**Anbindung der Datenquellen:** Während für das Konzeptsystem die Datenlandschaft insgesamt analysiert wurde, werden nun die (Produktiv-)Datenquellen iterativ angebunden und in der Benutzeroberfläche bereitgestellt. Ausgehend von den ersten beiden Quellen und deren Verknüpfungen wird das TR-System somit schrittweise vergrößert.

**Gesamttest und Verbesserungen:** Bei der iterativen Datenanbindung wurden jeweils Teilaspekte getestet. Nun wird noch einmal das gesamte TR-System von den Implementierern, den Entscheidern und den repräsentativen Endanwendern „auf Herz und Nieren“ getestet, um letzte Fehler zu identifizieren und zu beseitigen und um letzte Verbesserungswünsche zu sammeln und umzusetzen.

Nach einer erfolgreichen **Abnahme** des TR-Systems durch den Kunden kann die Überführung in den Produktivbetrieb stattfinden. Der Kunde gibt dazu den Zugriff für alle Endbenutzer frei (Rechtmanagement und **Live-Schaltung**) und kommuniziert die Zugänglichkeit der Anwendung. Alle berechtigten Benutzer können das TR-System nun für die tägliche Arbeit verwenden. Falls erforderlich werden **Schulungen** durchgeführt.

#### 4 Traceability am Beispiel: elektrisch betätigte Radbremse für PKW

Welchen praktischen Mehrwert hat die durchgängige Vernetzung? Tabelle 1 zeigt einige exemplarische Artefakte und Verbindungen, die im Forschungsprojekt opTRAC entlang eines idealisierten MBSE-Prozesses erstellt wurden. Das System-of-Interest ist eine elektrisch betätigte Radbremse [Op15].

Tabelle 1: Einige Entwicklungsartefakte des Beispielsystems

| Klasse             | Instanz   | Kurzbeschreibung (Textfeld)   |
|--------------------|-----------|---|
| Produktanforderung | P-ANF-017 | Begrenzung des kritischen Raumbedarfs   |
|                    | P-ANF-027 | Montage-, Service- und Reparaturfreundlichkeit, wartungsfreie/arme Komponenten.               |
| Systemanforderung  | S-ANF-016 | Die Stromregelung sollte möglichst in direkter Nähe zum elektromechanischen Wandler erfolgen. |
| Systementwurf      | S-DES-300 | Energieversorgungseinrichtung   |
| Umsetzung          | D-DES-005 | Energieversorgungseinrichtung + Typ: Kaufteil   |
| Baugruppe          | BG-1-2    | Energieversorgungseinrichtung   |
| Bauteil            | BT-1-2-1  | Batterie  |
| Testfall           | TC-009    | Elektrischer Energiespeicher und Leistungselektronik  |
| Änderungsanfrage   | CRQ-001   | Kann statt CellD-2 [...kleinere Batterie...] genutzt werden?                                  |

Die Sammlung besteht aus 11 Klassen mit 147 Objekten bzw. Instanzen und aus 20 Verbindungstypen zwischen Klassen (z. B. Produktanforderung → Systemanforderung) mit insg. 192 einzelnen, uni-direktionalen Verbindungen. Jede Verbindung (A → B) impliziert dabei eine weitere, inverse Verbindung (B ← A).

Eine automatische Analyse lieferte weitere Verbindungen, beispielsweise durch triviale Namensgleichheit einiger Baugruppen mit den Kurzbeschreibungen anderer Instanzen (vgl. Tabelle 1). In der Praxis kommen bei Analysen viele nichttriviale, mögliche Verbindungen dazu, z. B. zwischen Objekten aus verschiedenen Tools mit verschiedenen Bezeichnungsschemata.

#### 4.1 Benutzeroberfläche

Bild 2 zeigt die Benutzeroberfläche des exemplarischen TR-Systems. Im oberen Teil befindet sich ein Feld zur Freitextsuche. Im mittleren Teil werden Suchergebnisse entweder als Übersicht aufgelistet oder einzelne Treffer in einer Detailansicht angezeigt.

