

Integration of Biological Resources Information: CABRI & EBRCN projects

Paolo Romano
National Cancer Research Institute, Genova
(pao.lo.romano@istge.it)

Summary

- Some ideas on information integration in biology
- CABRI: a “one stop shop” for biological resources
- EBRCN: interconnected biological resources database
- Future developments

Evolution of information management

- Isolated, local data storage, mainframes
- Shared, local data storage, LAN
- Decentralized, not shared data storage
- Decentralized, shared data storage → integration
- The network as a local archive, although decentralized
→ data GRID
- Which integration?

Degrees of information integration

- Tightly integrated systems
 - Data: local warehouse
 - Applications: centralized or CORBA
 - Processes: static, repetitive services
 - Integration: early or predefined
 - Transparency: high
- Dynamicly (loosely) integrated systems
 - Data: decentralized, dynamic integration
 - Applications: Web Services
 - Processes: dynamic, based on users' requirements
 - Integration: on demand or data mining
 - Transparency: medium to low (interaction)

Integration longevity

- Integration needs stability
 - Standardization.....
 - Good domain knowledge
 - Well defined data
 - Well defined goals
- Integration fears:
 - Heterogeneity of data and systems
 - Uncertain domain knowledge
 - Fast evolution of data
 - Highly specialized data
 - Lacking of predefined, clear goals
 - Originality, experimentalism (“*let me see if this works*”)

Information in biology

- Biomedical research produces an increasing quantity of new information
- Some domains, like genomics and proteomics, contributes to huge databases
- Emerging domains, like mutation and variation analysis, polymorphisms, metabolism, and technologies, e.g., microarrays, will contribute with even huger amounts of data

Information in biology

- EMBL Data Library 74 (Mar 2003):
 - Sequences: 23.234.788, Bases: 30.356.786.718
 - Size: ~103 Gbytes, 8,8% more over 73
- GenBank Release 125.0 (Aug 2001):
 - Sequences: 12.813.516, Bases: 13.543.364.296
 - Size: 49,72 GBytes

Heterogeneity of databanks

- Only a few databanks are managed in an almost homogenous way by EBI, NCBI, DDBJ (sequence)
- Many databases are created by small groups or single researchers
- Secondary databases are of high quality (good and extended annotation, quality control)
- Many databases are highly specialized, e.g. by gene, organism, disease, mutation, etc...

Data banks are distributed

- Distributed data banks means:
 - Different DBMS
 - Different data structures
 - Different information
 - Different meanings
 - Different data distribution methods

Softwares

- Specialist softwares are essential for almost all analysis in molecular biology:
 - Sequence analysis, secondary and tertiary protein structure prediction, gene prediction, molecular evolution, etc...
- Softwares must be integrated with databases
 - Databases as input for softwares
 - Results as new data to record and analyze

Goals of the integration

- Integration is needed in order to:
 - Achieve a better and wider view of all available information
 - Carry out analysis and/or searches involving more databases and softwares in one step only
 - Carry out a real data mining

Integration of databanks

- Integration of databanks implies:
 - Accurate analysis and definition of involved “biological objects”
 - Analysis of available information / data
 - Identification of logical links between objects and and definition of related data links between dbs
 - Definition and implementation of common data interchange formats, methods, tools

Integration of biological information

In biology:

- o Goals and needs of researchers evolve very quickly according to new theories and discoveries
- o A pre-analysis and reorganization of the data is very difficult, because data and related knowledge vary continuously
- o Complexity of information makes it difficult to design data models which can be valid for different domains and over time

Integration methods

- Explicit (reciprocal) links (xrefs)
- Implicit links (e.g., names)
- Common contents (vocabularies)
- Object oriented models
- Relational schemas
- Ontologies

Explicit links

- From one record to one related record in a different database:
 - Direct link, unique, not reciprocal
 - Remote ID
 - Standard formats
 - Life Science ID
 - Standard I3C
- Limitations:
 - Manual annotation required
 - Meaning of the link is undefined
 - Need to be predefined, not dynamically defined

