# D.3.3. Public databases

# The NewProt SSP portal revolves around data and software brought together in the SSP (Self Service Portal). In Deliverable 2.1 the access to publicly accessible data is described. The CMBI has for many years been working on the MRS project, and an old seminar on this topic is added as an appendix to D.2.1. This seminar is old, but the MRS concepts haven’t changed.

A series of snap-shots is added as appendix to this deliverable text to illustrate that the MRS-SSP coupling actually worked. As described in D.2.1, this coupling proved easier than expected, so that we are pleased to announce the delivery of D.3.3 six months ahead of schedule.

# **MRSan introduction**

## **M.L. Hekkelman, G. Vriend CMBI 2006**

# **Three parts**

## Automatic data gathering and updating

## Indexing of data. Mostly, but not limited to, flat file and XML

## Searching the data and presentation of results

# **Design goals**

## Avoid errors by keeping it simple

## Fault tolerant, show must go on

## Platform independent

## Minimal hardware requirements

# **Data updating**

## Based on makefiles

## Minimal configuration

## Continues after an error

## Clear reporting

# **Indexing**

## Creates one big file containing data and indices

### No more out of sync situation

### Very clear if indexing succeeded or failed

### Atomic replacement of production file

### File is completely standalone without any dependency

## Creates in one go, or by splitting work over cluster nodes and merge results

## Can deal with update databanks and join databanks

## Uses Perl plug-in scripts to parse the flat files

# **Searching**

## Fast with lots of headroom to become even faster

## Boolean queries with optional operators and wildcards

## Future version will contain ranked queries

## Results are pretty printed by Perl plug-in scripts

## Blast built in, blast against the results of a Boolean query

# **Wrap up**

## MRS is a library with several front-ends (Perl/Java/C++)

## Flexible, fast and with lots of potential

## Has been in production for several years at the CMBI

# **Future developments**

## Create a distribution with documentation

## Extend the Web Services interface using gSoap

## Distributed data and queries