Automated Access to 100.000.000 Statistical Facts via Statline4 Web Services

Olav ten Bosch

Statistics Netherlands

UN-ECE conference, Bratislava

April 2005

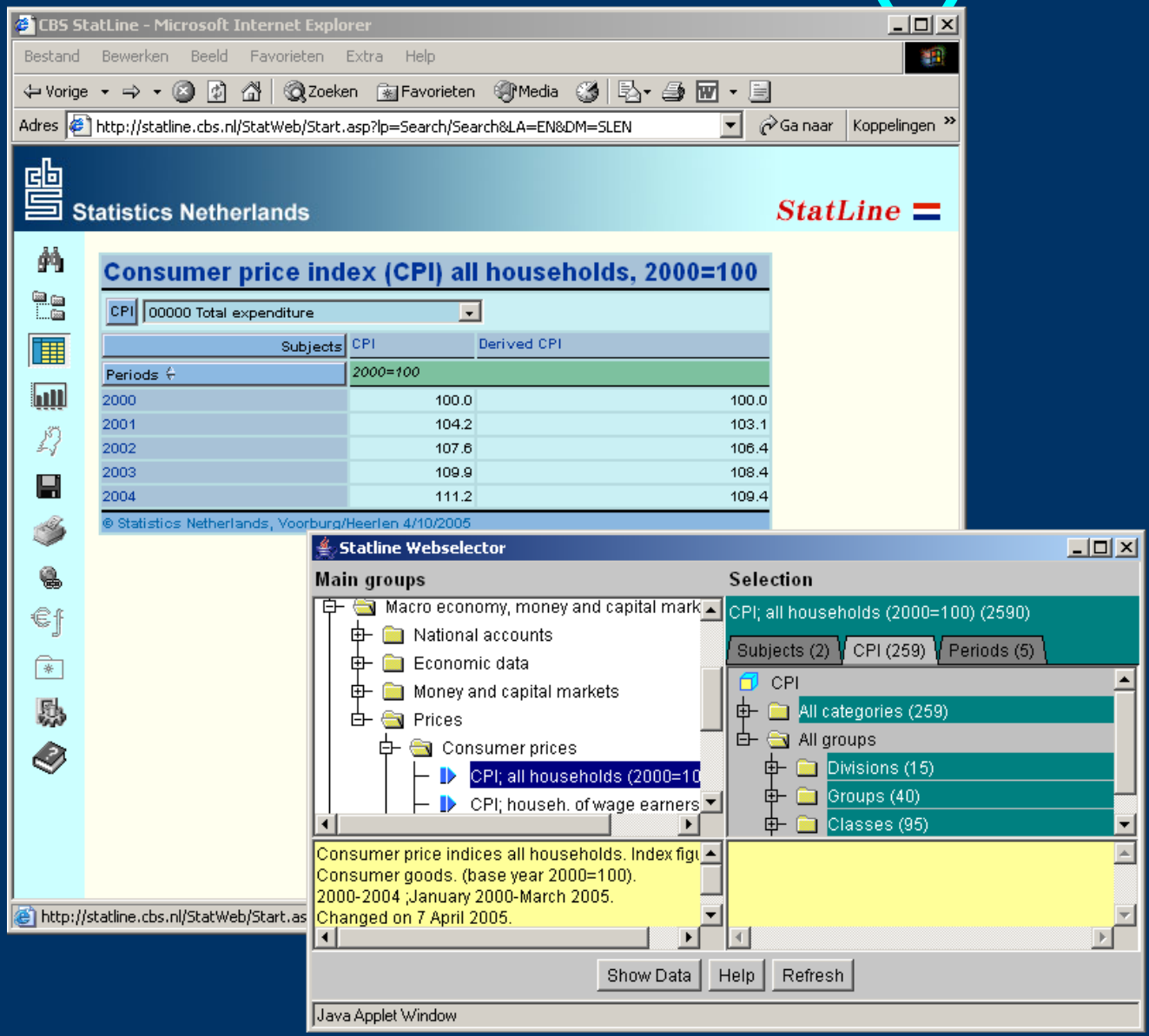
Contents

- 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)



