Critical Software S.A.

Especificação da Migração de Dados Históricos

Sistema de Informação de Saúde para Monitoria e Avaliação



©2014 Copyright Critical Software S.A. Todos Os Direitos Reservados.

REFERÊNCIA CONTRATO: CSW-2012-PRL-02976 Código Projeto: SISMA

Ref. Doc.: CSWMZ-SISMA-2013-DDS-

Data: 2013-05-30

Páginas: 17

Estado: Aprovado
Acesso: Público

Versão: 1.0



SALVAGUARDA - Documento sob contrato com Jembi .

O trabalho descrito no presente documento foi desenvolvido sob contrato com Jembi. A responsabilidade pelo seu conteúdo é do autor ou organização que o preparou.

Parceiros:





www.criticalsoftware.com



Aprovação				
Versão	Nome	Funções	Assinatura	Data
01	Paulo Grácio	Gestor Técnico		2013-07-08
01	Luis Gaspar	Gestor de Projecto		2013-07-08

Autores e Contribuintes				
Nome	Contacto	Descrição	Data	
Délio Nhantumbo	danhantumbo@criticalsoftware.com	Autor	2013-05-30	
Paulo Grácio	pgracio@criticalsoftware.com	Autor	2013-07-08	

Lista de Acesso

Acesso Interno

Dep. Qualidade, Equipa de Projeto

Accesso Externo

Equipa de projecto, Dr. Alessandro Campione e Mr. David Mendes, Mr Pascal Brandt, equipa MOASIS e Jembi.

O conteúdo do presente documento é copyright da Critical Software S.A.; Este documento é distribuído sob condição de não ser copiado no seu todo ou em parte, ou reproduzido de qualquer outra forma (seja por meio fotográfico ou qualquer outro meio) e o seu conteúdo não deverá ser em caso algum divulgado a qualquer pessoa ou entidade não referida nas listas de acesso interna e externa sem prévio consentimento por escrito das entidades envolvidas.

Histórico de Revisões				
Versão	Revisão	Data	Descrição	Autor
	0.01	2013-05- 30	Criação do documento	Délio Nhantumbo
	0.02	2013-07- 08	Contribuição e revisão do documento	Paulo Grácio



Histórico	Histórico de Revisões					
	0.03	2013-07- 30	Revisão final do documento	Paulo Grácio		
	0.03	2013-08- 12	Actualização do documento com comentários da Jembi mencionados na reunião de 09-08-2013.	Paulo Grácio		
			Actualização da secção <i>Testes e Validação</i> .			
01		2013-07- 08	Versão entregue ao cliente.	Luis Gaspar		

Índice

1	Int	troduçãotrodução	5
	1.1	Objectivo	5
	1.2	Âmbito	
	1.3	Audiência	5
	1.4	Definições e acrónimos	5
	1.5	Estrutura do documento	
	1.6	Documentos aplicáveis	6
	1.7	Documentos de referência	
2	Pr	rocesso de Migração de Dados	8
3	Me	etodologia	9
	3.1	ETL - Extract, Transform and Load	9
	3.1	1.1 Extracção	9
	3.1		
	3.1 3.1		
	3.2	Desenvolvimento	
	3.3	Testes e Validação	
	3.4	Execução da Migração	
4	Re	ecomendações	

Índice de Tabelas

Tabela 1: Definições	5
Tabela 2: Acrónimos	
Tabela 3: Documentos aplicáveis	7
Tabela 4: Documentos de referência	7

Especificação da Migração de Dados Históricos



Tabela 5 – Exemplo de mapeamento par a Unidades Organizacionais	Básico e
Tabela 7 – Exemplo de mapeamento de elementos de recolha de dados Module	o Básico
Anexo e SIS-MA.	11
Tabela 8 - Exemplo tabela temporária para extracção de dados	11
Índice de Figuras	
Figura 1 - Processos de Migração de dados	8
Figura 2 – Importação de dados manualmente	13
Figura 3 – Exemplo de processo de ETL usando TOS	



1 Introdução

1.1 OBJECTIVO

Este documento apresenta o modelo de Migração de Dados Históricos da actual aplicação SIS-MA.

O nível de detalhe que o documento apresenta tem como principal objectivo fornecer às pessoas que irão estar envolvidas na Migração de Dados históricos do sistema actual (Módulo Básico e Anexo), uma base estável para o entendimento de procedimentos para Migrar os dados para o novo sistema, SIS-MA com base no DHIS2.

