

Web Relaunch der Statistik Austria – Ein Erfahrungsbericht

Thomas Schrems, Volker Vãth
Silbury IT Solutions Deutschland GmbH, Fürth

1	Statistik Austria & Silbury	2
1.1	Vorstellung Website	2
2	Projektvorhaben & Anforderungen	2
2.1	Content-Repository, Versionierung.....	2
2.2	Collaboration	2
2.3	Intranet.....	2
2.4	Einfache Befüllung.....	2
2.5	Automatische Freigabeprozesse	3
2.5.1	Kontrolliert	3
2.5.2	Zur richtigen Zeit.....	3
2.5.3	Am richtigen Ort.....	3
2.6	Newsletter.....	3
3	Herausforderungen	3
3.1	IBM-Umfeld.....	3
3.2	SSXA.....	3
3.3	Performance.....	3
4	Umsetzung.....	4
4.1	Lösungsansatz	4
4.2	Content-Repository, Versionierung, Berechtigungen	4
4.3	Collaboration	4
4.4	Einfache Befüllung.....	4
4.5	Automatische Freigabeprozesse	5
4.5.1	Kontrolliert	5
4.5.2	Zur richtigen Zeit.....	5
4.5.3	Am richtigen Ort.....	5
4.6	Newsletter bei Freigabe gewisser Dokumente	6
4.7	Caching.....	6
4.8	SSXA.....	6
4.9	Migration	7
5	Résumé	8

1 Statistik Austria & Silbury

Die Bundesanstalt Statistik Österreich ist eine Organisation, die Statistiken aller Art erstellt, einschließlich zusammenhängender Analysen, Prognosen und statistischer Modelle. Mit dem zunehmenden Interesse der Öffentlichkeit an solchen Daten und der immer größer werdenden Datenflut, wurde die Notwendigkeit eines zentralen Dokumenten-Management-Systems mit automatisierten Prozessen und dessen Anbindung an eine Website immer deutlicher.

Angefangen bei der Konzeption und Inbetriebnahme des ersten Enterprise Content Management Systems 2004, über mehrere (Re)launches der Website und des Intranets, entwickelte Silbury für Statistik Austria bis heute Erweiterungen für Backend & Web und leistet Support.

1.1 Vorstellung Website

Die Website bietet, dank durchdachter Struktur, gezielten Zugriff auf die gewünschten Dateien der verschiedenen Themenbereiche. Den Nutzern sollen statistische Informationen, in leicht zugänglicher Form und zeitsparend zur Verfügung gestellt werden. Die Befüllung erfolgt über Metadaten automatisch (zeit)gesteuert und der Inhalt setzt sich überwiegend aus eingechekten Dokumenten, sowie manuell gepflegten Inhalten, über maßgeschneiderte Erweiterungen, direkt im Backend, zusammen.

Darüber hinaus gibt es einen kleinen Shop und ein eigens entwickeltes Newsletter-System.

2 Projektvorhaben & Anforderungen

Der Web-Relaunch wurde nicht nur durch den ablaufenden Support der alten WebCenter Content Version 10gR3 getriggert, sondern auch durch den Wunsch nach der Möglichkeit, modernere Features umsetzen zu können und mobile Endgeräte zu unterstützen.

2.1 Content-Repository, Versionierung

Es sollen verschiedenartige Dokumente nachvollziehbar versioniert und mit unterschiedlichen Zugriffsrechten ausgestattet sein, die zum Großteil auch in nativer (Excel) und konvertierter Form (HTML, PDF) automatisch an den entsprechenden Stellen der Website verfügbar gemacht werden.

2.2 Collaboration

Mehrere Redakteure sollen die in einer logischen Struktur eingelagerten Dokumente leicht wiederfinden und bearbeiten können.

2.3 Intranet

Neben dem Internetauftritt wurde auch das Intranet der Statistik Austria auf gleicher technologischer Basis erstellt, so dass sich die Mitarbeiter nicht mit mehreren unterschiedlichen Systemen auseinandersetzen müssen.