StatLine in a Nutshell (3)

StatLine4:

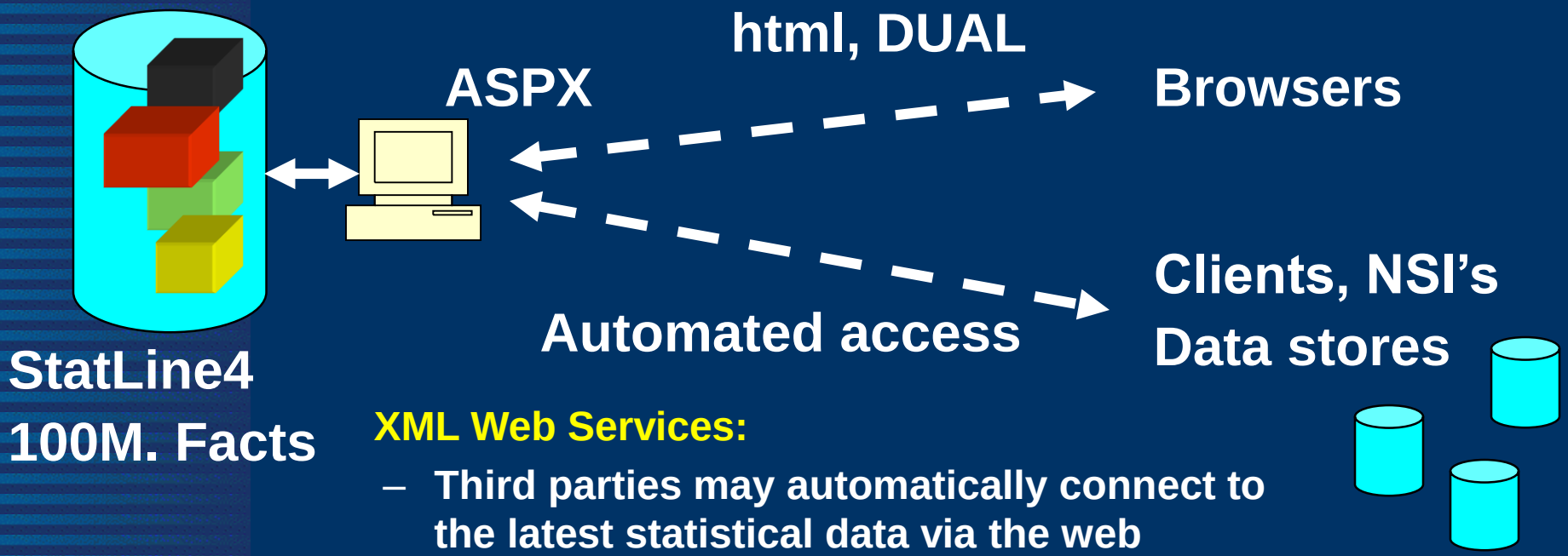
- **Mechanisms for standardization and coordination of metadata**
- **Facilities to handle changes in metadata. Multiple versions of statistical facts**
- **complete redesign in .NET**
- **“smart” user interface**

- **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)



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)

