

Erstellung von Word und Excel Berichten mit Eclipse BIRT

Henning von Bargaen
Triestram & Partner GmbH
Bochum

Schlüsselworte

Reports, Berichte, Eclipse BIRT, Word, Excel

Einleitung

Eclipse BIRT hat sich als eine gute Alternative zu Oracle Reports oder dem Oracle BI Publisher herausgestellt (siehe Vortrag DOAG Anwenderkonferenz und Ausstellung 2012).

Wenn Berichte passgenau gedruckt oder per Email versendet werden sollen, ist die Ausgabe im PDF Format ideal. Es gibt aber auch Fälle, in denen die Ausgabe entweder manuell oder automatisch weiter bearbeitet werden soll.

Für solche Fälle unterstützt BIRT die Ausgabe von Berichten im Microsoft Word und Microsoft Excel Format, außerdem auch OpenOffice Formate – und zwar deutlich besser als bei Oracle Reports.

Der Vortrag geht auf die Besonderheiten bei solchen Word- und Excel-Berichten ein. Nach Möglichkeit wird eine kleine Live-Demo gezeigt.

Funktionsweise von BIRT

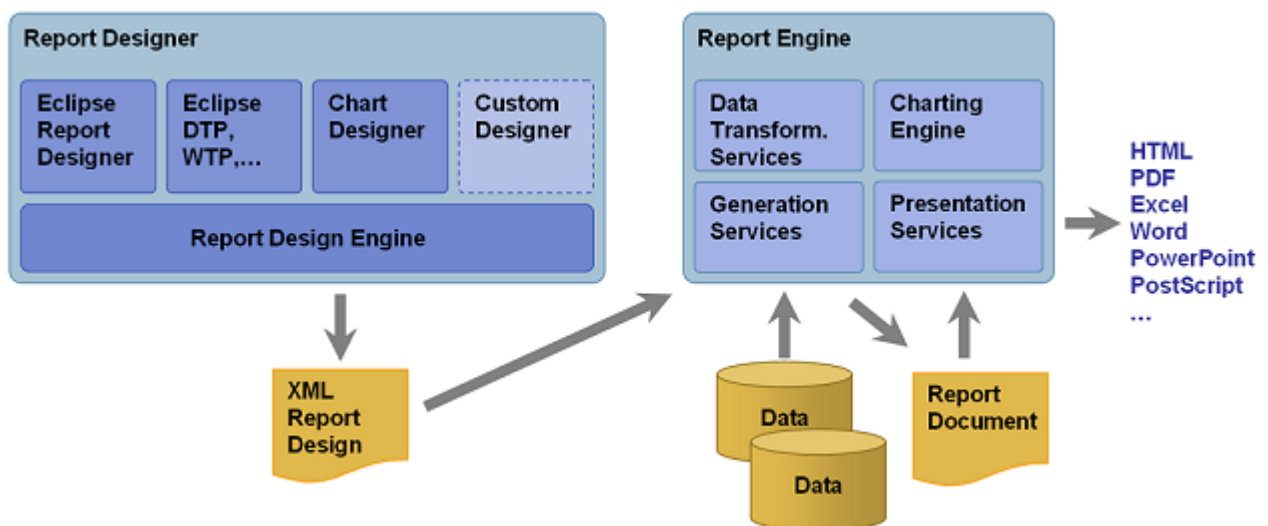


Abb. 1: Schema BIRT Funktionsweise

Es gibt neben Report Design Modulen (*.rptdesign) noch andere Module: Report Libraries (*.rptlibrary) und Javascript-Libraries (*.js). Auch Java-Klassen können verwendet werden.

Das Report Document beschreibt die Layoutstruktur des fertigen Berichts.

Die fertige Ausgabe wird von einem *Emitter* aus dem Report Document erzeugt (z.B. WordEmitter).

BIRT bietet zwei APIs, um Berichte zu erzeugen:

RunAndRenderTask: Erzeugt mit einem Emitter „direkt“ die fertige Ausgabe (das Report Document wird nur im Speicher erzeugt).