2.4 Einfache Befüllung

Da bei der Statistik Austria viele Personen mit dem Erstellen und dem Veröffentlichen von Statistiken betraut sind, muss der Veröffentlichungsprozess einfach gehalten sein. Die Einarbeitungszeit zum Veröffentlichen von Dokumenten auf der Website soll so gering wie möglich sein. Das Mittel der Wahl ist Microsoft Office, da es allen Redakteuren vertraut ist. Darüber hinaus werden an manchen Stellen auch WYSIWYG-Editoren eingesetzt und über Backend-Erweiterungen gepflegte Inhalte auf der Website angezeigt.

2.5 Automatische Freigabeprozesse

2.5.1 Kontrolliert

Neuer Content sollte nicht unbemerkt bzw. unkontrolliert in das System gelangen. Für alle neuen Dokumente muss mindestens eine Benachrichtigung an die Leitung der entsprechenden Direktion geschickt werden. Für spezielle Dokumente, sind vor der Freigabe im System mehrere Kontrollpunkte mit der Möglichkeit der Genehmigung bzw. Ablehnung des Inhaltes vorgesehen.

2.5.2 Zur richtigen Zeit

Wird ein Dokument im System freigegeben, so muss dies zu einer festgelegten Zeit auf die Minute genau im System und auf der Website veröffentlicht werden können.

2.5.3 Am richtigen Ort

Die veröffentlichten Informationen können auf der Website an mehreren Stellen benötigt werden. Die Frage des Veröffentlichungsortes auf der Website wurde automatisiert über Metadaten gelöst. Diese werden zur Kategorisierung des Contents, bereits beim Einchecken vergeben.

2.6 Newsletter

Besucher der Website sollen über unterschiedliche neue Veröffentlichungen, mit der Erscheinung im Web, automatisch über ein Newslettersystem per E-Mail informiert werden können, nachdem diese sich dafür angemeldet haben. Dabei können die Besucher aus verschiedenen Themenbereichen wählen und bekommen auch nur die Informationen, für die sie sich interessieren.

3 Herausforderungen

3.1 IBM-Umfeld

Die größte Herausforderung ist die Kompatibilität bzw. Verträglichkeit der beiden Welten "IBM" und "Oracle". Zwar sind alle benutzten IBM-Produkte seitens Oracle zertifiziert und unterstützt, jedoch oft nur in älteren Versionen. Darüber hinaus ist die Oracle Dokumentation an einigen wichtigen Stellen lückenhaft.

3.2 SSXA

Da mit der vorher eingesetzten UCM Version noch Site Studio 10gR3 eingesetzt wurde, war die Migration auf 11g mit SSXA (Grundlage ist hier Site Studio 10gR4) quasi einer Neuprogrammierung gleichzusetzen.

SSXA (Site Studio for External Applications) existiert im Moment in der Version 11.1.1.6. SSXA baut auf Site Studio auf, es sind jedoch nicht alle Backendfunktionen konsequent für SSXA erweitert worden, so dass verschiedene Funktionalitäten selbst implementiert oder über Drittanwendungen ersetzt werden mussten.

Es wurde versucht, möglichst wenige Schnittstellen von SSXA direkt im Sourcecode zu verwenden, um bei Bedarf nur mit geringem Aufwand auf eine andere Technologie wechseln zu können.