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

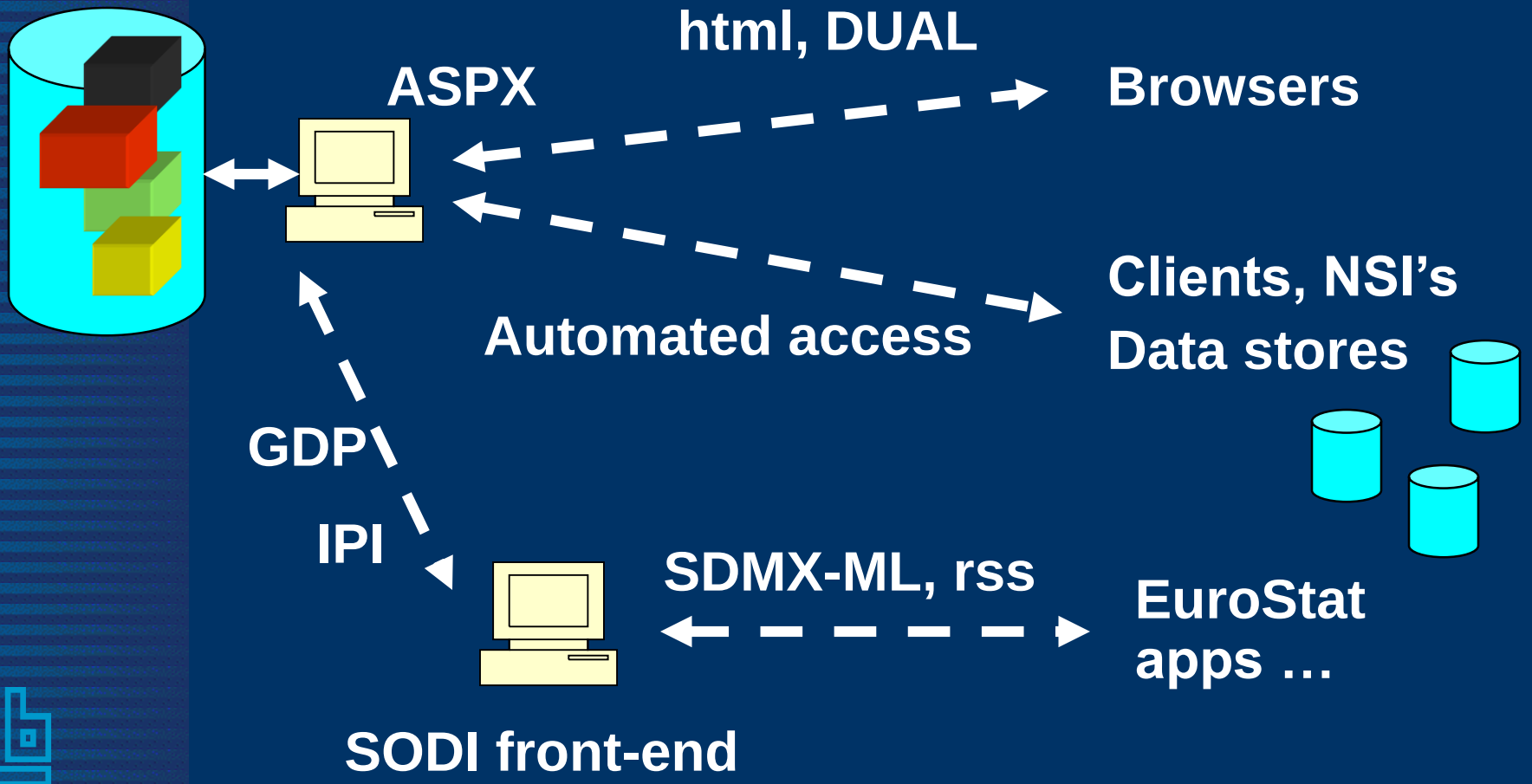
- ***Data* Web Services**
  - Facility to actually retrieve specific data sets
  - or just one statistical fact
  - Backwards compatible with current hyperlink mechanism (DUAL)

The SN SODI Pilot (1)