1.2 ÂMBITO

Este documento apresenta a o modelo de Migração de dados Históricos do actual sistema de acordo com o estabelecido na proposta na tarefa H – Especificação da Migração de Dados Históricos do actual sistema, Módulo Básico, para o SIS-MA.

1.3 AUDIÊNCIA

A audiência deste documento é constituída pelas entidades Jembi, mOASIS e MISAU responsáveis pela especificação e validação do sistema, bem como pela equipa de projecto que venha a ser designada para participar no projecto e pelos elementos da Critical Software que serão responsáveis pelos trabalhos a desenvolver.

1.4 DEFINIÇÕES E ACRÓNIMOS

A Tabela 1 apresenta a lista de definições usadas neste documento.

Nome	Descrição
Documento Aplicável	Um documento é considerado aplicável se complementar este documento. Todo o seu conteúdo é diretamente aplicável como se tivesse sido escrito como anexo deste documento.
Documento de Referência	Um documento é considerado de referência se neste documento existem referências ao seu conteúdo, mas este não é totalmente aplicável. Os documentos de referência são principalmente utilizados para fornecer leitura e recursos adicionais.

Tabela 1: Definições

A Tabela 2 apresenta a lista de acrónimos usados neste documento.

Acrónimo	Descrição
AD	Applicable Document (documento aplicável)
CSW	Critical Software, S.A.
DHIS	District Health Information System



Acrónimo	Descrição	
SIS-MA	Sistema de Informação de Saúde para Monitoria e Avaliação	
ETL	Extract, Transform and Load	
SDMX-HD	Statistical Data and Metadata eXchange	
DXF	Drawing Exchange Format	
XML	Extensible Markup Language	
HTML	HyperText Markup Language	
MS	Microsoft	
UIDs	Unique Identification Number	
IXF	Integration Exchange Format	
RD	Reference Document (documento de referência)	
TBC	To be confirmed (a confirmar)	
TBD	To be defined (a definir)	
TOS	Talend Open Studio	

Tabela 2: Acrónimos

1.5 ESTRUTURA DO DOCUMENTO

A secção 1 (Introdução) apresenta uma descrição geral do conteúdo deste documento.

A secção 2 (Processo de Migração de Dados) apresenta os processos de migração de dados para o novo SIS-MA.

A secção 3 (Metodologia) apresenta as metodologias que permite a migração de dados do módulo básico e anexo para o SIS-MA.

A secção 4 (Recomendações) apresenta as recomendações de forma sumária a serem tidas em consideração quanto à migração de dados.

1.6 DOCUMENTOS APLICÁVEIS

A Tabela 3 apresenta a lista de documentos aplicáveis a este documento.

Documento aplicável	Número do documento
[AD-1] Proposta Sistema de Informação de Saúde para	CSW-2012-PRL-02976-01



Documento aplicável	Número do documento
Monitoria e Avaliação, 2012-07-30, Critical Software SA	
[AD-2] Adenda à Proposta Sistema de Informação de Saúde para Monitoria e Avaliação, 2012-12-18, Critical Software SA	CSW-2012-PRL-02976-05

Tabela 3: Documentos aplicáveis

1.7 DOCUMENTOS DE REFERÊNCIA

A Tabela 4 apresenta a lista de documentos de referência a este documento.

Documento de referência	Número do documento
[RD-1] Ferramentas de Projecto	CSWMZ-SISMA-2013-MMO-00072
[RD-2] Modelo de Dados do Módulo Básico	CSWMZ-SISMA-2013-MMO-00083
[RD-3] Unidades Sanitárias do Módulo Básico	CSWMZ-SISMA-2013-MMO-00157
[RD-4] Plano e Especificação de testes	CSWMZ-SISMA-2013-TCS-00342

Tabela 4: Documentos de referência



2 Processo de Migração de Dados

O processo de migração de dados será feito com base em na tecnologia assente na plataforma DHIS2 base e será executado de modo estruturado e faseado. O processo de migração deve conter as seguintes fases:

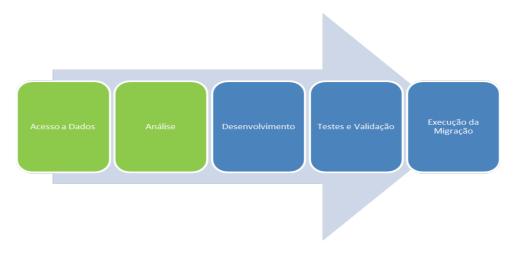


Figura 1 - Processos de Migração de dados

- Acesso a dados o Módulo Básico será a única fonte de dados a considerar para migração de dados de histórico.
- Análise análise dos dados existentes no módulo básico de forma a identificar qual a informação a migrar para o SIS-MA. Este trabalho de análise já foi iniciado, durante a Fase 1 – Estudo, estando o resultado desta análise sistematizado no documento [RD-2] Modelo de Dados do Módulo Básico. Esta actividade visa essencialmente compreender o conteúdo, estrutura e qualidade dos dados a migrar.
- Desenvolvimento desenvolvimento de um mecanismo de migração de dados de acordo com os requisitos de migração definidos na secção 3 - Metodologia.
- Testes e Validação execução de testes com vista a validar os dados migrados. Comparação antes e depois da migração para identificar eventuais inconsistências nos dados migrados.
- Execução da migração execução da migração em ambiente Piloto e ambiente de Produção.



3 Metodologia

A metodologia a ser utilizada na migração de dados assenta na utilização dos mecanismos de importação de dados já disponibilizados pelo DHIS2 base. Para este efeito, os dados actuais dos sistemas devem ser extraídos e transformados para o formato de importação/exportação do DHIS2 base, que é baseado em ficheiros de XML no formato DXF2¹. Os ficheiros são posteriormente importados, ou de forma automática, através da API nativa do DHIS2, ou de forma manual, no sistema central e/ou nos ambientes offline. Segue-se no documento a fundamentação referente às técnicas e processos de ETL utilizadas para o desenvolvimento da ferramenta de Migração de Dados.

3.1 ETL - EXTRACT, TRANSFORM AND LOAD

ETL, *Extract, Transform and Load* é um processo, suportado por ferramentas de *Software*, que se destinam a extracção, transformação e carregamento de dados. Estes dados podem ser originários de uma ou mais bases de dados, tendo como destino o sistema central do SIS-MA.



Embora o processo de migração suporte o carregamento de dados de diferentes base de dados, de forma faseada, para o sistema central do SIS-MA é recomendado que seja feita uma consolidação de dados na Base de Dados Central do Módulo Básico e só posteriormente executado o processo de migração.

É nossa recomendação que os dados sejam migrados sem qualquer correcção/transformação, podendo, caso necessário, proceder à correcção dos dados depois da migração usando os mecanismos de Qualidade de Dados disponibilizados pelo SIS-MA:

- Análise das regras de validação executar as regras de validação, a fim de revelar anomalias e erros nos dados existentes na base de dados.
- Análise de discrepâncias de Desvio Padrão Analisar valores discrepantes com base em desviospadrão. Valores discrepantes podem ser examinados e marcados para acompanhamento.
- Análise de discrepâncias de Min-Max Analisar valores discrepantes com base em valores mínimos e máximo. Valores discrepantes podem ser examinados e marcados para acompanhamento.

Esta abordagem visa essencialmente garantir que os dados são migrados exactamente como se encontram no Módulo Básico e só corrigidos se assim for decidido pelos Chefes de Programas Nacionais e/ou Responsável de Monitoria e Avaliação.

3.1.1 EXTRACÇÃO

Na fase inicial do processo de ETL, temos a extracção de dados de origem, neste caso, provenientes da base de dados de suporte Módulo básico e Anexo. São criados ficheiros DFX2, com os dados extraídos do MSAccess. Estes ficheiros podem ser carregados de forma automática ou manual no sistema central SIS-MA.

.

¹ Formato padrão DHIS 2 para carregamento de dados em bulk.

Especificação da Migração de Dados Históricos



O processo de extracção, regista a sua data de início, data de fim e argumentos utilizados num ficheiro de Log para um melhor controlo da migração.

É necessário que uma tabela de Excel de mapeamento seja alimentada pelos distintos códigos encontrados. Essa tabela será necessária para mapear automaticamente dados de um sistema para o outro, sendo facilmente detectados aqueles que ainda não possuem mapeamento efectuado.

