

# **Automated Access to 100.000.000 Statistical Facts via Statline4 Web Services**

***Olav ten Bosch***

***Statistics Netherlands***

**UN-ECE conference, Bratislava**

**April 2005**



# Contents

- **StatLine in a Nutshell**
- **Automated Access to Statistics: Why?**
- **Design of StatLine4 Web Services**
- **The SN SODI Pilot**
- **Conclusions**



# StatLine in a Nutshell (1)

## - StatLine:

- a statistical output database accessible via Internet
- about 100 Million statistical facts
- Multidimensional cubes with hierarchical dimensions organized by theme
- Search facility
- Tables, charts and maps
- DUAL = refer to any table, map or chart in StatLine within one URL
  - (example: give the 10 most recent statistical facts on subject x)



# StatLine in a Nutshell (2)

The screenshot shows the CBS StatLine website in Microsoft Internet Explorer. The browser's address bar displays the URL: <http://statline.cbs.nl/StatWeb/Start.asp?p=Search/Search&LA=EN&DM=SLEN>. The website header includes the Statistics Netherlands logo and the StatLine logo with the Dutch flag.

### Consumer price index (CPI) all households, 2000=100

CPI: 00000 Total expenditure

Subjects	CPI	Derived CPI
Periods	2000=100	
2000	100.0	100.0
2001	104.2	103.1
2002	107.6	106.4
2003	109.9	108.4
2004	111.2	109.4

© Statistics Netherlands, Voorburg/Heerlen 4/10/2005

The Statline Webselector window is overlaid on the bottom right. It shows a hierarchical tree of data groups under 'Main groups'. The selected group is 'CPI; all households (2000=100)'. The 'Selection' panel shows the selected group and its sub-groups: 'Subjects (2)', 'CPI (259)', and 'Periods (5)'. The 'CPI' group is expanded, showing 'All categories (259)', 'All groups', 'Divisions (15)', 'Groups (40)', and 'Classes (95)'. The 'Show Data' button is visible at the bottom of the window.

Statline Webselector

Main groups

- Macro economy, money and capital markets
  - National accounts
  - Economic data
  - Money and capital markets
  - Prices
    - Consumer prices
      - CPI; all households (2000=100)
      - CPI; househ. of wage earners

Selection

CPI; all households (2000=100) (2590)

Subjects (2) | CPI (259) | Periods (5)

- CPI
  - All categories (259)
  - All groups
    - Divisions (15)
    - Groups (40)
    - Classes (95)

Consumer price indices all households. Index figures for consumer goods. (base year 2000=100). 2000-2004 ; January 2000-March 2005. Changed on 7 April 2005.

Show Data | Help | Refresh

Java Applet Window



# StatLine in a Nutshell (3)

## StatLine4:

- Mechanisms for standardization and coordination of metadata
- Facilities to handle changes in metadata. Multiple versions of statistical facts
- complete redesign in .NET
- “smart” user interface
  