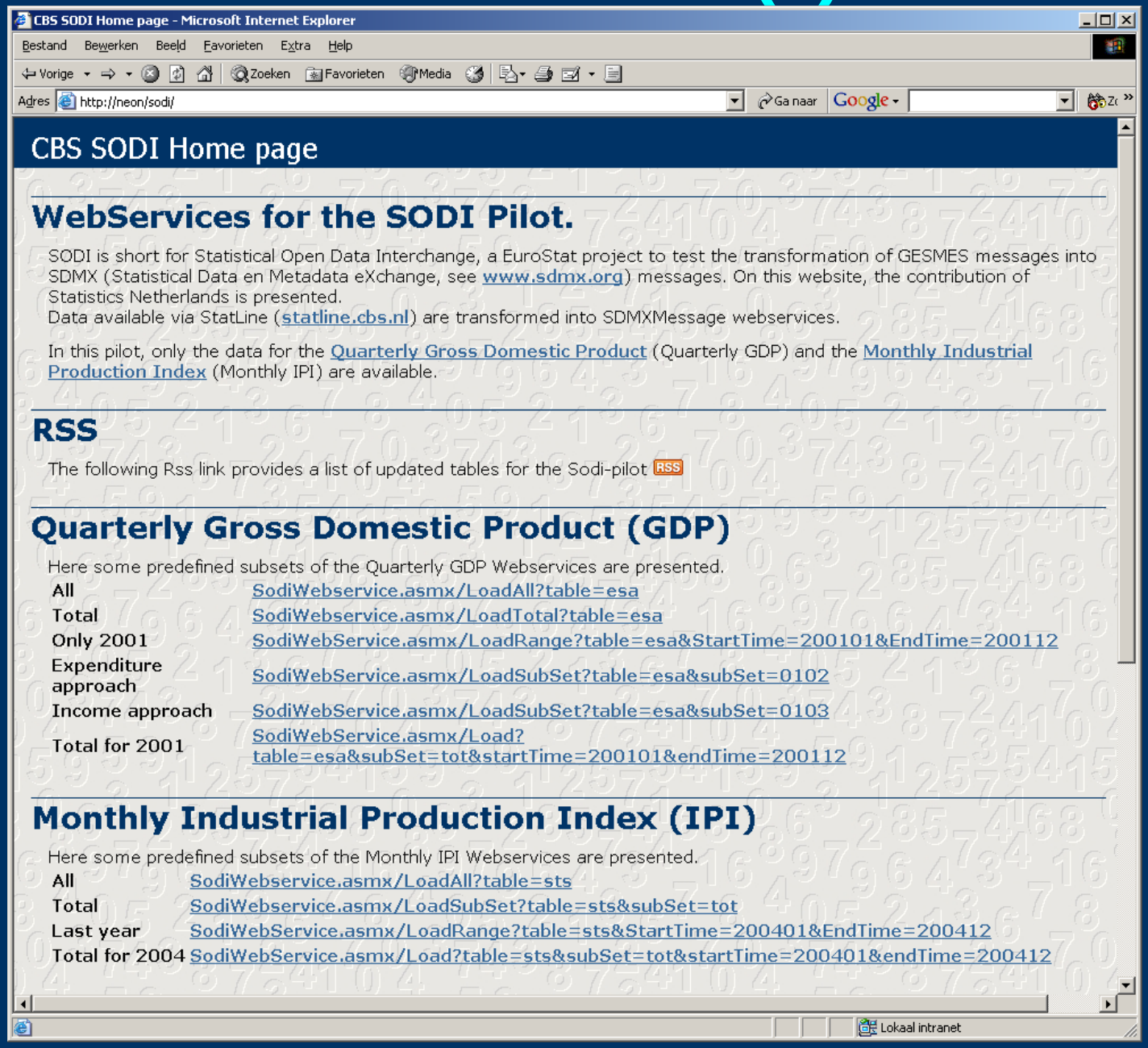
- EuroStat SODI (SDMX Open Data Interchange) Task Force
- Pilot for:
  - Quarterly GDP (Quarterly Gross Domestic Product)
  - Monthly IPI (Industrial Production Index)

- SN added dedicated Web Services:
  - Return results in SDMX-ML format
  - RSS feed for updates
  - Technical: yet another interface to StatLine

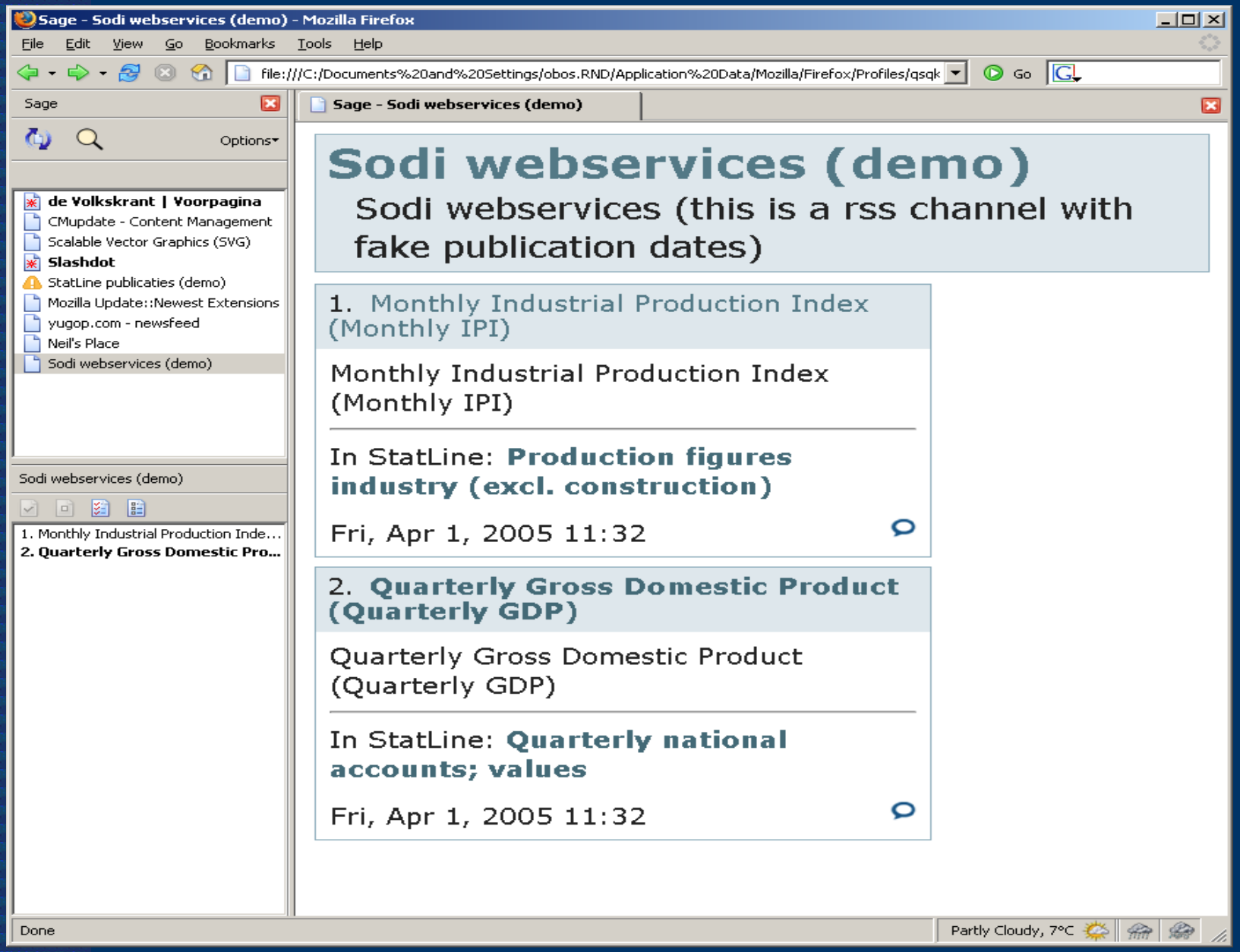
The SN SODI Pilot (2)