Este processo, ao ser executado uma segunda vez, sobrepõem os ficheiros, DXF2, existentes na área de staging,.

Seguem-se exemplos de tabelas de mapeamento:

Código de País	Código de Província	Código de Distrito	Código Unidade Sanitária	REF – SIS-MA Código
01	11	01	99	01110199
01	11	01	38	01110138
01	11	01	35	01110135
01	11	01	11	01110111
01	11	01	10	01110110

Tabela 5 – Exemplo de mapeamento par a Unidades Organizacionais

A Tabela 5 – Exemplo de mapeamento par a Unidades Organizacionais, poderá não ser necessária caso todos os códigos usados no SIS-MA sejam possíveis de gerar de forma automática tendo como base a informação existente no Módulo Básico.

No Módulo Básico e Anexo os dados são armazenados de forma distinta. No MB existe um tabela por cada ficha, DATA<Código da Ficha>. As primeiras 6 colunas são usadas como chave primária e as restantes colunas para os elementos de dados, em que cada coluna corresponde a uma variável do formulário.

Formulário	Coluna	REF – SIS-MA Data Element
B-14	A1	zWfxUDBWlil
B-14	A2	m5OF1GuXJXa
B-14	A3	CbHKC5l1Dtl
B-14	A4	E27Rue7PZ5b

Tabela 6 – Exemplo de mapeamento de elementos de recolha de dados Modulo Básico e SIS-MA.

No Módulo Básico Anexo os dados são armazenados na tabela FICHA_DADOS onde a coluna FORMULARIO_ID e VARIAVEL_ID identificam a ficha e a variável de recolha respectivamente.



FORMULARIO_ID	VARIAVEL_ID	REF – SIS-MA Data Element
B04	A10	f7n9E0hX8qk
B04	A11	f7r9E0yX7qk
B04	A12	d7n9v0hX8we
B04	A13	asdouy766ads

Tabela 7 – Exemplo de mapeamento de elementos de recolha de dados Modulo Básico Anexo e SIS-MA.

Estes dados devem ser obtidos do MS Access, que é a base de dados de suporte ao Módulo Básico e Anexo, sendo depois carregados para uma estrutura temporária com o seguinte formato:

dataElement	period	orgUnit	value
f7n9E0hX8qk	201201	DiszpKrYNg8	12
f7n9E0hX8qk	201201	FNnj3jKGS7i	14
f7n9E0hX8qk	201202	DiszpKrYNg8	16
f7n9E0hX8qk	201202	Jkhdsf8sdf4	18

Tabela 8 - Exemplo tabela temporária para extracção de dados



O processo não deve realizar quaisquer operações de escrita/modificação nas tabelas existentes, de forma a assegurar que os dados originais não são modificados.

3.1.2 TRANSFORMAÇÃO

Este é o processo responsável pelo tratamento e transformação de dados após o processo de extracção.

Durante a execução deste processo, caso existam valores por mapear, os elementos ignorados devem ser apresentados no log da migração, para posterior análise. O processo de transformação, regista igualmente a sua data de início, data de fim e argumentos utilizados na execução do processo no ficheiro de Log.

Este processo, ao ser executado uma segunda vez, apaga todos os ficheiros previamente gerados em directório configurado para o output da rotina.

A transformação de dados consiste na produção de um ficheiro temporário no formato *DXF2* DataValueSet que é o formato que facilmente interpretado pelo sistema SIS-MA. Tendo como exemplo o conteúdo da Tabela 8, o output gerado durante o processo de transformação seria o seguinte:

Especificação da Migração de Dados Históricos



Os ficheiros são colocados em pastas temporárias de acordo com a configuração existente.



Esta abordagem segue a recomendação da comunidade DHIS2 para importação de dados em *bulk*.

Durante a execução da migração devem ser gerados outputs e registados no ficheiro de *log*, de modo a manter o controlo da execução do processo de Migração. O registo de log deve permitir também manter o controlo e registo de todas operações para quando um problema ocorrer, ser fácil de identificar e corrigir. No caso de ocorrer um problema numa operação, deve ser possível identificar as causas do problema de forma a definir acções a serem tomadas para a resolução dos problemas em causa.

3.1.3 CARREGAMENTO

O processo de carregamento consiste no envio dos dados extraídos, tratados e transformados nas etapas anteriores, através da API WEB² existente no sistema central SIS-MA.