RunTask und *RenderTask*: Die *RunTask* erzeugt das Report Document als Datei (*.rptdocument). Die *RenderTask* erzeugt daraus die fertige. Das kann dann interessant sein, wenn die DB-Aufrufe relativ lange dauern und gleichzeitig verschiedene Ausgabeformate benötigt werden, weil dann die DB-Aufrufe nur einmal erfolgen müssen.

Ausgabe im Microsoft Word Format

Im Gegensatz zu Oracle Reports (lauter einzelne Positionrahmen) erzeugt der BIRT Word Emitter Dateien, die (nach Ausführung eines kleinen Makros) bearbeitet werden können wie jedes handgeschriebene Word-Dokument, z.B. können Absätze eingefügt oder gelöscht werden und die Seitenumbrüche können angepasst werden.

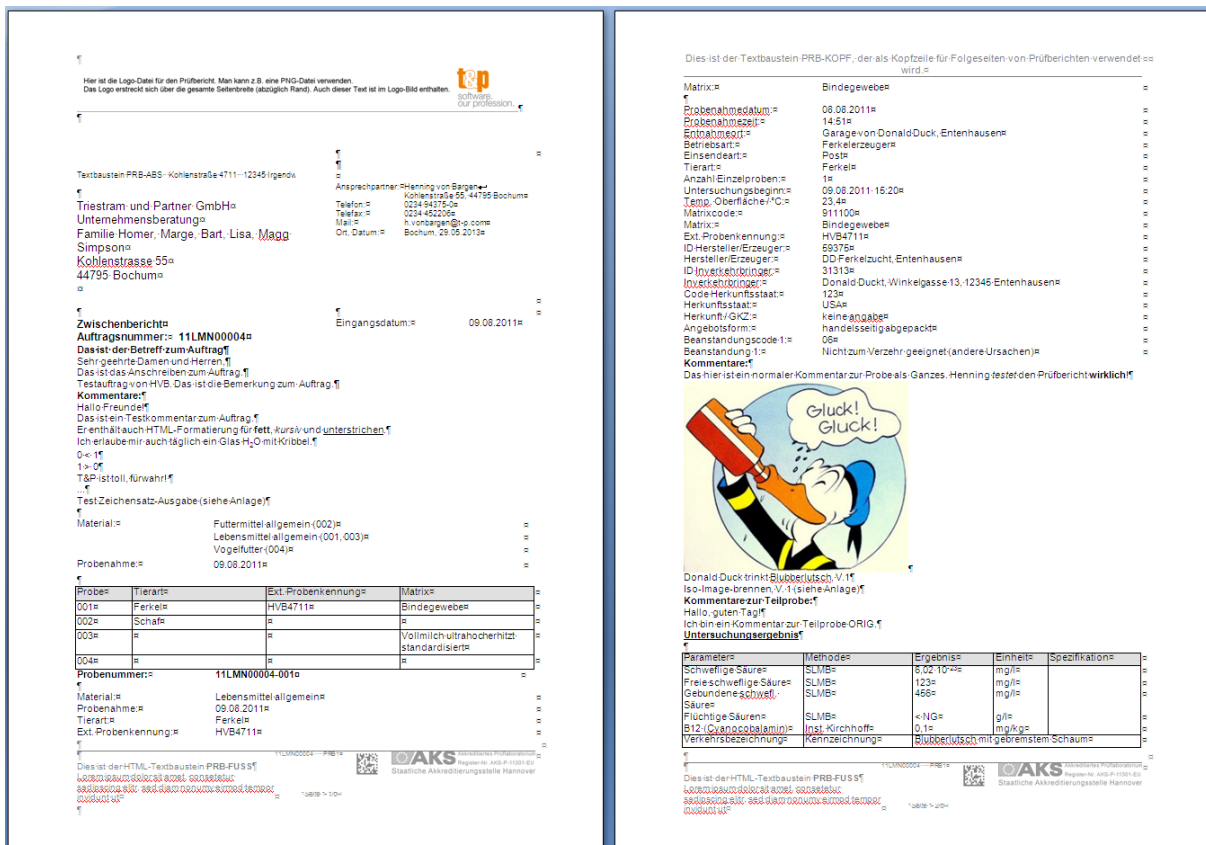


Abb. 2: Ausgabe im Word-Format (hier absichtlich mit Anzeige von Steuerzeichen)