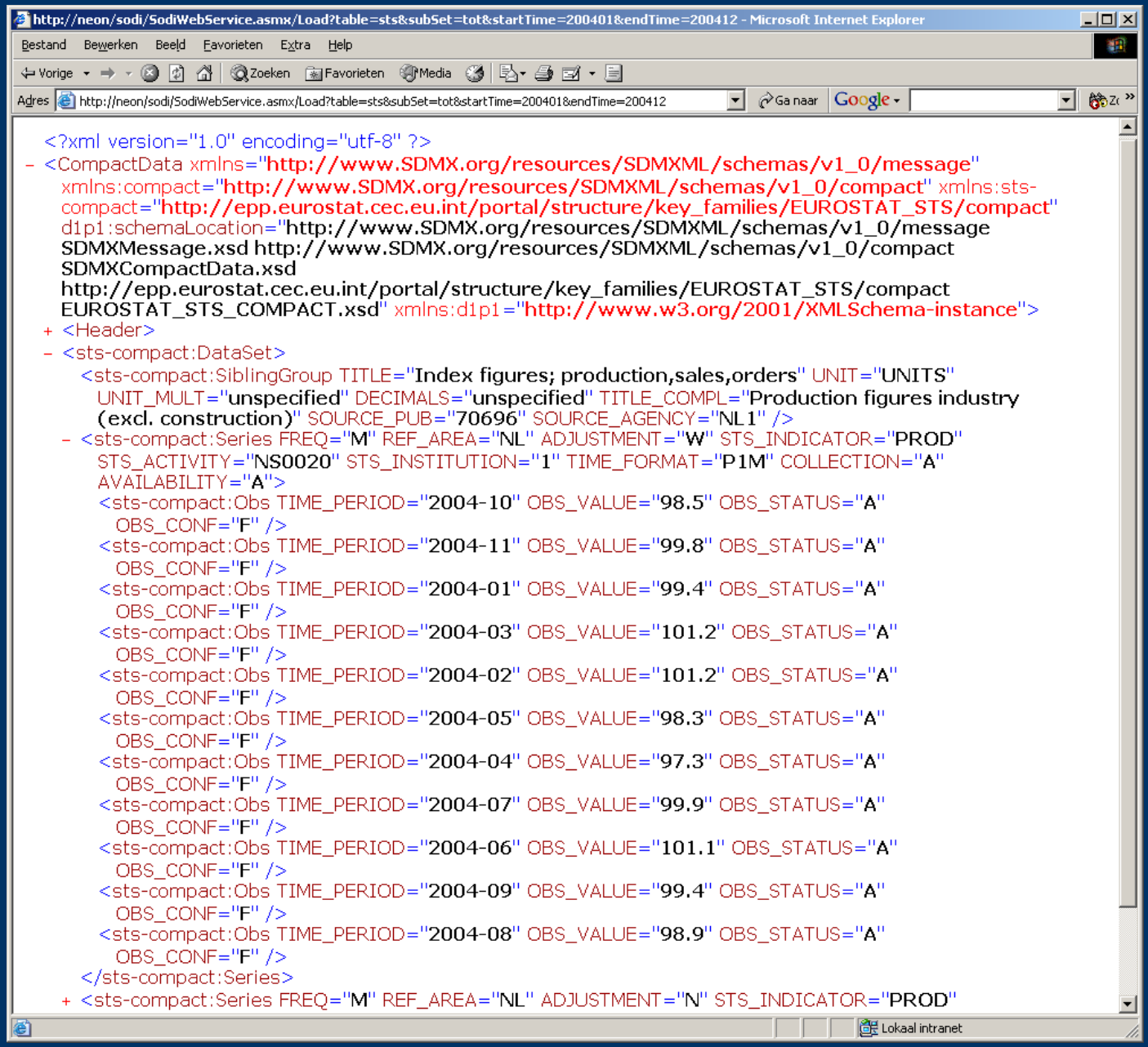
The SN SODI Pilot (3)



The SN SODI Pilot (4)



The SN SODI Pilot (5)



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