



KAP DION

## Kap Dion SDI 1.0 Produktbeschreibung

Der Inhalt des vorliegenden Dokuments beschreibt den Funktionsumfang von Kap Dion SDI in der Version 1.0. Kap Dion SDI (Stammdatenintegration) ermöglicht den Import, die Konvertierung und den Abgleich von Stammdaten. Ziel der Applikation ist die Optimierung der Datenqualität in den stammdatenführenden Anwendungen sowie die Verbesserung des stammdatenabhängigen Reportings und die Unterstützung der applikationsübergreifenden Kommunikation.

**Ansprechpartner:**

Carsten Schmitz

Geschäftsführer Kap Dion GmbH

Tel. +49 (152) 279 882 01

Tel. +43 (664) 734 800 24



## Inhalt

Inhalt .....	2
1. Einleitung .....	3
2. Übersicht .....	3
3. Funktionen im Detail .....	3
1.1. Definition des Stammdatenmodells .....	4
1.2. Stammdatenimport .....	4
1.3. Automatischer Stammdatenvergleich .....	4
1.4. Stammdatenzwillinge .....	5
1.5. Manueller Stammdatenabgleich .....	6
1.6. Auswahl eines führenden Objekts .....	6
1.7. Doublettenprüfung .....	7
1.8. Integration in Drittsysteme .....	7
1.1.1. Integration von Anwendungssystemen .....	7
1.1.2. Integration von Reportingsystemen .....	8
1.1.3. Integration von Kommunikations- bzw. Messagingsystemen .....	8
1.9. Weitere Funktionen .....	8
1.1.4. Suche .....	8
1.1.5. Benutzerverwaltung .....	9
1.1.6. Workflow .....	9
4. Voraussetzungen .....	9
1.10. Verfügbarkeit .....	9
1.11. Technische Voraussetzungen .....	9
1.12. Mengengerüst .....	9
1.13. Verwendete Software anderer Anbieter .....	9
5. Abbildungsverzeichnis .....	10



## 1. Einleitung

Das vorliegende Dokument bietet einen vollständigen Überblick über die in Kap Dion SDI 1.0 enthaltenen Funktionen, die notwendigen Voraussetzungen für den Einsatz von SDI sowie mögliche Beschränkungen in der Verwendung.

## 2. Übersicht

Mit Kap Dion SDI können Stammdaten effizient aus verschiedenen Datenquellen abgeglichen und in der weiteren Folge bereinigt werden. In einem ersten Schritt wird ein einheitliches Stammdatenmodell für einen Stammdatentyp definiert (1.1). Danach erfolgt der Import der Stammdaten aus einer oder mehreren Datenquellen (1.2). Nach erfolgtem Import und der möglichen Transformation der Stammdaten in ein definiertes Zielformat stehen die Stammdaten für einen automatisierten Datenvergleich zur Verfügung (1.3). Als Ergebnis des Stammdatenvergleichs werden sogenannte Stammdatenzwillinge gebildet (1.4) bzw. in einem weiteren Schritt ein manueller Stammdatenabgleich angestoßen (1.5). Die so gewonnenen Informationen können im weiteren Verlauf den Herkunftssystemen zur Bereinigung der Stammdaten (1.6), Reportingsystemen zu verbesserten Berichten oder Middleware-Systemen zur Systemkommunikation zur Verfügung gestellt werden (1.8).

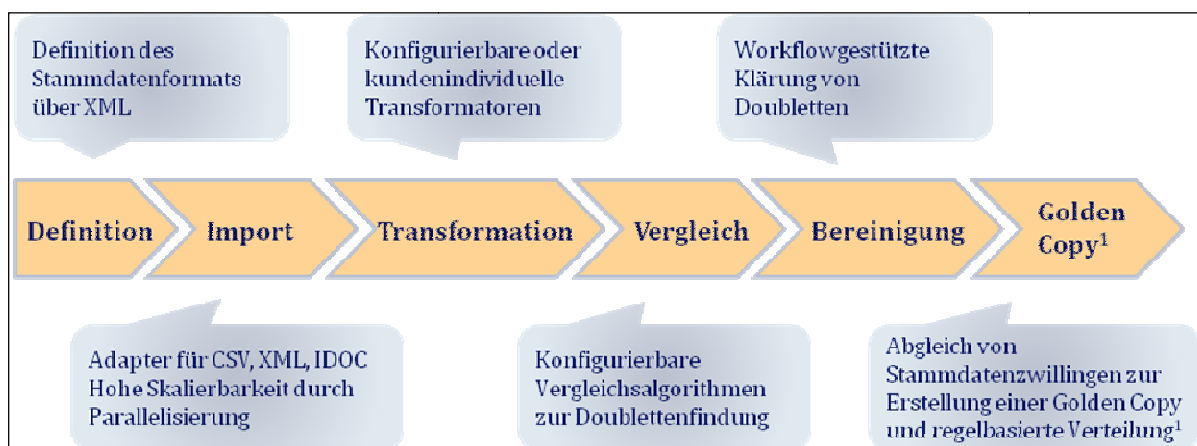


Abbildung 1: Prozessübersicht SDI<sup>1</sup>

Darüber hinaus kann Kap Dion SDI verwendet werden, um zukünftig das Anlegen redundanter Stammdaten zu verhindern (1.7).

## 3. Funktionen im Detail

Im folgenden Kapitel werden die einzelnen Funktionen und Prozessschritte von Kap Dion SDI 1.0 beschrieben.

<sup>1</sup> Erstellung einer Golden Copy und regelbasierte Verteilung geplant für SDI 2.0



## 1.1. Definition des Stammdatenmodells

Um Stammdaten in Kap Dion SDI bearbeiten zu können, ist zunächst die Definition eines einheitlichen Stammdatenmodells für jeden Stammdatentyp erforderlich. Dieses Modell dient einerseits zur Vergleichbarkeit von Stammdaten aus verschiedenen Herkunftssystemen, andererseits aber auch zur Unterstützung der Stammdatensuche (1.1.4).

## 1.2. Stammdatenimport

Stammdaten liegen in der Regel in verschiedenen Anwendungssystemen vor. Häufig laufen in regelmäßigen Abständen Programme ab, die die verschiedenen Stammdatenablagen in den unterschiedlichen Anwendungen miteinander synchronisieren. Viele Anwendungen besitzen daher schon die Möglichkeit Stammdaten in verschiedensten Formaten zu exportieren.

Kap Dion SDI verfügt über ein flexibles Adapterkonzept, um pro Anwendungssystem und Stammdatentyp einen spezifischen Stammdatenimport durchzuführen. Folgende Importmöglichkeiten sind in Kap Dion SDI verfügbar:

- File-Upload via CSV-File
- File-Upload via XML-File
- SAP IDOC<sup>2</sup>

Darüber hinaus können weitere Adapter auf Projektbasis zur Verfügung gestellt werden.

Beim Stammdatenimport wird eine Konvertierung der importierten Daten auf das für den Stammdatentyp definierte einheitliche Stammdatenmodell durchgeführt.

Anschließend stehen die Stammdaten für die weitere Bearbeitung im System zur Verfügung.

## 1.3. Automatischer Stammdatenvergleich

Ziel des Stammdatenvergleichs ist es, für einen Stammdatensatz mögliche Doubletten zu ermitteln, um diese Daten dann in der weiteren Folge bereinigen zu können.

Dazu werden zunächst für einen Stammdatensatz ein oder mehrere Indizes definiert, auf deren Basis dann mögliche Doubletten ermittelt werden können. Die Definition eines geeigneten Index hängt dabei von den Merkmalen der importierten Stammdaten ab. Es können verschiedenen Indizes miteinander kombiniert werden, wobei sowohl exakte Indizes (z.B. Personalausweisnummer) als auch unscharfe Indizes (z.B. Trigramme) verwendet werden können.

Als Ergebnis des Indexvergleichs wird für einen Stammdatensatz eine Menge möglicher Doubletten ermittelt.

In einem zweiten Schritt wird eine Übereinstimmungswahrscheinlichkeit zwischen den möglichen Doubletten und dem ursprünglichen Stammdatensatz ermittelt. Die Berechnungsmethodik kann dabei wieder frei konfiguriert werden und z.B. weitere Attribute in die Kalkulation beziehen.

---

<sup>2</sup> Verfügbar mit Kap Dion SDI 1.1



KAP DION SDI StammDatenIntegration

**Lieferanten**

**Kunden**

Doubletten  
Konfiguration  
Adapter

**Materialien**

**Warengruppen**

**Administration**

Status	Vergleich	Normalisierung	Loeschen	Ersetzen	Symbol Ersetzen	Formatieren
Id	Feld	Gewicht	Algorithmus	Parameters		
101	Name	20	StringComparisonDLAlgorithm			
102	Strasse	10	StringComparisonDLAlgorithm			
103	Stadt	10	StringComparisonDLAlgorithm			
104	PLZ	5	StringComparisonDLAlgorithm			
105	Land	10	StringComparisonDLAlgorithm			

Seite 1 von 3 Datensatz 1-5 von 11

ZURÜCK WEITER

**Feld Name** ▶ Speichern

Gewicht  
20

Algorithmus  
StringComparisonDLAlgorithm

Abbildung 2: Konfiguration des Vergleichsalgorithmus

Folgende Ergebnisse sind dabei möglich:

- Die Stammdaten stimmen überein. Die beiden Datensätze werden zu Stammdatenzwillingen (1.4) zusammengefasst.
- Die Stammdaten stimmen nicht überein.
- Der Vergleich kann nicht automatisch entschieden werden. Für einen manuellen Vergleich wird eine Aufgabe erzeugt (1.1.6), um die beiden Datensätze auf Gleichheit von einem Mitarbeiter prüfen zu lassen (1.5).

## 1.4. Stammdatenzwillinge

Eine Stammdatengruppe besteht aus Stammdatensätze, die das gleiche Stammdatenobjekt beschreiben.

### Beispiel

Herr Müller hat ein Konto bei einer Bank. In einer anderen Filiale der gleichen Bank eröffnet er ein weiteres Konto und wird als Kunde ein weiteres Mal mit seinen persönlichen Daten gepflegt.

Im Ergebnis wird Herr Müller doppelt als Stammdatensatz bei der Bank geführt, obwohl ihm eigentlich nur das neue Konto hätte zugewiesen werden müssen.

Die Stammdatenzwillinge bestehen in diesem Fall aus den beiden Stammdatensätzen für Herr Müller.

Stammdatenzwillinge können aus Stammdatensätzen des gleichen Herkunftssystems aber auch aus Stammdatensätzen unterschiedlicher Herkunftssysteme bestehen.

Bei Stammdatenzwillingen aus dem gleichen Herkunftssystem ist in der Regel die Auswahl eines führenden Objekts und Eliminierung der übrigen Zwillinge das Ziel ist.

Bei Stammdatenzwillingen aus unterschiedlichen Herkunftssystemen reicht häufig schon das Wissen um die Zuordnung, um z.B. Reportinganforderungen zu erfüllen.



## 1.5. Manueller Stammdatenabgleich

Wenn ein Stammdatenvergleich nicht automatisch entschieden werden kann, bietet SDI die Möglichkeit, mithilfe einer effizienten Benutzeroberfläche Stammdatenvergleiche manuell zu entscheiden. Dazu werden dem Benutzer diejenigen Stammdatenvergleiche zur Entscheidung vorgelegt, die im automatischen Stammdatenvergleich nicht entschieden werden konnten und eine entsprechende Aufgabe angelegt wurde.

The screenshot shows the KAP DION SDI interface. The top header includes 'KAP DION SDI' and 'StammDatenIntegration'. A left sidebar contains navigation options: Lieferanten, Kunden, Doubletten, Konfiguration, Adapter, Materialien, Warengruppen, and Administration. The main area displays a table of comparison tasks:

ID	Beschreibung	Status
10264	Bereinigung Stammdatengruppe 10263	offen
10094	Bereinigung Stammdatengruppe 10093	offen
10086	Bereinigung Stammdatengruppe 10085	offen
10082	Vergleich Kunde 1013 (CRM) und Kunde 1018 (CRM)	erledigt
10081	Vergleich Kunde 1012 (CRM) und Kunde 1018 (CRM)	offen

Below the table, there is a detailed comparison view for two customer records. The left record is for customer 1012 and the right for customer 1018. Both have CRM as the origin system and DUNS number 321. The comparison highlights differences in the 'Doublettengruppe' field: 'Erster Schritt' for 1012 and 'First Step' for 1018. A green button labeled '10263' is visible in the comparison view.

Abbildung 3: Vergleich zweier Kundenstammdatensätze

Folgende Möglichkeiten hat der Benutzer zur Entscheidung:

- Die Stammdaten stimmen überein. Die beiden Datensätze werden zu Stammdatenzwillingen (1.4) zusammengefasst.
- Die Stammdaten stimmen nicht überein

Nachdem die Entscheidung vom Benutzer getroffen wurde, ist die entsprechende Aufgabe erledigt.

## 1.6. Auswahl eines führenden Objekts

Sind Stammdatenzwillinge gebildet, so wird eine weitere Aufgabe zur Auswahl eines führenden Objekts angelegt, sobald zwei Stammdatenzwillinge aus dem gleichen Herkunftssystem stammen. Der Benutzer hat dann die Möglichkeit, einen der Stammdatenzwillinge als führendes Objekt auszuwählen. Als Folge dieser Auswahl erhalten die weiteren Stammdatenzwillinge aus diesem Herkunftssystem den Status ‚Inaktiv‘.



ID	Beschreibung	Status
10080	Vergleich Kunde 1009 (CRM) und Kunde 1011 (CRM)	offen
10077	Bereinigung Stammdatengruppe 10076	offen
10071	Bereinigung Stammdatengruppe 10070	offen
10065	Bereinigung Stammdatengruppe 10064	offen

Bitte wählen sie ein führendes Objekt aus	
Strasse: burggasse 105/4/2	
PLZ: A-3100	
Stadt: St. Pölten	
Land: Österreich	
TELEFON, FAX UND E-MAIL	
Telefon	01 103-37-500
Fax	01 103-37-505
e-Mail	g-prag@www.com

Abbildung 4: Auswahl eines führenden Objekts

## 1.7. Doublettenprüfung

Nachdem Doubletten in einem Anwendungssystem mit der in 1.6 beschriebenen Funktionalität eliminiert sind, unterstützt SDI das System bei der Vermeidung neuer Doubletten. SDI bietet eine Schnittstelle an, die auf Basis eines neu eingegebenen Stammdatensatzes auf etwaige Doubletten prüft, bevor der neue Stammdatensatz gesichert wird. Dazu wird der Stammdatensatz über eine geeignete Schnittstelle an SDI übertragen und dort mit den bereits in 1.3 beschriebenen Mitteln geprüft, ob der Stammdatensatz möglicherweise bereits existiert.

Auf dieser Basis kann dann der Benutzer entscheiden, ob er den neuen Stammdatensatz tatsächlich neu anlegen muss.

## 1.8. Integration in Drittsysteme

Kap Dion SDI bietet die Möglichkeit verschiedene Systeme in die Anwendung zu integrieren. Dazu zählen Anwendungssysteme, Reportingsysteme und Kommunikations- bzw. Messagingsysteme.

### 1.1.1. Integration von Anwendungssystemen

Es gibt drei Integrationspunkte von Anwendungssystemen:

- den Datenimport (1.2) zum Upload von Daten in SDI
- die Auswahl eines führenden Objekts (1.6)



- die Doublettenprüfung (1.7)

Bei der Auswahl eines führenden Objekts können wie beim Import variabel Adapter konfiguriert werden, die das Herkunftssystem über die Auswahl eines führenden Objekts informieren. Zum Beispiel können alle Stammdatenzwillinge außer dem führenden Objekt mit einem Sperrkennzeichen versehen werden.

Auch bei der Doublettenprüfung können verschiedene Adapter konfiguriert werden, die das Herkunftssystem über mögliche Doubletten informiert.

### 1.1.2. Integration von Reportingsystemen

Reportingsysteme sind natürliche Abnehmer der in SDI ermittelten Stammdateninformationen. So können auf Basis der Stammdatenzwillinge konsolidierte Berichte über mehrere Anwendungssysteme erstellt werden, z.B.:

- Gesamtumsatz pro Kunde
- Einkaufsgesamtsumme pro Lieferant
- Gesamtumsatz pro Produkt

Da die Stammdateninformationen in SDI in relationalen Datenbanktabellen vorliegen, können diese Informationen in der Regel über die Standardfunktionalität des Reportingsystems importiert werden.

### 1.1.3. Integration von Kommunikations- bzw. Messagingsystemen

Auch bei der applikationsübergreifenden Kommunikation zwischen Systemen (z.B. SOA-Infrastrukturen) spielen Stammdatenzwillinge eine wichtige Rolle, um den Nachrichtenaustausch zu ermöglichen. In der Regel erwartet das Empfängersystem eine Verwendung der lokalen Schlüssel (z. B. Kundennummer, Lieferantenummer, Produktnummer), um eine System-System-Kommunikation zu unterstützen. Dem Sendersystem liegt diese Information in der Regel nicht vor.

Kap Dion SDI bietet für diesen Zweck Schnittstellen an, die auf Basis der gewonnenen Informationen über Stammdatenzwillinge die entsprechende Mapping-Information zwischen den lokalen Schlüsseln bereitstellen können.

## 1.9. Weitere Funktionen

Zur Verwendung von SDI gibt es einige unterstützende Funktionalitäten, die Teil des Lieferumfangs von Kap Dion SDI sind.

### 1.1.4. Suche

Um Stammdatensätze in SDI effizient suchen zu können, wird eine textbasierte, konfigurierbare Suchmaschine verwendet (1.13). Dadurch ist die Suche nach Stammdatensätzen sowohl über Wildcards als auch über eindeutige Schlüssel möglich. Wird SDI in einem Cluster betrieben, ist eine JMS-Infrastruktur (Java Messaging) zur Synchronisation der Suchindizes erforderlich.



### 1.1.5. Benutzerverwaltung

SDI verfügt über eine eigene Benutzerverwaltung, über die Benutzer angelegt und gepflegt werden können. Dabei können Benutzern Stammdatenobjekten zur Bereinigung zugeordnet werden. Externe Userverwaltungen (z.B. LDAP) können ebenfalls integriert werden (auf Projektbasis).

### 1.1.6. Workflow

SDI verfügt über einen einfachen einstufigen, konfigurierbaren Workflow, der die Aufgaben der Bereinigung von Stammdatenzwillingen einzelnen Benutzern nach vordefinierten Kriterien zuordnen kann.

## 4. Voraussetzungen

Vor dem Einsatz von Kap Dion SDI ist die Prüfung aller Voraussetzungen gemeinsam mit der Kap Dion GmbH erforderlich. Ohne Lizenzierung durch Kap Dion GmbH darf Kap Dion SDI nicht verwendet werden.

### 1.10. Verfügbarkeit

Kap Dion SDI 1.0 ist ab 02/2010 verfügbar.

Kap Dion SDI 1.1 wird voraussichtlich ab 05/2010 verfügbar sein.

### 1.11. Technische Voraussetzungen

Kap Dion SDI ist eine Java 1.5-Applikation die zum Betrieb folgende Komponenten benötigt:

- Java 1.5 kompatibles Betriebssystem
- Java-Applikationsserver  
z.B. JBoss > 4.2, Tomcat > 6.0, IBM Websphere > 6.1
- Datenbank mit zugehörigem JDBC-Treiber  
z.B. IBM DB2 > 7.1, Oracle > 8.i, MSSQL > 2000, MySQL > 3.23
- Browser  
z.B. Microsoft Internet Explorer > 6.0, Firefox >3.0

Für die Benutzung der Applikation über einen Browser muss Javascript aktiviert sein.

### 1.12. Mengengerüst

Kap Dion SDI ist mit Datenmengen von 500.000 Stammdatensätzen getestet. Vor der Verwendung größerer Datenmengen ist Rücksprache mit Kap Dion GmbH zu halten.

### 1.13. Verwendete Software anderer Anbieter

Kap Dion SDI verwendet folgende frei verfügbare Softwarepakete:

- Hibernate 3.2.6 (<https://www.hibernate.org>)
- Hibernate Search 3.1.0 (<https://www.hibernate.org/410.html>)
- Spring 2.5.1 (<http://www.springsource.org/about>)



- Google Webtoolkit 1.5.3 (<http://code.google.com/intl/de-DE/webtoolkit/overview.html>)
- GWT Validation 1.0 (<http://code.google.com/p/gwt-validation/>)
- Log4J 1.2.12 (<http://logging.apache.org/log4j/1.2/index.html>)

Für die Funktionsfähigkeit der Software anderer Anbieter übernimmt die Kap Dion GmbH keine Gewähr.

## 5. Abbildungsverzeichnis

Abbildung 1: Prozessübersicht SDI .....	3
Abbildung 2: Konfiguration des Vergleichsalgorithmus .....	5
Abbildung 3: Vergleich zweier Kundenstammdatensätze .....	6
Abbildung 4: Auswahl eines führenden Objekts.....	7