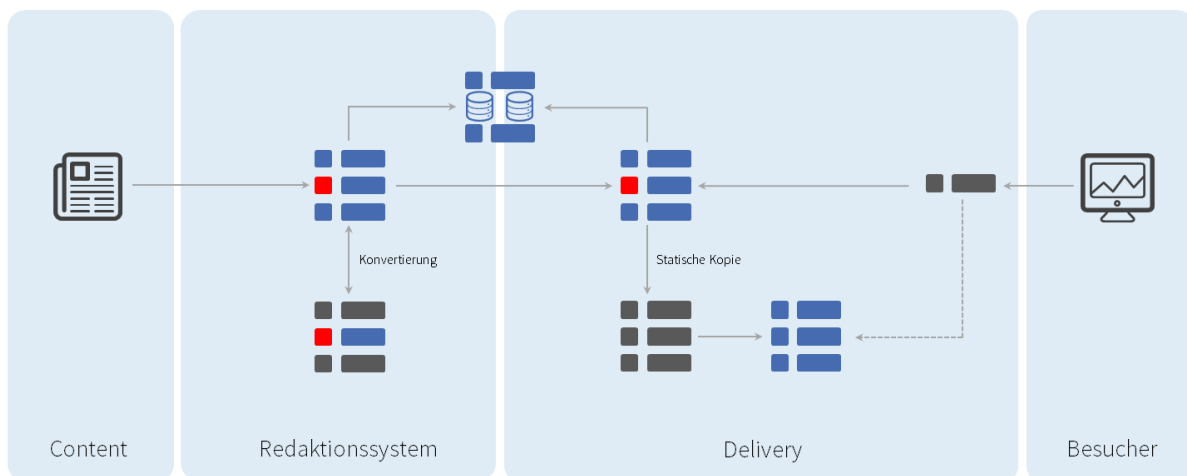
3.3 Performance

Die Hardware-Performance ist ausreichend dimensioniert.

Softwareseitig hat sich jedoch das Rendern der jsp-Seiten als zu langsam für den produktiven Einsatz erwiesen. SSXA bietet zwar einige Caching-Mechanismen an, die sich aber primär dem Problem des schnellen Contentbezugs aus dem WebCenter Content widmen.

4 Umsetzung

4.1 Lösungsansatz



Für die Dokumentenverwaltung, Konvertierung und Webseitenausgabe wird Webcenter Content inkl. Site Studio for External Applications eingesetzt und ist durch sog. Custom Components auf die Anforderungen der Statistik Austria angepasst. Eine statische Kopie der Website dient als Fallback - z.B. im Wartungsfall.

4.2 Content-Repository, Versionierung, Berechtigungen

Als Grundlage für die Lösung wurde WCC als ECM-System eingesetzt, womit die Basisfunktionalitäten, die an ein Content Management System gestellt werden (Zugriffsrechte, Versionierung, Office-Integration), erfüllt waren.

Über Gruppen, die aus dem bei Statistik Austria vorhandenen LDAP bezogen werden, wird gesteuert, wer welche Dokumente bearbeiten darf. Es findet ein Mapping auf die in WCC vorgesehenen Rollen statt, die den Zugriff auf die Security Groups steuern.

4.3 Collaboration

Eine umfangreiche Printpublikation musste bisher aufwendig durch Zusammentragen aller Einzelbestandteile erstellt werden. In der von Silbury entwickelten Lösung werden Dokumente, die für diese Publikation bestimmt sind, direkt bei der Anfertigung an der richtigen Stelle im System abgelegt. So entsteht die umfangreiche Publikation in seiner endgültigen Form bereits durch die Zusammenarbeit aller beteiligten Mitarbeiter – ein spürbares Mehr an Effizienz.

Im Anschluss können alle Daten der Publikation mit einem Mausklick an die Druckerei übertragen werden, die den Band schließlich in Papierform und als CD-ROM ausgibt.

4.4 Einfache Befüllung

Die Webseiten werden von den Redakteuren in Word verfasst. Vorgegebene Formatvorlagen stellen die einheitliche Konvertierung nach HTML sicher. Das passiert zuerst durch das Bordmittel "Dynamic Converter" und im zweiten Schritt durch einen von Silbury entwickelten Parser, um spezielle Anforderungen an das HTML zu erfüllen (Linksyntax für interne Verlinkungen, Unterscheidung des Dokumententyps).

“Formatvorlagen” für Tabellen. In Excel gibt es keine Formatvorlagen. Daher wurde von Statistik Austria ein eigenes xml-Tabellenformat für die HTML-Ansicht entwickelt, welches als Plugin von Excel geladen wird und ebenfalls vom Silbury-Parser unterstützt wird.