Common contents

- From one record to many records in a different database, through a text search:
 - Implicit link, reciprocal
 - Automatically determined
 - Vocabulary terms
 - Standard vocabularies
- Limitations:
 - Use of vocabularies in manual annotation
 - Meaning of the link is undefined
 - Scope definition needed
 - Text mining

CABRI: Objectives

Common Access to Biological Resources and Information

- Setting Quality Management Guidelines
- Distributing biological resources of the highest quality
- Integrating searches and access to catalogues
- One-stop-shop for quality resources
- Ad hoc search (CABRI Simple Search)
- Shopping cart (pre-ordering facility)

CABRI: Partners and resources

Partners:

- INSERM (coordination)
- BCCM, CBS, DSMZ, ECACC, HGMP-RC, ICLC, NCCB (resources)
- HGMP-RC, IST, CERDIC (ICT)

Resources:

- Microorganisms (bacteria, yeasts, fungi)
- Cells (animal and human cell lines, hybridomas, HLA typed B lines)
- Plasmids, phages, viruses, DNA probes
- Overall, more than 90.000 items in catalogues

CABRI: Resources

	DP	B/A	F/Y	PL	PH	PC	PV	AC	HYB	BC
BCCM		X	X	X						
CABI		X	X							
CBS		X	X							
CIP		X								
DSMZ		X	X	X	X	X	X	X		
ECACC	X							X	X	X
ICLC								X		
NCCB		X		X	X					
NCIMB		X								

CABRI: why SRS

- Yes because
 - Manages heterogeneous databases
 - Flat file format
 - Simple and effective interface
 - Internal and external links
 - Link operator
 - Easily expandible (new databases)
 - Flexibility in creation of indexes

CABRI: why SRS

- No because
 - Local databases, not remote (updates)
 - “Difficult” language (Icarus)
 - Commercial software (not free)

CABRI: data structure

For each material, three data sets identified:

- Minimum Data Set (MDS): essential data, needed to identify individual resources
- Recommended Data Set (RDS): all data that are useful to describe individual resources
- Full Data Set (FDS): all data available on the resources

CABRI: data structure

For each information, data input and authentication guidelines, including:

- Detailed textual description of the information
- In-house reference lists of terms and controlled vocabularies
- Predefined syntaxes (e.g., Literature, scientific names)

CABRI: Data sets

Data set	Field label	Catalogues
MDS	Strain_number	All
MDS	Other_collection_numbers	All
MDS	Name	All
RDS	Race	All
MDS	Organism_type	All
MDS	Restrictons	All
MDS	Status	All
MDS	History	All
RDS	Misapplied_names	All
RDS	Substrate	All
RDS	Geographic_origin	All
RDS	Sexual_state	All
RDS	Mutant	All
FDS	Genotype	DSMZ
.....

CABRI: Name field

Field	Name
Description	<p>Full scientific and most recent name of the strain. It includes:</p> <ul style="list-style-type: none">▪ Genus name and species epithet▪ Subspecies▪ Pathovar▪ Authors of the name▪ Year of valid publication or validation▪ Approbation of the name
Input process	<p>Enter full scientific name as given by depositor and confirmed (or changed) by collection. Names of authors of the name, year of valid publication or validation and approbation are included after a comma.</p> <p>Values for approbation:</p> <p>AL = approved list, c.f.r. IJSB 1980 VL = validation list, in IJSB after 1980 VP = validly published, paper in IJSB after 1980</p> <p>Reference list: DSMZ list of bacterial names</p>
Required for	MDS

CABRI: Reference paper field

Field	Reference paper
Description	Original paper [if available]
Input process	<p>New entries: JournalTitle Year; Volume(issue): beginning page#-ending page#</p> <p>The title is abbreviated following international standard rules (ISSN). Abbreviations are without dot. Authors and title of the article are not mentioned.</p> <p>The reference can be followed by the Pubmed ID enclosed within square brackets as follows: [PMID: 1234567], where '1234567' is the Pubmed ID of the paper</p>
Required for	MDS

