

1° CONVEGNO ITALIANO DEGLI UTENTI DI STATA

**OUTPUT DI STATA
IN MARK-UP LANGUAGES:
GENERAZIONE AUTOMATICA
DI TABELLE IN HTML E LATEX**

**Rosa Gini
Jacopo Pasquini**

Osservatorio di Epidemiologia
Agenzia Regionale di Sanità Toscana

Roma, 25 Ottobre 2004



OUTPUT DI STATA IN MARK-UP LANGUAGES:

Cos'è un linguaggio di tipo MARK-UP?

Un linguaggio di markup dei documenti è un insieme di elementi (chiamati "tag") che hanno una o più delle seguenti funzioni:

- Descrizione della struttura del documento
- Descrizione del contenuto del documento
- Controllo di come il documento si presenta all'utente

“A way to add computer-understandable information to text files. Certain parts of the text file are interpreted as mark-up instead of content. This markup may contain instructions for the computer. The interpretation of those instructions is defined by the semantics of a particular markup language”
(www.watchfire.com)

FORMATTAZIONE + CONTENUTI + CONTROLLO

TESTO → COMPILAZIONE → OUTPUT

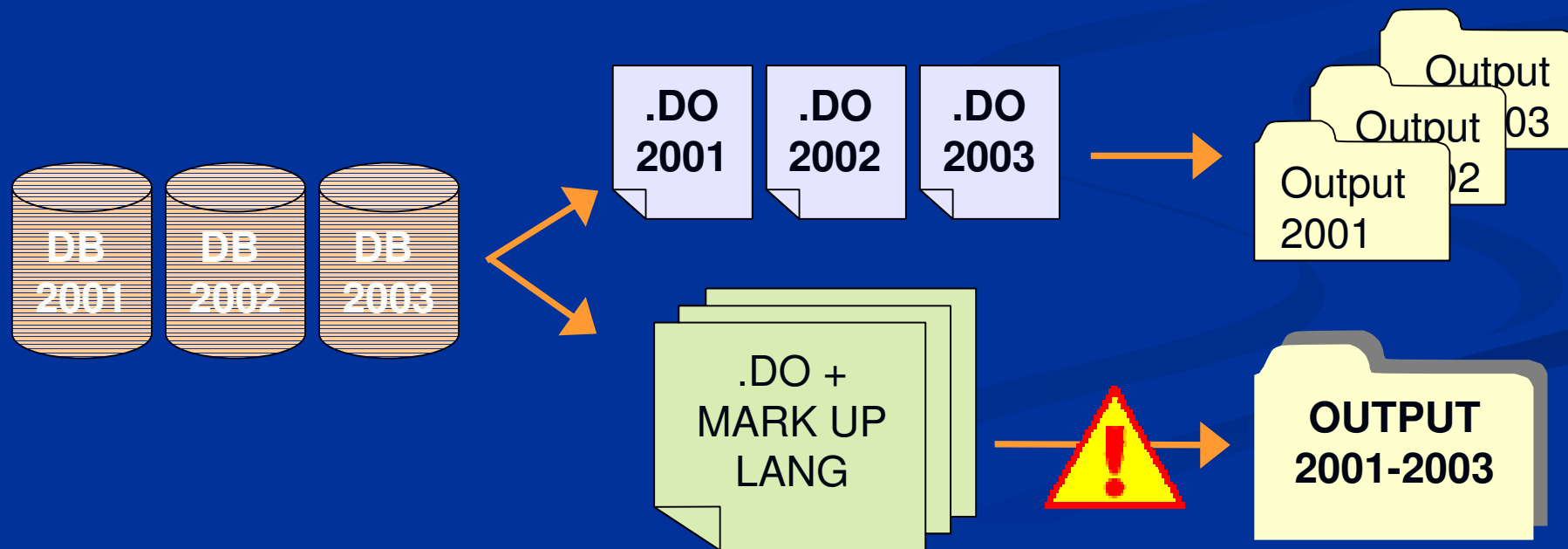
OUTPUT DI STATA IN MARK-UP LANGUAGES:

ESEMPI DI LINGUAGGI MARK-UP:

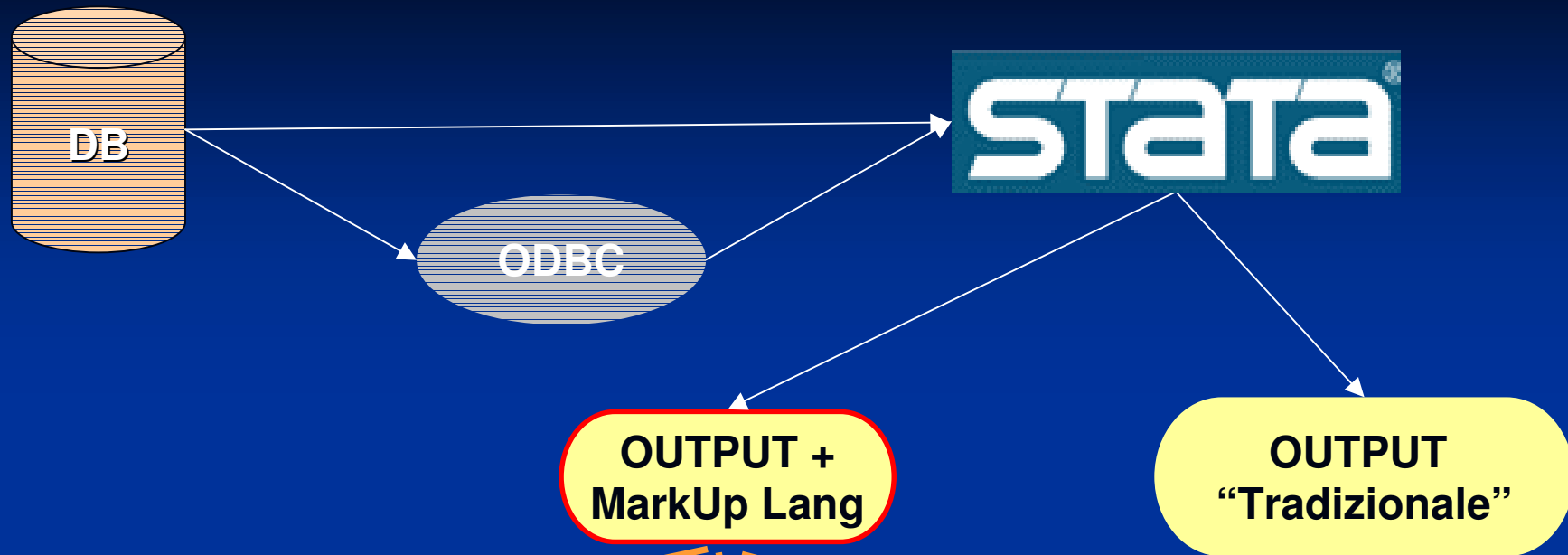
- CHTML : (Compact Html – DoCoMo telephones)
- CML (Chemical Markup Language)
- DocBook (Standard XML - creazione di documentazione)
- GML (Geographic Markup Language - sistemi GIS)
- HDML (Hendheld Device markup Language - PDA)
- HTML (HyperText Markup Language – sviluppo web)
- inkML (Ink Markup Language – per input da penne elettroniche)
- MathML (mathematical Markup Languages)
- MusicXML (Music XML)
- RDF (Resource Description Framework – metadata model)
- RSS (Really Simple Syndication – web Syndication)
- SGML (Standard Generalized Martkup Language – metalinguaggio MU)
- SMIL (Synchronized Multimedia Integration Language – pres. multimediali da XML)
- SML (Spacecraft Markup Language)
- SOAP (Simple Object Access Protocol – comunicazioni tra software)
- SVG (Scalable Vector Graphics)
- TEX e LaTeX (editor scientifico e programmazione)
- VRML (Virtual Reality Model language – modelli 3D sul Web) (diventa X3D)
- WML (Wireless Markup Language)
- Xforms (XML Form – realizzazione di Form per raccolta dati sul web)
- XML (Extensible Markup Language – per trasporto dati sul web)

OUTPUT DI STATA IN MARK-UP LANGUAGES:

- SCOPO:**
- Utilizzare STATA per generare **OUTPUT COMPLESSI**
- OUTPUT COMPLESSI = RISULTATI ANALISI + "REPORTISTICA"**
- COME:**
- Includendo il linguaggio **MARK-UP** nell' **OUTPUT**
- QUANDO:**
- In presenza di flussi di dati correnti
 - Per esigenze di formattazione dell'output
 - In caso di analisi ripetute
 - Per grandi moli di output (grafici, tabelle ...)