O envio de dados para o SIS-MA será feito usando o cURL³ que permite o carregamento de forma simples dos dados no sistema central. Exemplo de envio de dados:

```
curl -d @datavalueset.xml "http://apps.dhis2.org/demo/api/dataValueSets" -H "Content-Type:application/xml" -u admin:district -v
```

Dependendo das necessidades do processo montado, esta etapa pode ser realizada uma única vez ou de forma faseada.

Existe também a opção de fazer o upload do ficheiro de forma manual através da aplicação, usando o serviço de importação de dados, de acordo com a Figura 2 – Importação de dados manual.

http://www.dhis2.org/doc/snapshot/en/user/html/dhis2_user_manual_en_full.html , 25.9. Sending large bulks of data values

³ http://curl.haxx.se/docs/manpage.html



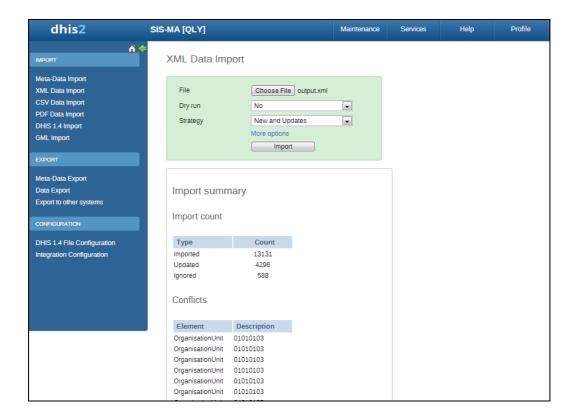


Figura 2 – Importação de dados manualmente

Depois de importados os dados de forma correcta, sem que existam conflitos ou dados ignorados, os formulários devem ser fechados para os períodos previamente importados. O exemplo seguinte apresenta os dados necessários para o fecho de um Formulário, num determinado Período, para uma Unidade Organizacional.

```
<?xml version='1.0' encoding='UTF-8'?>
<dataValueSet xmlns="http://dhis2.org/schema/dxf/2.0" dataSet="NxK5l1INAp1"
completeDate="2013-07-23" period="201201" orgUnit="HuB171Ny0HZ"/>
```

3.1.4 CONFIGURAÇÕES

De forma a tornar o processo de migração o mais flexível possível, será necessário ter um mecanismo de configuração, que permita definir o seguinte:

- Caminho para o ficheiro da base de dados do módulo básico. Default: C:\HISVB\Moz\Data\ His1.mdb\
- Pasta temporária para armazenar os ficheiros de output gerados.
- URL da Web API, Default: URL do sistema central SIS-MA
- · Utilizador válido no sistema SIS-MA
- · Password válida no sistema SIS-MA

Especificação da Migração de Dados Históricos



- · Data de início para migração de dados
- Data de fim para migração de dados
- · Códigos das fichas a migrar dados
- Modo de execução:
 - extraction: apenas efectua o processo de extracção de dados, carregando os dados pretendidos para a tabela temporária.
 - o transform: efectua o processo de extracção de dados e gera os ficheiros de output
 - dryrun: apenas obtem um resumo da importação no SIS-MA, sem que sejam guardados os dados.
- Estratégia de importação:
 - o new_and_updates: importa todos os dados
 - o new: importa apenas dados novos
 - o updates: importa apenas actualizações

sourcedatabase= C:\HISVB\Moz\Data\ His1.mdb
tmpdir= C:\SIS-MA\migration\tmp\
url=http://apps.dhis2.org/demo/api/dataValueSets
username=admin
password=xpto
start=012012
end=121012
codes=B01-B;B04
execution=dryrun
strategy= new_and_updates

As configurações apresentadas podem ser modificadas durante a fase de desenvolvimento da Migração caso se justifique com o objectivo de tonar o processo de migração de dados o mais robusto possível.



3.2 DESENVOLVIMENTO

O desenvolvimento do processo de migração será feito tendo como base *Talend Open Studio for Data* Integration⁴, TOS. Recorrendo a este Software, teremos acesso a um conjunto de ferramentas que permitem aceder, transformar e carregar os dados no SIS-MA de forma flexível.