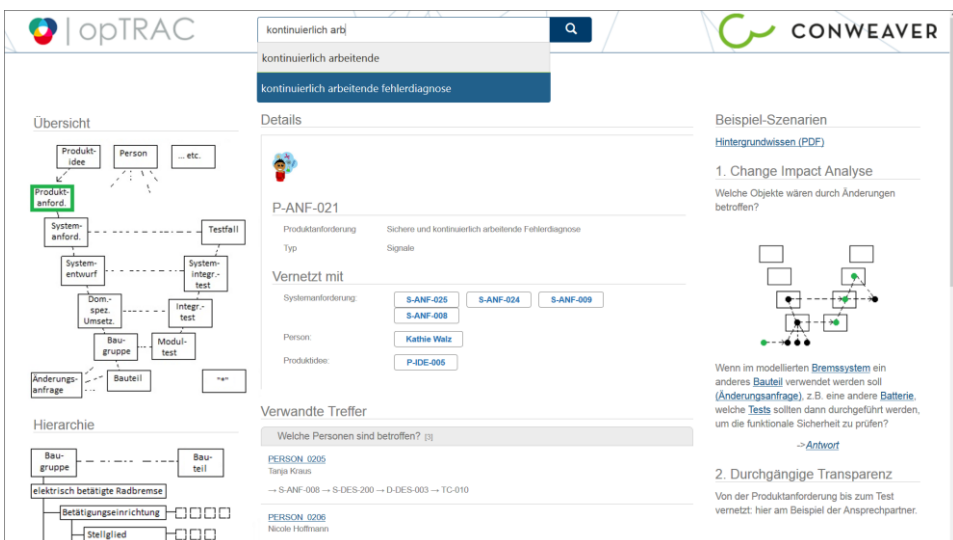


Bild 2: Screenshot einer Benutzeroberfläche

Ein Benutzer kann wie folgt recherchieren: Freitextsuche nach Nummern (S-ANF-021), nach Begriffen (Fehlerdiagnose), nach Objekttypen (Produktanforderung), oder mittels Übersicht oder Hierarchie (Abb. 2, linke Felder). Von der Detailansicht ausgehend kann er mittels Mausklick weiter navigieren zu verlinkten Dokumenten, benachbarten Objekten oder zu Ergebnissen von komplexen, konfigurierten Suchen (*Verwandte Treffer*). Letztere gehen von der aktuellen Instanz aus und folgen sukzessive mehreren Verbindungen: Änderungsanfrage P-ANF-021 → Systemanforderung(en) → domänenspezifische Umsetzung(en) → etc. (vgl. Bild 4).

#### 4.2 Anwendungsszenario: Anforderungs- und Testmanagement

Gegeben sei eine Produkt- oder Systemanforderung. Die Frage, welche Tests bzw. welche Mengen an Testergebnissen relevant zum Abgleich mit einer solchen

Anforderung wären, ist oft schwierig zu beantworten. Im bekannten V-Modell befinden sich Anforderungen links oben (top-down Entwicklung) und Tests verschiedener Granularität im rechten Ast (bottom-up Integration).

Ein TR-System kann hier über mehrere Schritte automatisiert die indirekten Verbindungen herausfinden helfen. Einerseits sind die Testpläne selbst als Ergebnis der Recherche wichtig. Andererseits muss auch immer der Mensch als Experte mitbedacht und einbezogen, vom TR-System als möglicher Ansprechpartner angezeigt werden.

Als eine Art *Globalanalyse* der Vernetzungsstruktur könnte auch eine automatisierte Berechnung von Überdeckungsgraden stattfinden: Welche Anforderungen sind durch welche Tests abgedeckt? Welche Anforderungen werden weder direkt noch indirekt durch Tests adressiert?