- Automated access through *Web Services*

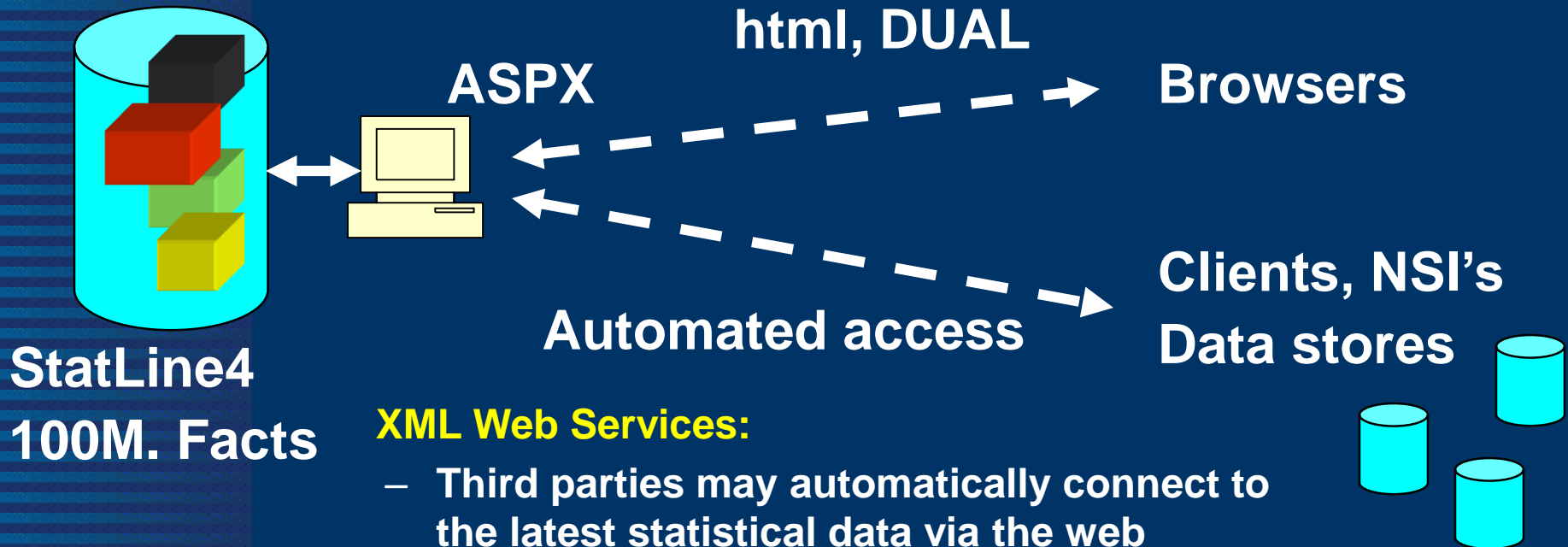


# Automated Access Why? (1)

- The Web evolves into a loosely coupled web of data repositories: statistics should be part of that.....
- We see clients linking deeply into our statistical database (deeplinking): we should make this easy.....
- Clients tend to query for certain specific statistical figures regularly: we should help them retrieve this content automatically.....
- NSI's and others institutes should be able to exchange statistical data automatically...



# StatLine4 Web Services (1)



StatLine4  
100M. Facts

## XML Web Services:

- Third parties may automatically connect to the latest statistical data via the web
- Service based on Industry standards & proven technology
- Web service is self-describing:
  - “what can I get here and how should I ask it”



# StatLine4 Web Services (2)

## - **Search Web Services**

- Inspired by Google API and uses Lucene
- Integrate StatLine search results into your own web page or application ([www.cbs.nl](http://www.cbs.nl))

## - **Update Web Services**

- RSS feeds on three levels:
  - Statistical Database as a whole (100 M. facts)
  - Within a specific Theme
  - Within a specific Cube

## - **Data Web Services**

- Facility to actually retrieve specific data sets
- or just one statistical fact
- Backwards compatible with current hyperlink mechanism (DUAL)