Mit BIRT können verschiedene Word-Formate erzeugt werden:

Office 2003 XML oder Office 2010 DOCX ab BIRT Version 3.4.0 (voraussichtlich Ende Juni 2013)

Das binäre Format von Office 97 kann nicht erzeugt werden.

Schwierigkeiten bei der Word-Ausgabe von BIRT

Wie bereits angedeutet sollte erst ein kleines Word-Makro laufen, um das von BIRT generierte Dokument für die weitere Verarbeitung aufzubereiten, denn die rohe Word-Ausgabe von BIRT ist dafür noch nicht geeignet. Das liegt hauptsächlich an den nachfolgend beschriebenen Eigenheiten:

Scheinbar überflüssige Absätze: Siehe Bug 224261 – „BIRT empty table bands produce white space in output“. Der Word Emitter erzeugt für jede Tabellenzeile (genauer: Tabellen-Zelle oder Listen-Zeile) einen Absatz, auch wenn die Zeile keinen Inhalt hat. Das ist auch richtig so, führt aber natürlich dazu, dass eine leere Zeile entsteht. Beim PDF Emitter hat eine leere Zelle eine Höhe von 1px und wird daher in der Ausgabe normalerweise gar nicht bemerkt. Bei richtiger Herangehensweise schon während der Entwicklung des Berichts lässt sich das vermeiden (siehe Best Practices unten).

Überflüssige Tabellen: Der Word Emitter generiert offenbar für jede Zelle einer Tabelle implizit eine geschachtelte Word-Tabelle. Das ist anscheinend notwendig für die Unterstützung von Margin, Border, Padding usw. Es macht aber eine spätere Nachbearbeitung des Word-Dokuments unnötig schwierig. Der große Vorteil von Word-Dokumenten im Vergleich zur PDF-Ausgabe ist ja gerade, dass der Anwender später noch Texte ändern, einfügen und löschen kann und dass er die Seitenumbrüche noch optimieren kann. Das Einfügen von festen Seitenumbrüchen wird durch die vielen geschachtelten Tabellen schwierig.

Unterer Rahmen von Tabellen fehlt: Manchmal fehlt bei Tabellen die einen Rahmen enthalten, der untere Rahmen bei der letzten Tabellenzeile.

Word-Makro zur Aufbereitung des Word-Dokuments

In unserem Produkt wird nach dem Öffnen des Word-Dokuments auf dem Client anschließend ein Word-Makro gestartet, um die geschachtelten Tabellen aufzulösen und die Rahmenprobleme zu beseitigen, so dass dem Anwender anschließend ein „wie von Hand erstelltes“ Dokument zur Verfügung steht.

Das Makro wandelt zunächst im gesamten Dokument alle Tabellen, die nur aus einer Spalte bestehen, mit dem Word-Basic-Befehl `ConvertToText` in normalen Text um.

Anschließend prüft es die Rahmenlinien aller „echten“ Tabellen (3 oder mehr Spalten) und korrigiert sie.

Leider ist es nicht ohne weiteres möglich, im Word-Dokument den einzelnen Layout-Objekten Namen in Form von Textmarken zu übergeben (nur ansatzweise doch über TOC Entries).

Excel-Berichte mit dem Spudsoft Excel Emitter

Wir verwenden den Spudsoft Excel Emitter, ein kostenloses Plugin für BIRT. Er erzeugt Office 2003 XLS oder XLSX. In der kommenden Version BIRT 4.3.0 ist geplant, dass der Spudsoft Emitter direkt in BIRT integriert wird. Es gibt auch noch andere Excel Emitter, der Spudsoft Emitter funktioniert aber unserer Meinung nach am besten.