#### **4.3 Anwendungsszenario: Change Impact Analysis**

Potenzielle Auswirkungen von Änderungen an einem System-of-Interest zu überblicken wird mit steigender Systemgröße manuell unmöglich. Soll beispielsweise im elektrischen Bremssystem eine andere Energiequelle verwendet werden (vgl. Tabelle 1 für einige relevante Artefakte zu dieser Fragestellung), so wäre *ohne* ein TR-System aufwändiges Suchen in den domänenspezifischen Tools, in Datenbanken, in Dokumenten notwendig, um eventuell betroffene Komponenten, Anforderungen, Tests, oder auch Ansprechpartner herauszufinden. Ein TR-System spart an dieser Stelle enorme Zeitaufwände und hilft insbesondere in der Unterstützung fachübergreifender Fragestellungen.

### **5 Vorgehen bei der Datenanbindung**

Wie funktioniert nun real die Anbindung einzelner Datenquellen? Jede Quelle durchläuft dazu vier technische Schritte (vgl. Bild 1 Mitte und Bild 3 Rechts). Anschließend wird die Anbindung getestet und verbessert; je nach Komplexität der Inhalte wird ein Workshop beim Kunden angeboten. Dann wird die nächste Datenquelle angeschlossen.

Der bekannte Begriff *ETL* (Extract – Transform – Load) deckt sich ungefähr mit Schritt Nr. 1 (Import) und teilweise Schritt Nr. 2 (einfach Analysen und Filter). Hier kommen die semantische Vernetzung und die Konfiguration komplexer Suchstrategien hinzu.

Die CONWEAVER GmbH bildet die vier technischen Schritte in einer zentralen Workbench ab, in der die häufigsten Funktionen mittels „Drag & Drop“ zu Verarbeitungsketten für Datenflüsse zusammengesetzt und parametrisiert werden. Fehlt eine notwendige Funktion, so kann sie meist durch bestehende zusammengesetzt werden. Aktuell enthält das Framework über 400 solcher Basisfunktionen bzw. Module. Aufwändiges Programmieren ist somit selten bis gar nicht mehr notwendig.

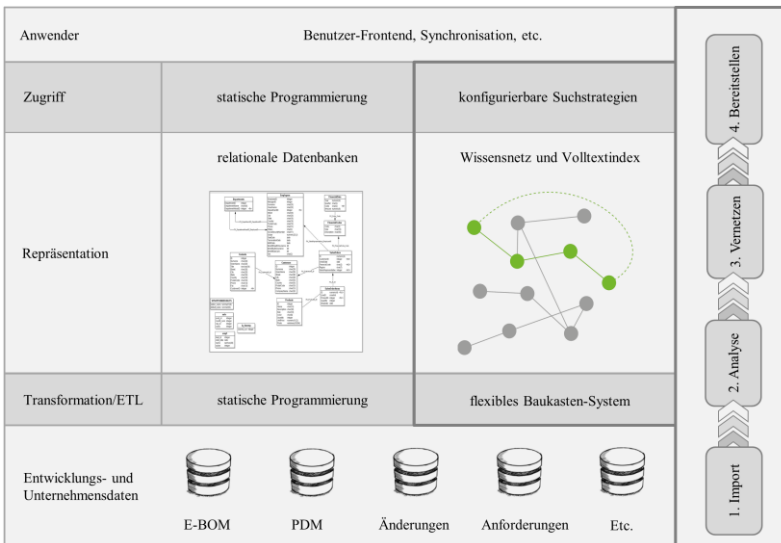


Bild 3: Unterschiede zwischen statischer Datenintegration und dem CONWEAVER-Ansatz

### 5.1 Import mit Konnektoren

Konnektoren setzen entweder auf Schnittstellen der relevanten Tools auf (Teamcenter, Sharepoint, Exchange-Server, SQL-Datenbanken, Doors, Windchill, SAP etc.), oder sind zum schnellen Auslesen optimiert, z. B. Crawler zum Auslesen von Metadaten und Inhalten eines Dateisystems.