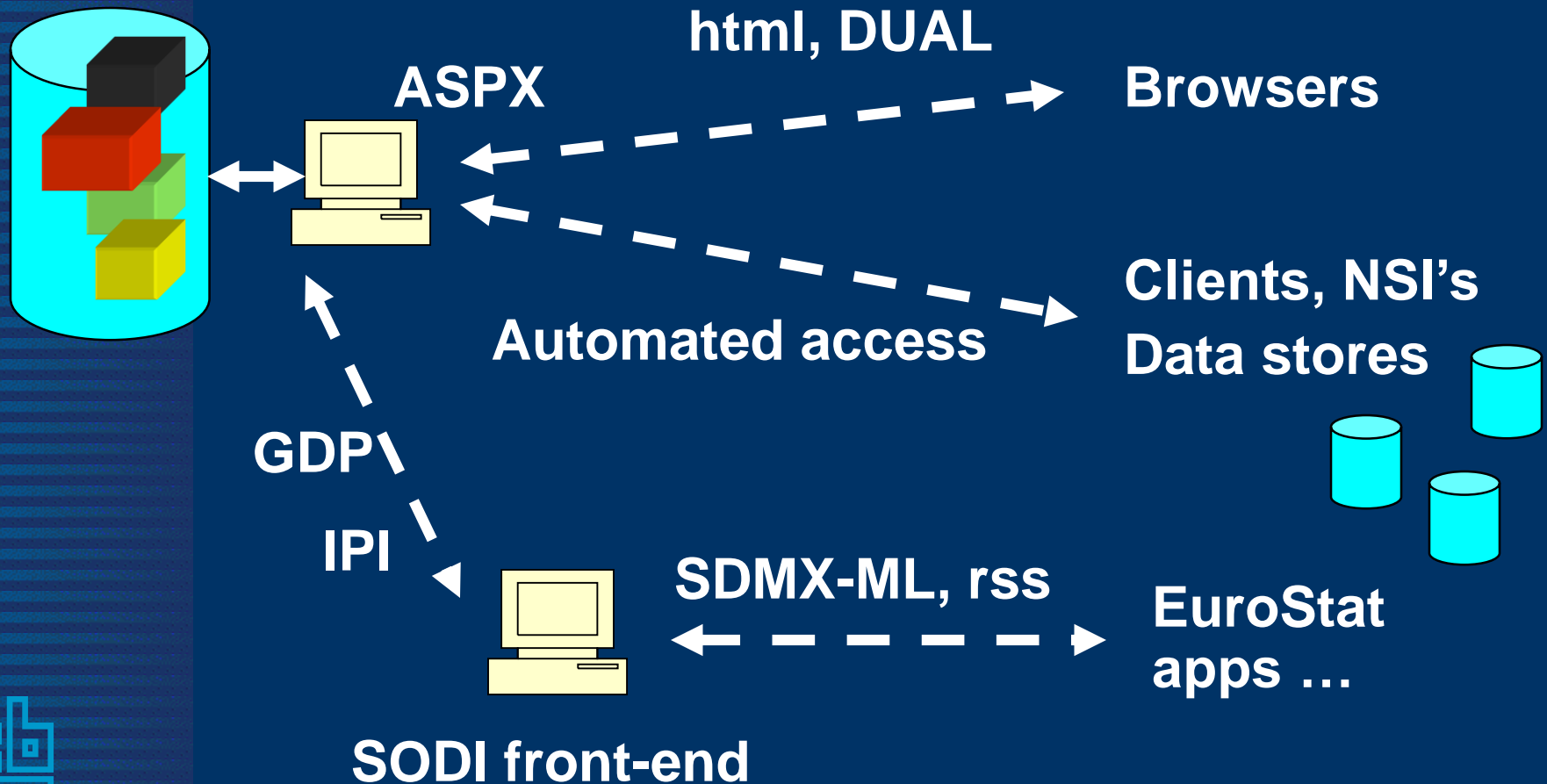


# The SN SODI Pilot (1)

- EuroStat SODI (SDMX Open Data Interchange) Task Force
- Pilot for:
  - Quarterly GDP (Quarterly Gross Domestic Product)
  - Monthly IPI (Industrial Production Index)
- SN added dedicated Web Services:
  - Return results in SDMX-ML format
  - RSS feed for updates
  - Technical: yet another interface to StatLine



# The SN SODI Pilot (2)



# The SN SODI Pilot (3)

CBS SODI Home page - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten Media

Adres <http://neon/sodi/> Ga naar Google

## CBS SODI Home page

### WebServices for the SODI Pilot.


SODI is short for Statistical Open Data Interchange, a EuroStat project to test the transformation of GESMES messages into SDMX (Statistical Data en Metadata eXchange, see [www.sdmx.org](http://www.sdmx.org)) messages. On this website, the contribution of Statistics Netherlands is presented.

Data available via StatLine ([statline.cbs.nl](http://statline.cbs.nl)) are transformed into SDMXMessage webservices.

In this pilot, only the data for the [Quarterly Gross Domestic Product](#) (Quarterly GDP) and the [Monthly Industrial Production Index](#) (Monthly IPI) are available.

---

### RSS

The following Rss link provides a list of updated tables for the Sodi-pilot 

---

### Quarterly Gross Domestic Product (GDP)

Here some predefined subsets of the Quarterly GDP Webservices are presented.

All	<a href="#">SodiWebservice.asmx/LoadAll?table=esa</a>
Total	<a href="#">SodiWebservice.asmx/LoadTotal?table=esa</a>
Only 2001	<a href="#">SodiWebService.asmx/LoadRange?table=esa&amp;StartTime=200101&amp;EndTime=200112</a>
Expenditure approach	<a href="#">SodiWebService.asmx/LoadSubSet?table=esa&amp;subSet=0102</a>
Income approach	<a href="#">SodiWebService.asmx/LoadSubSet?table=esa&amp;subSet=0103</a>
Total for 2001	<a href="#">SodiWebService.asmx/Load?table=esa&amp;subSet=tot&amp;startTime=200101&amp;endTime=200112</a>

---

### Monthly Industrial Production Index (IPI)

Here some predefined subsets of the Monthly IPI Webservices are presented.