Strain_number LMG 1(t1)
Other_collection_numbers CCUG 34964;NCIB 12128
Restrictions Biohazard group 1
Organism_type Bacteria
Name *Phyllobacterium rubiacearum*, (ex Knsel 1962) Knsel 1984 VL
Infrasubspecific_names -
Status Type strain
History <- 1973, D.Knsel
Conditions_for_growth Medium 1, 25C
Form_of_supply Dried
Isolated_from Pavetta zimmermannia
Geographic_origin Germany, Stuttgart-Hohenheim
Remarks Stable colony type isolated from LMG 1. See also *Agrobacterium* sp.
LMG 1(t2)

Strain_number LMG 1(t2)
Other_collection_numbers -
Restrictions Either Biohazard group 1 or Biohazard group 2
Organism_type Bacteria
Name *Agrobacterium* sp.
Infrasubspecific_names -
Status -
Other_names *Phyllobacterium rubiacearum*, (ex Knsel 1962) Knsel 1984 VL
History <- D.Knsel (*Phyllobacterium rubiacearum*)
Conditions_for_growth Medium 16, 28C
Form_of_supply Dried
Isolated_from Pavetta zimmermannia
Geographic_origin Germany, Stuttgart-Hohenheim
Remarks One (t2) out of two stable colony types isolated from the original culture LMG 1.

CABRI: integration

For each catalogue:

- SRS and HTML links to reference dbs
(media, synonyms, hazard, etc...)

For each material:

- Common data structure and syntax
- Integrated searches/results through SRS

CABRI: Extra features

CABRI Simple Search:

- Search by ID(s), name(s), all other fields
- Search by name(s) with synonyms support

CABRI Shopping cart:

- Set of mixed javascripts and perl scripts
- Pre-order facility (email or fax)

CABRI: Simple Search

Synonyms' support

- Only allowed for micro-organisms
- Managed through a perl script
- First searched terms are matched against synonyms' reference dbs with getz
- When available, names are added to the initial search and a new search is carried out
- Results are then displayed and a link to synonyms' dbs is added

CABRI: Contents updating

Catalogues

- New catalogues from CIP, CABI Bioscience, NCIMB
- Catalogues updated 1 to 2 times a year
- Now 27 catalogues for more than 130.000 items

CABRI HyperCatalogue

- Goal: improved visibility from search engine
- A hypertext including a set of static HTML indexes by material and catalogue
- ca. 48,000 HTML files and ca. 92 Mbytes
- Flat files -> Relational DB -> HTML
- MySQL, perl + PHP
- Detailed description through SRS (shopping cart)

CABRI: Contents updating

SRS links on Other_collection_numbers

- Identification of duplications
- Identification and removal of errors
- Problems with multiple links between couples of libraries

Three links in place for each pair of catalogues (on SRS 6):

- from Strain_number to Other_collection_numbers (direct)
- from Other_collection_numbers to Strain_number (reverse)
- from Other_collection_numbers to Other_collection_numbers (ext)

Strains in CIP which are also in DSMZ bacteria collection:

```
(( cip_bact < ( cip_bact_el < dsmz_bact_el )) |  
  ( cip_bact < ( cip_bact_rl < dsmz_bact_rl ))) |  
  ( cip_bact < dsmz_bact ))
```

CABRI: Recent developments

CABRI Simple Search

- Results are first grouped by collection (ordering problem)

Current solution:

5. execute global query, including all catalogues
6. retrieve single hits in the session file user.par,
7. return single hits figures,
8. return available results for global query, upon request
9. execute single queries, upon request

CABRI: Recent developments

CABRI GlobalSearch

- Free text search of the whole site or subsections
- Based on ht://dig, soon available

CABRI CD ROM

- Version 1.05 (April 2003)
- Only includes the HyperCatalogue, external links for SRS