### 5.2 Analyse der Rohdaten

Jede Datenquelle kann eine ganze Reihe von Informationsobjekten – mehr oder weniger strukturiert – enthalten. Eine Anbindung an einen LDAP-Server (Lightweight Directory Access Protocol) liefert z. B. Benutzernamen und Zugriffsrechte. Ein Fileshare hingegen kann alle möglichen Arten von Dateien enthalten, von denen die wenigsten relevant für die Umsetzung der Anwendungsfälle sind; Filter und (automatisierbare) Nacharbeiten sind gefragt. Befinden sich in den Dateien eingescannte Zeichnungen von Bauteilen, so kann eine OCR-Analyse (Optical Character Recognition) dabei helfen textuelle Informationsobjekte wie Bauteilnummern, technische Bezeichner, etc. zu erkennen. Metadaten einer Datei oder eines Datenbankeintrags liefern oft weitere Informationen, z. B. Aktualität (letzter Änderungszeitpunkt) und Ansprechpartner (letzter Bearbeiter).

### 5.3 Vernetzung der extrahierten Informationsobjekte

Alle textuellen Objekte können direkt in einem Volltextindex geladen werden und sind direkter Suche mit Verfahren des Information Retrieval zugänglich. Zusätzlich helfen



die Metadaten und die extrahierten, *feingranularen* Informationsobjekte bei dem Abgleich mit bestehenden Objekten aus anderen Systemen (=Duplikate) und der Abbildung verschiedener Schemata in ein *gemeinsames, systemübergreifendes Strukturmodell*.

Solch ein Strukturmodell kann vereinfacht als Graph verstanden werden: Knoten repräsentieren Informationsobjekte, Kanten stellen Verbindungen zwischen diesen Objekten dar: *Baugruppe* <wird-getestet-von> *Modultest*. Ein Strukturmodell definiert zuerst ein *Schema*, bestehend aus Objekttypen und Verbindungstypen. Die vielen tausenden Einzelobjekte und ihre Verbindungen werden als Instanzen bezeichnet. Repräsentationen solcher Art werden auch *Ontologie*, *semantisches Netz* oder *Wissensnetz* genannt.

### 5.4 Bereitstellung der Objekte durch fachspezifische Suchstrategien

Für das Konzept-System wurde mit dem Kunden eine grafische Benutzeroberfläche abgesprochen, die nun mit jeder weiteren angeschlossenen Datenquelle „mit Leben erfüllt“ wird. Textuelle Objekte werden über einen Volltextindex bereitgestellt. Feingranulare Informationsobjekte im Wissensnetz liefern zusätzliche, direkt verbundene Objekte.

Auf der Schemaebene eines Wissensnetzes werden die komplexen Suchen, die mehreren Pfaden folgen können, konfiguriert. Die Anzahl der Objekte bei der Ausführung einer solchen Suche kann dabei je Zwischenschritt größer und kleiner werden. Diese Suchkonfigurationen sind oft ähnlich zu den früheren, manuellen und zeitaufwendigen, sequentiellen Recherchestrategien der Endanwender auf ihren verschiedenen (nicht-integrierten) IT-Tools.