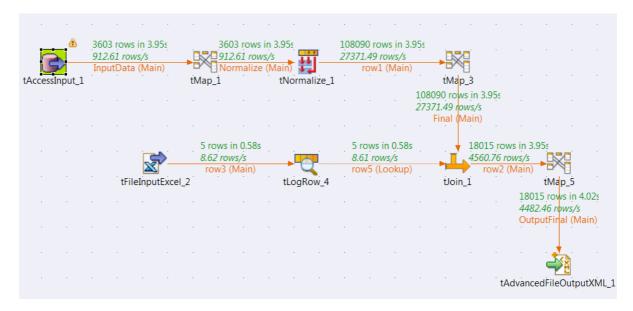


Figura 3 – Exemplo de processo de ETL usando TOS.

O TOS disponibiliza um conjunto de componentes que vão facilitar a produção da migração, sendo de destacar:

- Ligação a base de dados Access;
- Carregamento de dados para tabelas temporárias;
- Produção de ficheiros xml com output de dados;
- Produção de Log com erros e de controlo de execução;

3.3 TESTES E VALIDAÇÃO

Na secção *Migração de dados históricos* do documento Plano e Especificação de Testes [RD-4], são detalhados os testes que serão realizados com o objectivo de validar a correcta migração de dados de histórico.

http://www.talend.com/download/data-integration http://www.talend.com/products/data-integration

Especificação da Migração de Dados Históricos



3.4 EXECUÇÃO DA MIGRAÇÃO

A execução da migração será feita pela equipa de projecto utilizando uma base de dados central, com os dados consolidados, provenientes dos Distritos e Províncias, seguindo o fluxo de dados actualmente em vigor para o Módulo Básico. Desta forma, o SIS-MA ficará com o mesmo conjunto de dados existente no Módulo Básico, facilitando a validação de cálculos de indicadores e produção de relatórios.

Especificação da Migração de Dados Históricos



4 Recomendações

O processo de migração nunca deverá ser feito sobre a base de dados original do Módulo Básico, deve ser feita uma cópia de forma a assegurar que, mesmo que algo corra mal durante a migração, os dados não são comprometidos. Esta acção tem como principal objectivo assegurar que o processo de migração de dados do Módulo Básico para o SIS-MA não causa qualquer impacto na operação do Módulo Básico.





COIMBRA HEADQUARTERS, PORTUGAL

Parque Industrial de Taveiro, Lote 49 3045-504 Coimbra, Portugal Tel.: +351 239 989 100 Fax: +351 239 989 119

LISBON OFFICE, PORTUGAL

Campus do Lumiar, Edifício M8 Estrada do Paço do Lumiar 1649-038 Lisboa, Portugal Tel.: +351 217 145 430 Fax: +351 217 145 432

OPORTO OFFICE, PORTUGAL

Rua Eng.º Frederico Ulrich, nº 2650 4470-605 Moreira da Maia, Portugal

Tel.: +351 229 446 927/8 Fax: +351 229 446 929

SÃO PAULO, BRAZIL

Parque Tecnológico UNIVAP Av. Shishima Hifumi 2911 - Urbanova S. José dos Campos, S. Paulo SP - CEP: 12244-000, Brazil Tel: +55 12 3949 2512

SINGAPORE OFFICE, SINGAPORE

3 Temasek Avenue, #31-02 Centennial Tower, Singapore, 039190 Tel: +65 68 3677 98 Fax: +65 68 3677 96

CHICAGO OFFICE, USA

70 West Madison St, Ste 5750 Chicago Illinois, USA, 60602 Tel.: +1 (312) 772-6232

SOUTHAMPTON OFFICE, UK

2 Venture Road Southampton Science Park Southampton SO16 7NP, UK Tel.: +44 (0) 23 8011 1339 Fax: +44 (0) 870 762 2487

YEOVIL OFFICE, UK

Yeovil Innovation Centre Barracks Close, Copse Road Yeovil, Somerset BA22 8RN, UK Tel.: +44 (0)1935 385911 Fax: +44 (0)870 762 2487

MAPUTO OFFICE, MOZAMBIQUE

Rua Pereira Marinho, 179 Bairro da Sommerchield Maputo, Mozambique Tel.: +258 826 038 600

LUANDA OFFICE, ANGOLA

Rua Pedro Félix Machado, nº16, Mutamba Luanda, Angola

