CABRI: mirrors

Main site <http://www.cabri.org/>

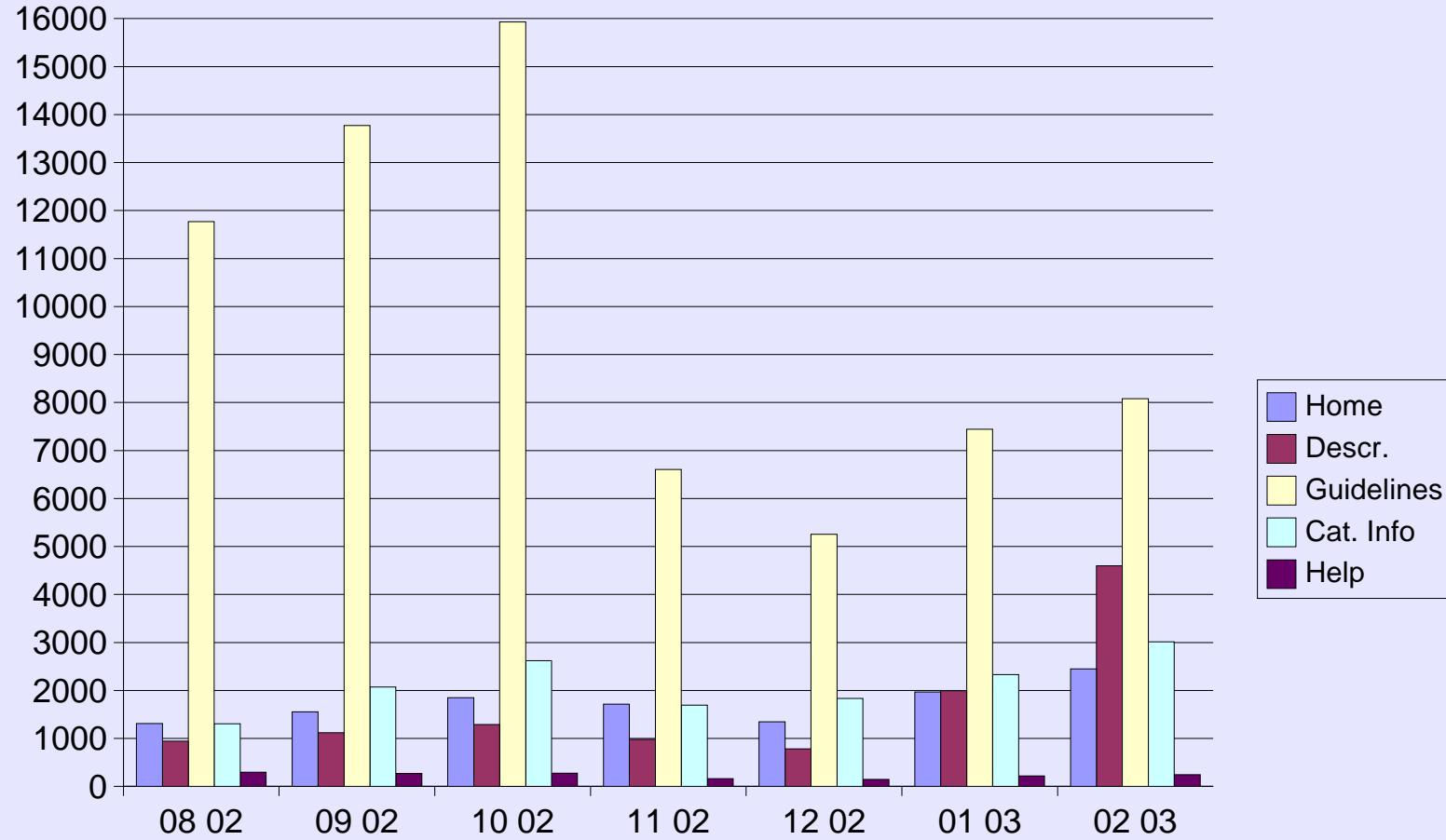
Four mirrors currently available

- Belgium: <http://www.be.cabri.org/>
- France: <http://www.fr.cabri.org/>
- Italy: <http://www.it.cabri.org/>
- China: <http://www.cn.cabri.org/>