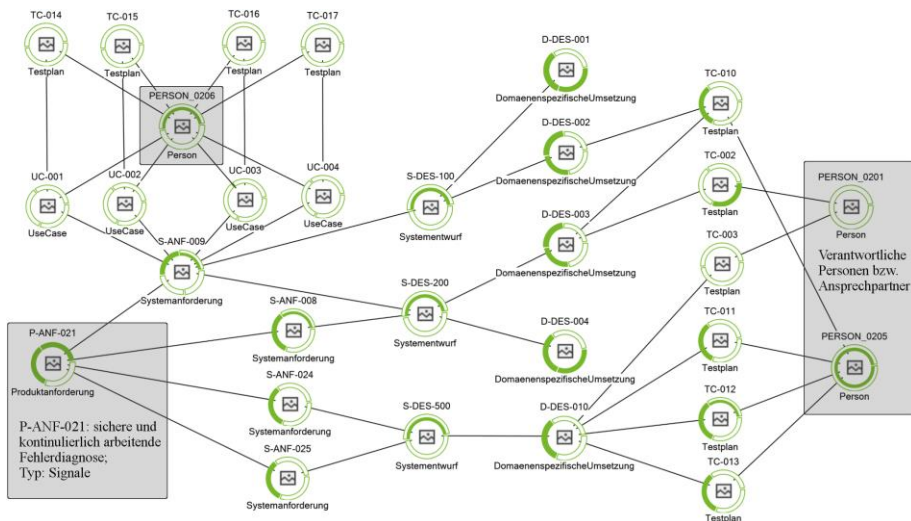


Bild 4: Beispielinstanzen zu einer komplexen Suchkonfiguration

Bild 4 zeigt alle Instanzen für eine solche Suche: Ausgehend von der Produkthanforderung P-ANF-021, werden über verschiedene Verbindungen verschiedene Arten von Tests gefunden, mit denen insgesamt drei Ansprechpartner verbunden sind.

Die gefundenen Tests und Ansprechpartner werden als *Verwandte Treffer* in der Benutzeroberfläche dargestellt (vgl. Bild 2) und sind ein Beispiel für die Unterstützung fachspezifischer, automatisierbarer Suchstrategien: Gegeben eine Produkthanforderung (hier: P-ANF-021), finde alle Tests, die relevant zum Abgleich mit dieser Anforderung sein könnten bzw. Ansprechpartner (Experten), die bei der Beantwortung solcher Fragestellungen weiterhelfen könnten.

## 6. Das wichtigste in Kürze

Gutes Informations- und Datenmanagement ist eine wichtige Voraussetzung für digitales Engineering. Durchgängige Vernetzung unterstützt Experten bei fachübergreifenden Fragestellungen. IT-Integration sollte dazu frühzeitig die Endanwender einbeziehen und, ausgehend von identifizierten Strukturbrüchen und relevanten Datenquellen, iterativ arbeiten, um ein erstes Konzeptsystem Schritt für Schritt als operatives TR-System an die Unternehmensrealität anzupassen.

Semantische Methoden nutzen Graph-ähnliche Repräsentation als gemeinsames Strukturmodell und ermöglichen die Automatisierung komplexer Suchstrategien. Die CONWEAVER GmbH entwickelt kundenspezifische TR-Systeme und verwendet dazu aktuell über 400 Module (für Konnektoren, zur Analyse, zur Vernetzung, zur Konfiguration), um Programmieraufwände zu minimieren.

## Literaturverzeichnis

- [Ad15] BKCASE Editorial Board. 2015. The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 1.3.2. R.D. Adcock (Editor-in-Chief). Hoboken, NJ: The Trustees of the Stevens Institute of Technology. [www.sebokwiki.org](http://www.sebokwiki.org), abgerufen am: 25.06.2015.
- [Ca15] Caspar, J. et al.: Traceability: vom Konzept zur Umsetzung. Ein Leitfaden für die IT-Integration, 2015. [www.optrac.de](http://www.optrac.de) → Downloads, zuletzt abgerufen am: 25.06.2015.
- [Es07] Estefan, Jeff A.: Survey of model-based systems engineering (MBSE) methodologies. IncoSE MBSE Focus Group 25 (2007): 8.
- [Mi14] MITRE Systems Engineering Guide – Collected wisdom from MITRE’s systems engineering experts, 2014. [www.mitre.org/publications/systems-engineering-guide/](http://www.mitre.org/publications/systems-engineering-guide/), abgerufen am: 11.06.2014.
- [Op15] Ein Bericht zu dem im Forschungsprojekt opTRAC entwickelten mechatronischen Beispielsystem wird demnächst online verfügbar sein auf: [www.optrac.de](http://www.optrac.de) → Downloads.