OUTPUT DI STATA IN MARK-UP LANGUAGES:



IMPAGINAZIONE,
LAYOUT,
PRESENTAZIONE,
GESTIONE FILE ...

FORMATTAZIONE

RISULTATI DELLE ANALISI
(VALORI, GRAFICI,
TABELLE ...)

+ CONTENUTI

OPERAZIONI LOGICHE,
PROGRAMMAZIONE ...

+ CONTROLLO

OUTPUT DI STATA IN MARK-UP LANGUAGES:

CASE STUDIES (AGENZIA REGIONALE SANITA' TOSCANA)

1. PUBBLICAZIONE DATI ANZIANI SU WEB

Dati da SDO, RMR, SPA, anni 1999-2002, aggregati per ASL (12). Partendo da una tabella creata con una cascata di query SQL, il programma di stata genera: per ciascuno dei 15 indicatori prodotti, 8 grafici e 5 tabelle (120 GRAFICI e 40 TABELLE), per ogni aggregazione temporale e geografica (2 anni X 12 ASL), e li salva in una albero di directory per un file html. Ogni anno la procedura popola il sito con i nuovi dati e genera la struttura del sito html.

2. ANALISI DI QUESTIONARIO SUI COMPORAMENTI DI ACCUDIMENTO DEI NEONATI

Ogni mese un programma .do produce un output che presenta in forma tabellare e grafica tutte le descrittive, alcune bivariante e due modelli multivariati, in un unico documento cartaceo e in una presentazione a schermo (LaTeX).

3. INDAGINE ISTAT "MULTISCOPO": accodamento archivio anni 1993 - 2002

Il tracciato record dell'indagine multiscopo ISTAT di ciascun anno viene esaminato (file read), trasformato in archivio e confrontato con gli archivi degli anni precedenti, producendo come output:

- 1) due tabelle in formato csv che chiedono l'intervento dell'utente sui casi di accodamento dubbi
- 2) procedura ORACLE che accoda correttamente le nuove variabili alle variabili degli anni precedenti
- 3) un ipertesto di documentazione dell'archivio accodato

Conclusioni

Risorse

Problematiche

Problematiche: dettagli di supporto

Problematiche: indicazione fondamentale

- ▶ Per poter sfruttare le tecniche proposte, è necessario conoscere il linguaggio scelto per la formattazione dell'output.

- ▶ Per poter sfruttare le tecniche proposte, è necessario conoscere il linguaggio scelto per la formattazione dell'output.
- ▶ Sia per \LaTeX che per \HTML , grazie alle rispettive comunità di supporto e sviluppo, è possibile reperire sul *web* una notevole quantità di risorse e di documentazione.

- ▶ Questa modalità di utilizzo di *Stata* presenta moltissime opportunità;

- ▶ Questa modalità di utilizzo di *Stata* presenta moltissime opportunità;
- ▶ per supportarla tuttavia sarebbero utili alcune indicazioni.

Gestione dei formati grafici

Sarebbe utile una gestione maggiormente completa dei formati grafici (pdf, png)

Gestione trasparente ODBC

- ▶ Tutti i comandi di *Stata* possono essere dati completamente via linea di comando

Gestione trasparente ODBC

- ▶ Tutti i comandi di *Stata* possono essere dati completamente via linea di comando
- ▶ **tranne la suite odbc!**

Gestione trasparente ODBC

- ▶ Tutti i comandi di *Stata* possono essere dati completamente via linea di comando
- ▶ tranne la suite odbc!
- ▶ ...sarebbe invece molto utile poter accedere alle sorgenti ODBC direttamente da un file do.

Indicazione fondamentale

- ▶ Non è prassi universale dei programmatori di *Stata* rendere disponibile in macro le informazioni necessarie a ricostruire l'output completo dei comandi!

Indicazione fondamentale

- ▶ Non è prassi universale dei programmatori di *Stata* rendere disponibile in macro le informazioni necessarie a ricostruire l'output completo dei comandi!
- ▶ Questo costringe a modificare il codice dei comandi, o a ricorrere ad algoritmi alternativi e più complessi.

Indicazione fondamentale

Riteniamo sarebbe utile che la comunità di supporto e sviluppo di *Stata* divenisse consapevole di potenzialità di questa natura, e che l'abitudine a lasciare disponibili in macro i risultati dei calcoli, almeno a livello di opzione, si diffondesse in modo sistematico.