Mirroring through rsync (password protected)

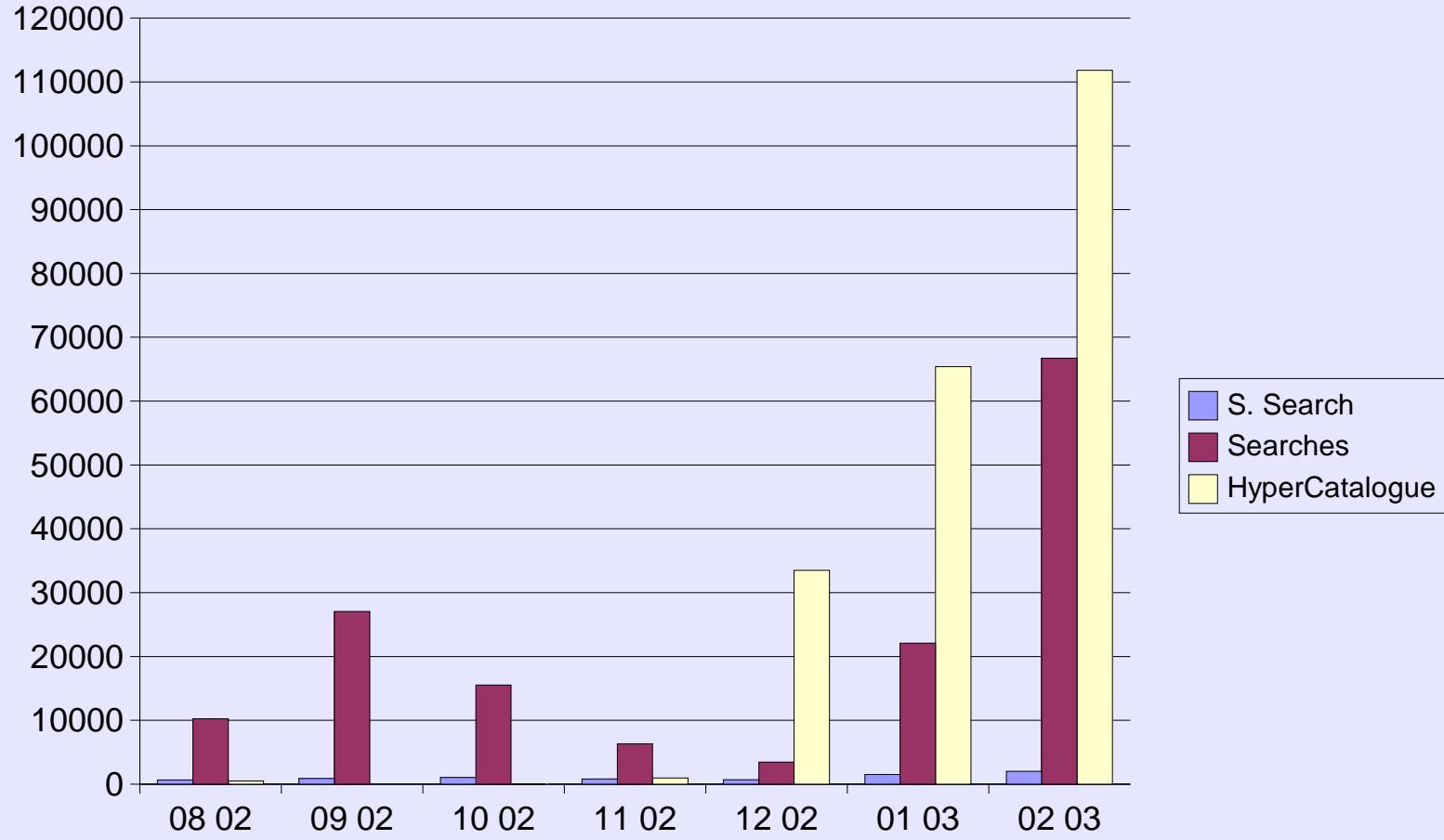
Analysis of Logs and Hits Count

Hits CABRI Home, Descr, Guidelines (ago 02-feb03)



Analysis of Logs and Hits Count

Hits Search, Simple Search and HyperCat (ago 02-feb03)



EBRCN: Extending integration

European Biological Resource Centres Network

Wp1 Co-ordinate European BRC policies, prepare a co-ordinated European response to international initiatives on biodiversity and become the European focal point for BRCs

Wp2 Develop new and maintain existing quality standards for European BRCs

Wp3 Establish a framework to maximise complementarity and minimise duplication among European BRCs

Wp4 Introduce new techniques in Information Technology to the EBRCN to add value to current catalogue information and enhance accessibility

Wp5 Collate and disseminate relevant information to the BRCs

EBRCN: Workpackage 4

Workpackage 4

“Introduce new techniques in information technology to the EBRCN to add value to current catalogue information and enhance accessibility”

Objective

Link catalogue data to literature, to nucleotide and to related genetic databases

EBRCN: new links

For all catalogues:

- Links to Medline through Pubmed ID
- Links to representative EMBL records

For selected catalogues:

- Links to plasmids' maps (plasmids)
- Links to microscope images (microorganisms)
- Links to other dbs under evaluation

Interconnected Biological Resource database

EBRCN: Linking to EMBL

- Test for linking to EMBL Data Library through SRS, without explicit IDs, gave negative results:
 - Links are different for different materials and can use various EMBL fields:
 - Organism (micro-organisms), Division (viruses and plasmids), Feature Table (definition of the source through Key, Qualifier, Description)
 - Annotation and indexing problems

EBRCN: EMBL links variability

- Annotation problems:
 - CBS 100.20 can be annotated as CBS 100.20 or CBS100.20
 - CBS 12345 can be annotated as CBS12345
- Indexing problems:
 - CBS 100.20 is indexed as CBS, 100 and 20
 - The dot is not included and is used as a separator
 - CABRI unique index key is “CBS 100.20”

EBRCN: Linking to EMBL (ii)

Examples of search:

- Query: Fungi & source & cbs 100.20

```
( ( ([emblrelease-FtKey:source] &
      [emblrelease-FtQualifier:strain] &
      ( ( [emblrelease-FtDescription:cbs] &
          [emblrelease-FtDescription:100] ) | 
          [emblrelease-FtDescription:cbs100] ) &
          [emblrelease-FtDescription:20]) )
< [emblrelease-Organism:fungi*] )
```

EBRCN: Linking to EMBL (iii)

A possible approach:

- Identify xrefs for linking from EMBL to CABRI catalogues, based on CABRI IDs
- A huge number of EMBL records could be linked to a single CABRI item
- Add links in EMBL and use these links when linking from CABRI (search by means of SRS)
- CABRI Ids included in EMBL data library and distributed with it

EBRCN: Linking to EMBL (iv)

- Agreement with EBI (list of xrefs)
- Work do be done after uploading CABRI catalogues to EBI
- Table of xrefs returned to collections
- Possible well known “wrong” EMBL sequence removed
- Links from plasmids catalogues to EMBL managed differently (using current remarks)

EBRCN: Extracted databases

- Extracted databases made available for SRS based sites in academic/no-profit Institutes
- Selected meaningful subset of information: MDS+link to main CABRI site
- FTP site with data and SRS syntax/structure files
- Agreement with EBI
- SPEAR ?

CABRI & EBRCN: what next?

Following SRS developments:

- SRS 5.1 -> SRS 7.1
- Flat file -> XML
- Other ??
- Can we abandon the simple search?

Adding contents:

- New catalogues (continuous effort)
- New materials
- New links to external dbs

CABRI & EBRCN: farer in time

New mixed architecture

- Decentralized catalogues/sites for bigger BRCs
- Centralized catalogues for smaller collections
- Unique interface (federated system?)

Evolution of biodiversity systems

- XML based tools
- Web Services

GBIF: Objectives

Global Biodiversity Information Facility (GBIF)

- Implementation of an international interlinked network for biodiversity information sources
- Central Secretariat and wide participation offered to governments and international agencies, through a Memorandum of understanding
- Initial focus on species- and specimens-level
- National nodes, organisms' nodes, etc...
- European Network for Biodiversity Information

GBIF: Tools

xml! xml! xml!

- Web Services as reference information source
- WSDL for describing Web Services
- UDDI for implementing registries of Web Services
- ABCD (Access to Biological Collection Data) schema as reference data structure

XML (eXtensible Markup Language)

- Markup language designed to add semantics in the text:
 - Overcome some HTML limitations
 - Simple definition and implementation (Document Type Definitions - DTDs)
 - Scalable, new DTDs based upon previous
- Feasible for software applications:
 - Well formed
 - Standardized vocabularies
 - Parsing and data extraction automatically performed

ID AA415057 standard; RNA; EST; 337 BP.
AC AA415057;
SV AA415057.1
DT 27-OCT-1997 (Rel. 53, Created)
DT 14-DEC-1999 (Rel. 62, Last updated, Version 2)
DE Mg0001 RCW Lambda Zap Express Library Pyricularia grisea cDNA clone RCW1
DE 5', mRNA sequence.
KW EST.
OS Magnaporthe grisea
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
RN [1]
RP 1-337
RA Wu S.-C., Bernstein B.D., Darvill A.G., Albersheim P.;
RT "Expressed sequence tags of the rice blast fungus grown on rice cell
RT walls";
RL Unpublished.
DR UNILIB; 863; 863.
CC Contact: Sheng-Cheng Wu
CC CCRC
CC University of Georgia
CC 220 Riverbend Road, Athens, GA 30602-4712, USA
CC Tel: 706 542 4446
CC Fax: 706 542 4412
CC Email: wusc@bscr.uga.edu
CC Seq primer: T3.

FH	Key	Location/Qualifiers
FH		
FT	source	1..337
FT		/db_xref="taxon:148305"
FT		/db_xref="UNILIB:863"
FT		/note="Vector: Lambda Zap; Messenger RNAs prepared from Magnaporthe grisea grown at 23C in the dark with constant gyratory shaking (100 rpm) in Vogel's medium containing 0.5% isolated rice cell walls as the sole carbon source"
FT		/organism="Magnaporthe grisea"
FT		/strain="CP987"
FT		/clone="RCW1"
FT		/clone_lib="RCW Lambda Zap Express Library"
FT		/tissue_type="Mycelium"
FT		/dev_stage="Day 5 post-inoculation"
XX		
SQ	Sequence 337 BP; 56 A; 111 C; 74 G; 96 T; 0 other;	
	ctttttcaat cagcccgaga actcctggtt gggaaaaat cctgttctga cagctacttg	60
	tcatcgata gcccgttctt tgggtccaga taccacaagc ctgggacatt gatttcccag	120
	caactcttc aaaatggat tattagcctc ctcacgatcc ctcgcgcgtt cgcttggtcc	180
	ccttgcttca cgctctcgag ctttcagag cagtgctgt tcccggttccc tctcgactgc	240
	caccgctcgc ggccaggca aatccacaac tctcctgagg cccgcggccg ccacgaggac	300
	aagcaggttgc ttgtcgactg gttcgccctt tcgttatt	337
//		

```
<interpro id="IPR000002">
  <name>FIZZY/CDC20 domain</name>
  <type>Domain</type>
  <abstract> This domain is found in proteins ...</abstract>
  <examplelist>
    <example>
      <protein sptr_ac="Q12834" />Mammalian protein, p55CDC
    </example>
    <example>
      <protein sptr_ac="Q09649" />
    </example>
  </examplelist>
  <publist>
    <publication pub_id="PUB00006167">
      <authorlist>Shirayama M., Toth A., Galova M., