



## A short presentation of EPGRIS,

at the N.I. Vavilov Institute of Plant Industry

by Dag Terje Endresen, The Nordic Gene Bank.

The three year EU founded project EPGRIS (European Plant Genetic Resources Information-Structure) aim to produce an European PGR Search Catalog (EURISCO) including passport data of collections maintained *ex situ* in Europe. The project started in October 2000. The participating countries are divided in four regions: north, south, east and west. Region 1 include Denmark, Estonia, Finland, Germany, Iceland, Latvia, Lithuania, Norway, Poland, Russia and Sweden. The Nordic Gene Bank is the coordinator in region 1.

Today there are several European search cataloges for different crops within the ECP/GR networks. One of the motivations behind the EPGRIS project is to simplify the data flow. The data flow for the Central Crop Data Bases (CCDB) and for EURISCO is illustrated below in figure 1 and figure 2.

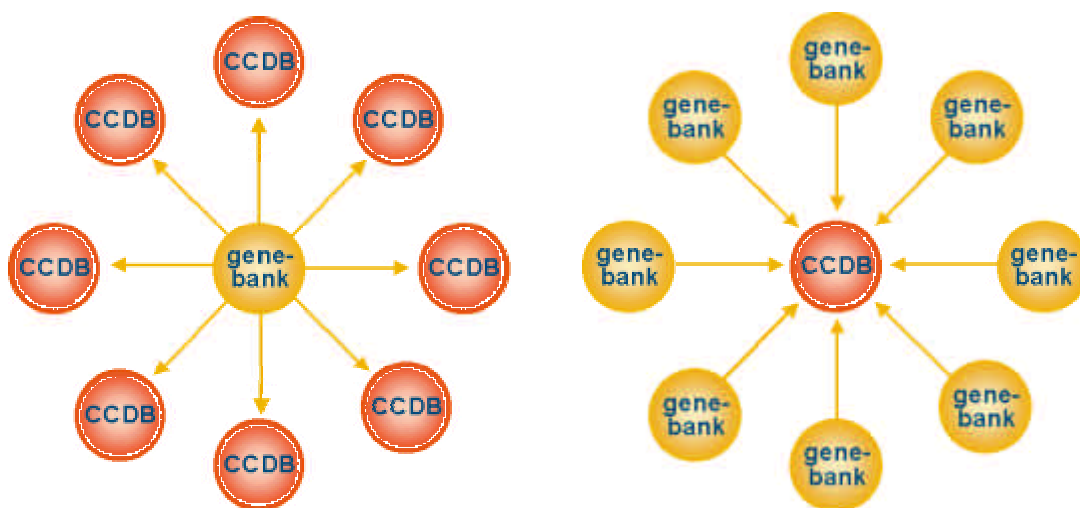


Figure 1. Information flow for the central crop databases today. Each gene bank upload data to many Central Crop Data Bases (CCDB). Each Central Crop Data Base receive data sets from many gene banks.



Figure 2. Information flow with EURISCO. National inventories hold all PGR data set of national gene banks and upload the data set directly to the multi crop search catalog EURISCO.

Each of the regions are expected to have a separate regional process to produce their subset of the passport data to be included in EURISCO. The methods of collecting passport data could be different for each country. Some gene banks don't have a permanent internet connection and must actively upload their data set, while others are permanently online and may provide their data set online. EPGRIS will use the IPGRI/FAO Multi-Crop Passport Descriptor List (MCPDL) for data exchange. Some gene banks will transform the data from their documentation system into the MCPDL themselves, others will wish to make their 'raw' local data set available (together with a decoding table) and leave for the regional coordinator to transform the data to the MCPDL format. The goal is however to have an automatic procedure where the national inventories present their data directly to EURISCO in the standardised format.

EPGRIS will initially use only the flat ascii file for file exchange. For the Nordic countries and especially Russia the ascii file will be insufficient as national characters will need to be transliterated. For the collection of passport data in region 1 we have agreed to accept different character sets (as long as the used character set is specified). The data sets will next be combined and stored in the unicode character set before transliterated and submitted to EURISCO. EPGRIS will need to solve the character set 'problem' and when it does region 1 will this way already have routines to provide correct characters.

For data exchange preferably the flat text file should be used. In region 1 it is agreed to temporarily accept even other file formats like dbf and excel to get 'fresher' and more complete data faster. Many gene banks in the EPGRIS region 1 still keep their original data in these formats. To produce a flat text file would in these cases mean manual work. NGB will produce automatic or 'half'-automatic script routines for the transformation to text files coded in unicode (and ascii) to reduce the 'manual' work involved.

The flat ascii text file for data exchange is very sensitive to 'illegal' characters. One misplaced carriage return or tab could ruin the record or even possibly all the data below the 'illegal' character. One solution (or improvement) suggested by Germany is to use XML (Extensible Markup Language). XML is spreading in use to be one the standard data exchange formats on the internet. This will probably eventually be the most common exchange format also for EURISCO. XML is designed to make it easy to interchange structured documents over the Internet. For gene bank passport data an accession could in XML terminology be an 'entity' or 'object' and the descriptors 'attributes' or 'properties'. To allow the computer to check the structure of a data set a Document Type Definition (DTD) could declare each of the permitted entities, elements and attributes, and the relationships between them.

The common search catalog (EURISCO) will be an important tool to improve collaboration between European gene banks. It will be easier to compare and combine data sets for analysis, extracting interesting PGR material or to find possible duplicates. And a search catalog of passport data is a good starting point for adding more descriptive data to the common 'data-pool' or for linking from the search catalog to more detailed information provided online by the individual gene banks or national inventories. Another interesting task could be to link EURISCO to other large PGR search catalogs like GRIN or SINGER.

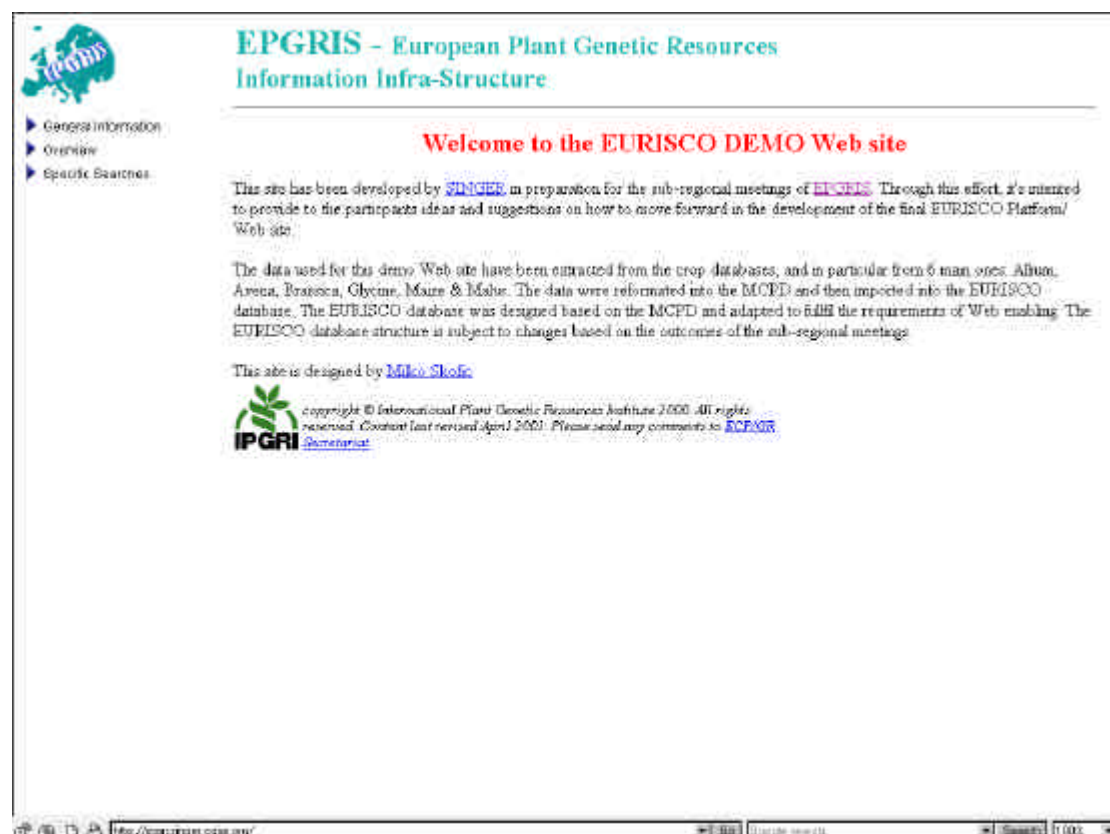


Figure 3. Demo of the search catalog has been published at <http://ipgri.singer.cgiar.org>

A first demo of the EURISCO search catalog was prepared for the sub-regional meetings and may still be accessed at <http://ipgri.singer.cgiar.org>, figure 3. NGB has an example on how the transliteration of characters might work demonstrated online

at [http://www.ngb.se/epgris/epgris\\_test2.php](http://www.ngb.se/epgris/epgris_test2.php). For the page to work your browser must of course have support for unicode for the 'No transliteration' option and Latin 1 for the 'Latin-1' option. MS Windows 2000, MS Windows NT 4, MS Windows XP and Linux systems all have support for unicode. Even older systems like MS Windows 95 and MS Windows 98 will show the text properly with unicode-aware software such as Netscape, Internet Explorer, or Outlook Express, to name a few. You also need a font with all the characters represented. Today a modern computer system will only miss the very most exotic characters and all the European languages should be presented fine. See figure 4 and figure 5. The example use the UTF-8 unicode schema. Another page show online export of the complete or a selection of the NGB *ex situ* collection to the IPGRI/FAO Multi-Crop Passport Descriptors. See figure 6. For more information on the EPGRIS project see the home page at <http://www.ecpgr.cgiar.org/epgris/>

### EPGRIS Test Unicode/Transliterated Output

---

Table	Transliteration
<input type="text" value="Agropyron (VIR)"/>	<input type="text" value="No transliteration"/> <input type="button" value="Search"/>

Note: Tables included are i 'raw' format, i.e. as they were submitted by data provider.  
Conversion to common format comes later ... Besides that, this is only a test so don't use this data.

---

Your query gave 172 results:

**SciCod, CulNam, OriReg, OriRem**

[AGROCRIS, N 447, КАЗАХСТАН, УРАЛЬСКАЯ ОБЛ.](#)  
[AGROCRIS, N 477, КАЗАХСТАН, УРАЛЬСКАЯ ОБЛ.](#)  
[AGROCRIS, N 637-57, УРАЛЬСКАЯ.](#)  
[AGROCRIS, N 694, КАЗАХСТАН, УРАЛЬСКАЯ ОБЛ.](#)  
[AGROCRIS, БАСОВСКИЙ ЛУГОВО, СТАВРОПОЛЬСКИЙ.](#)  
[AGROCRIS, БРОДСКИЙ ШИРОКОК, ОРЕНБУРГСКАЯ.](#)  
 AGROCRIS Л.

Figure 4. An example of search result presented without transliteration in Unicode.  
[http://www.ngb.se/epgris/epgris\\_test2.php](http://www.ngb.se/epgris/epgris_test2.php)

### EPGRIS Test Unicode/Transliterated Output

---

Table	Transliteration
<input type="text" value="Agropyron (VIR)"/>	<input type="text" value="Latin-1"/> <input type="button" value="Search"/>

Note: Tables included are i 'raw' format, i.e. as they were submitted by data provider.  
Conversion to common format comes later ... Besides that, this is only a test so don't use this data.

---

Your query gave 172 results:

**SciCod, CulNam, OriReg, OriRem**

[AGROCRIS, N 447, KAZAXSTAN, URAL'SKAYa OBL.](#)  
[AGROCRIS, N 477, KAZAXSTAN, URAL'SKAYa OBL.](#)  
[AGROCRIS, N 637-57, URAL'SKAYa.](#)  
[AGROCRIS, N 694, KAZAXSTAN, URAL'SKAYa OBL.](#)  
[AGROCRIS, BASOVSKIJ LUGOVO, STAVROPOL'SKIJ.](#)  
[AGROCRIS, BRODSKIJ SHIROKOK, ORENBURGSKAYa.](#)  
 AGROCRIS D.

Figure 5. An example of search result presented with transliteration to Latin 1.  
[http://www.ngb.se/epgris/epgris\\_test2.php](http://www.ngb.se/epgris/epgris_test2.php)

Export to Multicrop passport descriptor list (MCPDL) search

Family name =

Genus name =

Scientific name =

Country of origin =

Genus	Species	Subtaxa	Authority	Accenumb	Accename	Origcty
gennam	spenam	subnam	autnam	accnum	accnam	oricou
Avena	sativa		L.	4758	TULUNSKII	RUS
Avena	sativa		L.	5129	VIDVIGENEZ	RUS
Avena	sativa		L.	6362	HIBINI II	RUS
Avena	sativa		L.	6363	VOLGANENSIS	RUS
Avena	sativa		L.	8772	CLAV 2424; KVL 1388	RUS
Avena	sativa		L.	9811	MOSKOVSKII A. 315	RUS

Total hits 6

Select preferred format for the full report of selected NGB accessions in MCPDL format below

report as xls     
  report as html     
  report as tab separated

Figure 6. The Nordic Gene Bank has online a tool for export of all or a selection of the NGB collection to the multicrop passport descriptor list format at <http://www.ngb.se/epgris/mcpdl/>

References:

- \* EPGRIS home page, <http://www.ecpgr.cgiar.org/epgris>
- \* EURISCO DEMO Web site, <http://ipgri.singer.cgiar.org>
- \* EPGRIS Demo at the Nordic Gene Bank, <http://www.ngb.se/epgris>
- \* SINGER, <http://www.singer.cgiar.org>
- \* GRIN, <http://www.ars-grin.gov/npgs>
- \* EPGRIS Power point presentation, (unpublished, filename: EPGRIS.ppt, 26 July 2001, 419 kB)