All	<a href="#">SodiWebservice.asmx/LoadAll?table=sts</a>
Total	<a href="#">SodiWebservice.asmx/LoadSubSet?table=sts&amp;subSet=tot</a>
Last year	<a href="#">SodiWebService.asmx/LoadRange?table=sts&amp;StartTime=200401&amp;EndTime=200412</a>
Total for 2004	<a href="#">SodiWebService.asmx/Load?table=sts&amp;subSet=tot&amp;startTime=200401&amp;endTime=200412</a>

Lokaal intranet



# The SN SODI Pilot (4)

Sage - Sodi webservices (demo) - Mozilla Firefox

file:///C:/Documents%20and%20Settings/obos.RND/Application%20Data/Mozilla/Firefox/Profiles/qsqk

Sage

Sodi webservices (demo)

## Sodi webservices (demo)

Sodi webservices (this is a rss channel with fake publication dates)

- Monthly Industrial Production Index (Monthly IPI)**  
Monthly Industrial Production Index (Monthly IPI)  

---

In StatLine: **Production figures industry (excl. construction)**  
Fri, Apr 1, 2005 11:32
- Quarterly Gross Domestic Product (Quarterly GDP)**  
Quarterly Gross Domestic Product (Quarterly GDP)  

---

In StatLine: **Quarterly national accounts; values**  
Fri, Apr 1, 2005 11:32

Done

Partly Cloudy, 7°C



# The SN SODI Pilot (5)

```
http://neon/sodi/SodiWebService.asmx/Load?table=sts&subSet=tot&startTime=200401&endTime=200412 - Microsoft Internet Explorer
Bestand Bewerken Beeld Favorieten Extra Help
Vorige Zoeken Favorieten Media
Adres http://neon/sodi/SodiWebService.asmx/Load?table=sts&subSet=tot&startTime=200401&endTime=200412 Ga naar Google
<?xml version="1.0" encoding="utf-8" ?>
- <CompactData xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v1_0/message"
  xmlns:compact="http://www.SDMX.org/resources/SDMXML/schemas/v1_0/compact" xmlns:sts-compact="http://epp.eurostat.cec.eu.int/portal/structure/key_families/EUROSTAT_STS/compact"
  d1p1:schemaLocation="http://www.SDMX.org/resources/SDMXML/schemas/v1_0/message
  SDMXMessage.xsd http://www.SDMX.org/resources/SDMXML/schemas/v1_0/compact
  SDMXCompactData.xsd
  http://epp.eurostat.cec.eu.int/portal/structure/key_families/EUROSTAT_STS/compact
  EUROSTAT_STS_COMPACT.xsd" xmlns:d1p1="http://www.w3.org/2001/XMLSchema-instance">
+ <Header>
- <sts-compact:DataSet>
  <sts-compact:SiblingGroup TITLE="Index figures; production,sales,orders" UNIT="UNITS"
    UNIT_MULT="unspecified" DECIMALS="unspecified" TITLE_COMPL="Production figures industry
    (excl. construction)" SOURCE_PUB="70696" SOURCE_AGENCY="NL1" />
  - <sts-compact:Series FREQ="M" REF_AREA="NL" ADJUSTMENT="W" STS_INDICATOR="PROD"
    STS_ACTIVITY="NS0020" STS_INSTITUTION="1" TIME_FORMAT="P1M" COLLECTION="A"
    AVAILABILITY="A">
    <sts-compact:Obs TIME_PERIOD="2004-10" OBS_VALUE="98.5" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-11" OBS_VALUE="99.8" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-01" OBS_VALUE="99.4" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-03" OBS_VALUE="101.2" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-02" OBS_VALUE="101.2" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-05" OBS_VALUE="98.3" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-04" OBS_VALUE="97.3" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-07" OBS_VALUE="99.9" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-06" OBS_VALUE="101.1" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-09" OBS_VALUE="99.4" OBS_STATUS="A"
      OBS_CONF="F" />
    <sts-compact:Obs TIME_PERIOD="2004-08" OBS_VALUE="98.9" OBS_STATUS="A"
      OBS_CONF="F" />
  </sts-compact:Series>
+ <sts-compact:Series FREQ="M" REF_AREA="NL" ADJUSTMENT="N" STS_INDICATOR="PROD">
```



# Conclusions

## Designing a Statistical Web Service:

- Start simple; use examples
- Keep Web Services efficient and effective
- Support services on a detailed *and* a general level (example: information on updates)

## SL4 approach versus SODI approach:

- general web services (*all* statistics) versus dedicated web services (web service per domain)
- Easy general access versus complete and exact communication
- Both are necessary

## Web Service technology does *not* solve the metadata matching problem:

- Better metadata standards or semantic techniques





# Further Reading

- **Test page StatLine4**
  - <http://neon.vb.cbs.nl/statweb4/Intro/?LA=en>
  
- **SODI page of Statistics Netherlands:**
  - <http://neon.vb.cbs.nl/SODI>
  
- **SDMX page:**
  - <http://www.sdmx.org>