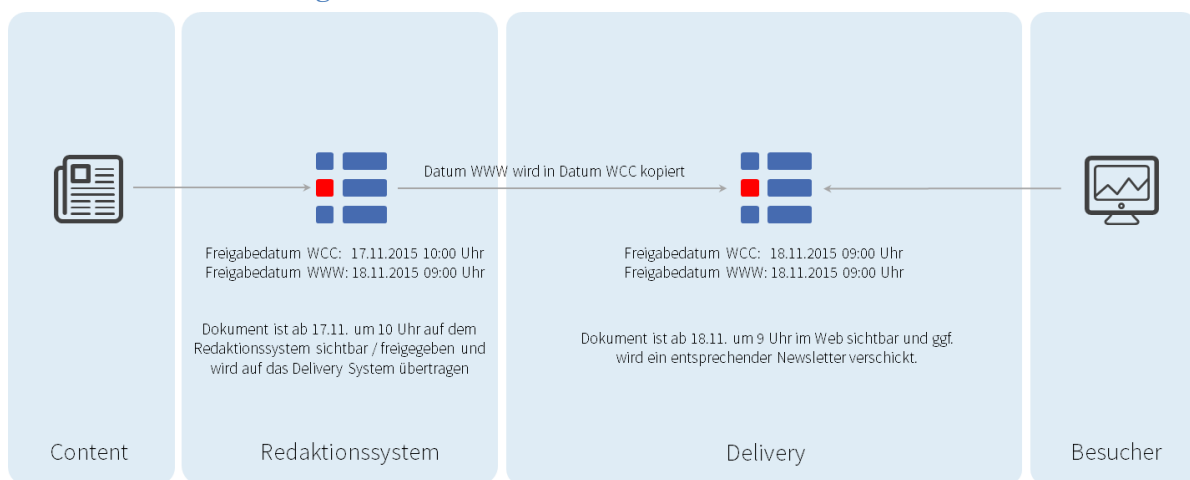
Die Zuweisung der Webseiten zu den Navigationspunkten und die Erstellung neuer Navigationspunkte erfolgt nach dem Site Studio-Standard. Auch die Zuweisung von Datafiles zu Placeholdern z.B. auf der Startseite der Statistik Austria ist über die Bordmittel von WCC/SS realisiert.

4.5 Automatische Freigabeprozesse

4.5.1 Kontrolliert

Zur Freigabe von Pressemitteilungen wurde ein Workflow mit Bordmitteln aus dem WCC-Standard realisiert. Zusätzlich gibt es zahlreiche reine Benachrichtigungsworkflows.

4.5.2 Zur richtigen Zeit



Zur pünktlichen Veröffentlichung eines Dokuments wird der WCC-Standard-Prozess zum Veröffentlichung eines Dokuments verwendet. Dieser Prozess setzt das “veröffentlicht”-Flag eines Dokuments, sobald das Datum aus dem Standard-Metadatum “Freigabedatum” erreicht ist. Über einen zweistufigen Ablauf wird dieses Verhalten zu Veröffentlichung auf der Website genutzt.

1. Der Redakteur erstellt/bearbeitet ein Dokument und setzt die Metadaten “Freigabedatum WCC” und “Freigabedatum WWW”
2. Sobald das Dokument auf der Redaktionsinstanz freigegeben ist und das Datum “Freigabedatum WCC” erreicht ist,
3. wird es von der Redaktionsinstanz auf die Delivery-Instanz repliziert (Archiver) - dabei wird das “Freigabedatum WWW” in das “Freigabedatum WCC” geschrieben
4. Sobald auf der Delivery-Instanz nun das Datum “Freigabedatum WCC” erreicht ist, ist das Dokument im Web sichtbar

Schon vor der Freigabe im Web kann der Redakteur das Ergebnis auf der Webansicht der Redaktionsinstanz überprüfen und gegebenenfalls noch Änderungen vornehmen.

4.5.3 Am richtigen Ort

Mit den “Webkategorien” wurde ein Metadatum erstellt, um z.B. ein Dokument auf mehreren Seiten anzeigen zu können. Die Webkategorien bilden einerseits die Hierarchie der Website ab, andererseits können Zweige der Hierarchie dupliziert. Die Webkategorien sind sprachübergreifend und bieten so die Möglichkeit der Vereinheitlichung der Hierarchie zwischen der deutschen Ausgabe der Website

und der englischen Ausgabe. Diese Möglichkeit ist beim Einsatz von SSXA als Webframework für WCC nicht vorgesehen.

Es gibt zwar bereits ein Metadatum, das in etwa die gleiche Funktion besitzt (Websitesections), aber es kann nur eine Section pro Content-Item angegeben werden. Jedem Contentobjekt können mehrere Webkategorien zugewiesen werden.

Nun können die Contentobjekte durch ihre Metadaten (Webkategorie und Sprache) automatisiert auf der Website angeboten werden. Weitere Metadaten (z.B. der Dokumenttyp) steuern, in welchem Reiter der Content dargestellt wird.

4.6 Newsletter bei Freigabe gewisser Dokumente

Journalisten und Bürger sind an Neuveröffentlichungen von Statistiken eines gewissen Themas interessiert. Dafür bietet der Webauftritt www.statistik.at eine Seite, auf welcher sich Nutzer für einen Newsletter registrieren können. Eine feine Granulierung der Themengebiete stellt sicher, dass man nur Neuigkeiten über die Statistiken erhält, die für einen selbst von Belang sind.

Zur Realisierung der Newsletterfunktion wurde WCC um eine Backendkomponente erweitert. Diese Komponente wird während des Freigabeprozesses eines Dokuments aufgerufen. Für den Abgleich, ob ein Dokument für einen registrierten Benutzer von Belang ist und verschickt werden soll, wurde ein eigenes Metadatum eingeführt.

4.7 Caching

Zur Steigerung der Performance wurde ein zusätzlicher Cache implementiert.

Allerdings entspricht ein verzögertes Sichtbarwerden von neuen Dokumenten im Web (z.B. über den Cachetimeout) nicht den Anforderungen der Statistik Austria, da diese die gesetzliche Verpflichtung hat, bestimmte Dokumente (z.B. den Verbraucherpreisindex) zu definierten Uhrzeiten live zu stellen. Das Löschen des kompletten Caches kommt nicht in Frage, da zu diesen Uhrzeiten ein besonders hoher Traffic auf der Seite herrscht und diese Lastspitze durch eine komplette Cacheleerung zusätzlich erhöht würde.

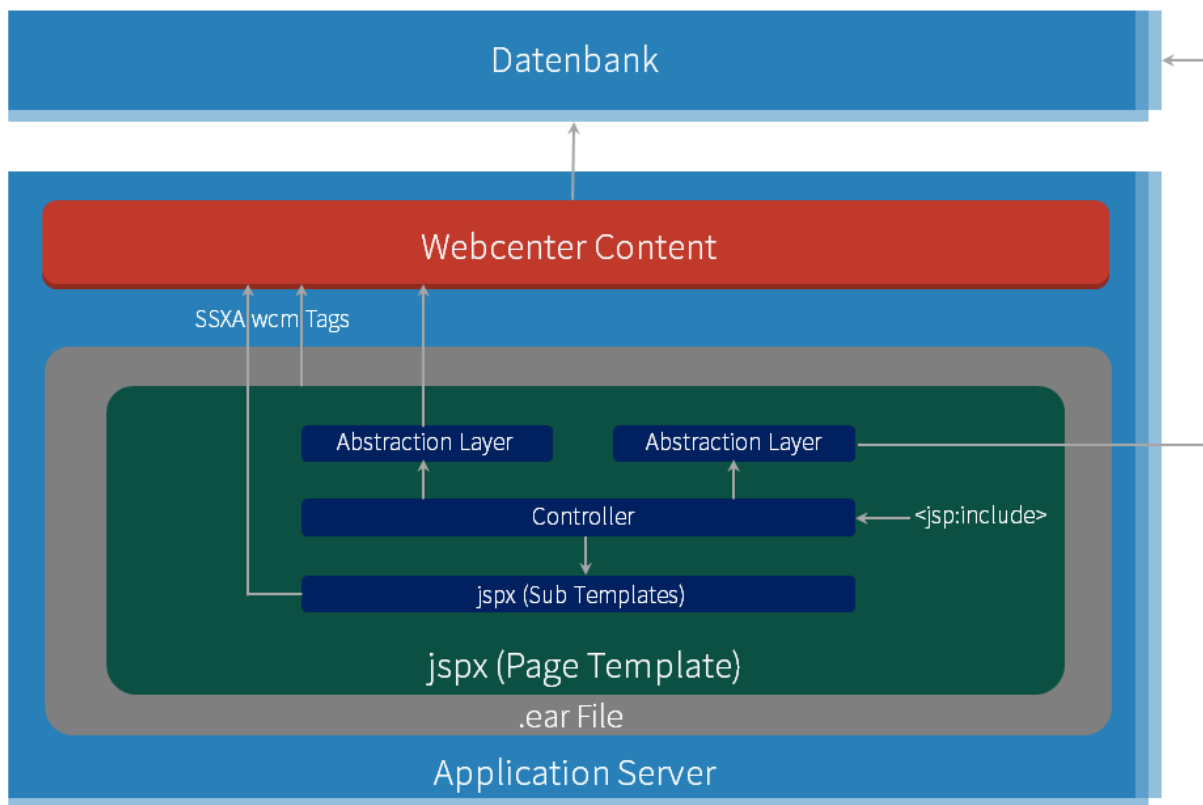
Die Lösung ist eine WCC-Backend-Komponente, die vom Freigabeprozess aufgerufen wird. Diese Komponente steuert die dynamische und gezielte Entfernung der betroffenen Cacheeinträge, so dass neu freigegebene Dokumente sofort im Web sichtbar werden.

Anhand der Metadaten des eingetragenen Dokuments (Dokumenttyp und Webkategorie) werden die URLs ermittelt, welche aus dem Cache der Webinstanz zu entfernen sind. Mittels eines Webserviceaufrufes, übermittelt das Backend dem Frontend diese URLs.

4.8 SSXA

In den jsp-Dateien wird nur eine Auswahl der wcm-tags verwendet (z.B. zur Erzeugung von URLs oder zur Platzierung von "Placeholdern"). Zum Beziehen von Content aus dem WCC wird das wcmproxy-Servlet von SSXA verwendet.

Die Trennung von Logik und Präsentation wurde mittels zusätzlich eingezogener Abstraktionsschicht über einen Controller erreicht.



In Site Studio existiert das Tool Site Studio Publisher, welches in einer früheren Version von www.statistik.at zur statischen Replikation genutzt wurde. Für SSXA steht dieses Tool nicht mehr zur Verfügung, weshalb ein externes Tool zur Erstellung der statischen Kopie eingesetzt wurde. Diese Kopie kann als Fallbacklösung ausgeliefert werden. Dadurch kann ein Großteil des Contents auch in Wartungsfenstern zur Verfügung gestellt werden. Dynamische Seiten, auf denen ein Server zur Verarbeitung nötig ist (wie z.B. Anmeldung für Newsletter oder Formulare), funktionieren in diesem Zeitraum nicht, werden aber durch entsprechende Hinweisseiten ersetzt.

4.9 Migration

Um möglichst keine Performance-Probleme im Betrieb zu provozieren und vorab die Übertragung des Contents auf eventuell fehlerhaftes Verhalten während der Übertragung zu prüfen (Probleme mit Metadaten / Archiver), wurde der meiste Content während eines Wochenendes auf die neuen Systeme übernommen. Am Go-Live Tag war dann nur die Übertragung eines kleinen Deltas nötig.

Die Migration umfasste Dokumente, deren Metadaten, den Webseitenbaum und einige Custom-Tabellen (z.B. die Newsletterabonnements). Schon während der Entwicklung wurden Datafiles, deren Format geändert wurde, konvertiert. Auch von diesen Dateien musste nur eine kleine Zahl am Go-Live Tag mit aktualisiertem Inhalt erneut konvertiert werden.

Mit einem theoretischen Arbeitsstopp von nur 4 Stunden für die Redakteure war die Unterbrechungszeit sehr gering.

5 Résumé

Für die Anforderungen des Kunden ist WebCenter Content, in Verbindung mit SSXA eigentlich die ideale Wahl, da mit vertrauten Werkzeugen (MS Office) erstellte Dokumente per Konvertierungsregeln in HTML oder PDF transformiert werden können und dann zur einfachen Veröffentlichung im Web zur Verfügung stehen. Dennoch gibt es zwei entscheidende Punkte, die zu bedenken sind:

1. Die Zukunft von SSXA ist ungewiss, da schon länger keine neue Version erschienen ist.
2. Die Harmonie zwischen den Welten Oracle und IBM ist nicht perfekt. Sei es Application Server, Datenbank oder Betriebssystem. Es ist im Grunde immer zum entsprechenden Oracle-Produkt zu raten, da dadurch nicht nur die Einrichtung und Wartung um ein vielfaches erleichtert wird, sondern auch die Unsicherheiten bezüglich Zertifizierung zukünftiger WebCenter Versionen passe sind. Außerdem sind nie alle theoretisch verfügbaren Funktionen nutzbar, da mit einer IBM-Landschaft um WebCenter Content einige Features nicht verfügbar sind.