Der Spudsoft Excel Emitter erzeugt keine überflüssigen schmalen Zeilen und Spalten, da er nicht versucht, das Layout möglichst pixelgenau zu übertragen (z.B. Margin, Padding), sondern sich auf das Wesentliche beschränkt.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Kreis	Anzahl Bestände, von denen Abortproben eingesendet wurden	Anzahl eingesendeter Aborte ungeachtet der Ursache	Feten	Lochien/ Nachgeburt	1. Blutprobe	2. Blutprobe	Aspergillus fumigata	BHV1-gB-DNA	BTV-RNA (Typ 6)	BYD/MD-Antigen	Bakterien	Brucella
2	Entenhausen	2	2	1	1	1	0	1	0	0	0	4	0
3	Gänsdorf	3	3	0	3	0	0	0	0	0	0	0	0
4	Gesamt	5	5	1	4	1	0	1	0	0	0	4	0

Abb. 3: Mit BIRT generierter Excel-Bericht

Features des Excel Emitters

Das Zusammenfassen (Merge) von mehreren Zeilen oder Spalten ist möglich.

Mehrzeilige Texte innerhalb einer Zelle mit automatischem Zeilenumbruch sind möglich. Voraussetzung für den automatischen Zeilenumbruch ist `whitespace: preformatted` bei der Zelle oder dem Data Item und eine fixe Spaltenbreite.

Das Erzeugen von mehreren Arbeitsblättern (Sheets) in einer Datei durch einen Bericht ist möglich. Ein expliziter Seitenumbruch in BIRT führt zu einem neuen Sheet; der Sheet Name ist steuerbar anhand des Layoutobjekts.

Auch gedrehter Text ist möglich (z.B. um 60° oder 90°). Dazu muss eine User Defined Property `ExcelEmitter.Rotation` vom Typ Integer bei der Cross Tab Cell angelegt werden, der Wert ist der Rotationswinkel.

Ein nicht zu unterschätzendes Feature: Mit der Option `ExcelEmitter.TemplateFile` ist es möglich, eine vorhandene Excel-Datei mit mehreren Sheets als Vorlage zu benutzen und dann mit BIRT nur die Daten in bestimmten Sheets zu aktualisieren. Man kann also z.B. in einem Arbeitsblatt „Daten“ Beispieldaten eintragen und diese im Arbeitsblatt „Auswertung“ mit einer Pivot-Tabelle auswerten. Diese Datei verwendet man als Template und der Bericht erzeugt dann eine Excel-Datei mit den tatsächlichen Daten im Arbeitsblatt „Daten“ und der Pivot-Tabelle im Arbeitsblatt „Auswertung“. Zu beachten ist dabei, dass die Daten in der Pivot-Tabelle (bzw. in Formeln) nicht automatisch aktualisiert werden, da die Datei „an Excel vorbei“ geändert wurde. Mit entsprechenden Einstellungen in der Vorlage (Häkchen bei „Beim Öffnen aktualisieren“ bei der Pivot-Tabelle) und im Bericht (`ExcelEmitter.ForceRecalculate`).

Diagramme wurden bisher nicht getestet, sollten aber funktionieren (allerdings als *Bilder* in Excel).

Schwierigkeiten bei Cross Tabs (Matrixberichte)

Wenn Excel Dateien erzeugt werden sollen, geht es häufig um Matrixberichte. Bei BIRT werden dafür Cross Tabs verwendet. Ganz allgemein gibt einige Einschränkungen für Cross Tabs – ähnlich wie für Matrixberichte bei Oracle Reports). Aufgefallen sind insb. folgende Probleme:

Cross Tabs können nachträglich schlecht erweitert werden (Stand 3.7.1). Es ist manchmal einfacher, anhand eines Backups das Cross Tab aus dem Data Cube neu zu erstellen und zu verfeinern.

Nach dem Versuch, in der Überschrift eine zweite Zeile hinzuzufügen, waren die Überschriften der letzten 3 Spalten um eine Zeile verrutscht.

Anstelle von Cross Tabs können auch speziell formulierte SQL Abfragen verwendet werden, um eine Matrix zu simulieren – das ist allerdings etwas umständlich.

Spezielle Probleme beim Spudsoft Excel Emmitter

Drop Cell funktioniert nicht (bei Group Header Cells). Obwohl für die übrigen Zellen im Group Header Padding 0px eingestellt ist, funktioniert `Drop Cell: Detail nicht`: für den Group Header wird immer eine eigene Zeile in der Excel-Ausgabe erzeugt. Es gibt dafür einen Workaround. die gewünschte Funktionalität kann durch eine Anpassung des Berichts trotzdem erreicht werden.

Hinweis: `Drop Cell` wird Master-Detail-Listenberichte im Stil von Oracle Reports benötigt.

Ohne Drop Cell:

Department	Employee
10	
	SCOTT
	JONES
20	
	SMITH

Mit Drop Cell:

Department	Employee
10	SCOTT
	JONES
20	SMITH

Kurzer Hinweis zum PDF Emmitter

Der PDF Emmitter ist sozusagen die Referenz (unterstützt die meisten Features). Allerdings gibt es ein Problem mit der Seitenzählung. Wir verwenden daher einen eigenen iText-basierenden PostProcessor.

Der BIRT Bericht sagt dem PostProcessor nur, wo und wie die Seitenzahlen ausgegeben werden sollen. Der PostProcessor kennt die Seitenzahl und die TOC-Struktur und schreibt dann die Seitenangaben auf die Seiten. Damit sind auch die bedingte Ausgabe von Seitenzahlen (z.B. nur auf der ersten Seite, nur in einem bestimmten Abschnitt) und gruppierte (abschnittsweise) Seitenzählung möglich.

Best Practises bei der Erstellung von BIRT Berichten für Word- und Excel-Ausgabe

Goldene Regel: keine überflüssigen Strukturen im Layout zu verwenden, da sich diese bei der Word-Ausgabe eher störend auswirken als bei der PDF-Ausgabe.

Entscheidung Table Item oder List Item: Bei Master-Detail-Berichten sind häufig nur die Detail-Daten in Tabellenform, während die Master-Daten eher „Formularform“ haben. Typisches Beispiel: Master: Rechnung mit Empfänger, Rechnungsnummer, Datum usw. Detail: Rechnungspositionen in Tabellenform. Für die Rechnung sollte ein List Item gewählt werden, für die Positionen ein Table Item.

Für Tabellenzeilen oder Listenzeilen, die möglicherweise komplett leer sein können, einen passenden Visibility-Ausdruck bei der Row definieren, damit im Word-Dokument keine Leerzeilen entstehen. Das ist manchmal nicht ganz einfach, wenn die benötigte Information zwar innerhalb der Row, aber nicht für die Row selbst verfügbar ist. Workaround in solchen Fällen: Die Bedingung vorher bzw. weiter oben berechnen und in einer Report Variablen zwischenspeichern.

Für anspruchsvollere Aufgaben bei der Seitennummerierung von PDF-Ausgaben kann ein iText-Nachbrenner verwendet werden. T&P hat dazu eine Lösung entwickelt (bei Interesse bitte melden).

Die HTML-Vorschau oder Preview im Designer nur für die Suche nach Skriptfehlern verwenden, nicht für die Optimierung des Layouts.

Das Layout zunächst mit der PDF-Ausgabe entwickeln und testen. Dabei die Macken des Word Emitters kennen und die hier beschriebenen Regeln berücksichtigen. Erst wenn das Layout soweit fertig ist, die Word-Ausgabe prüfen. Notwendige Korrekturen mit Hilfe eines Word-Makros durchführen.

Kontaktadresse:

Henning von Barga
Triastram & Partner GmbH
Kohlenstraße 55
D-44795 Bochum

Telefon: +49 (0) 234 94375-0
Fax: +49 (0) 234 452206
E-Mail h.vonbarga@t-p.com
Internet: www.t-p.com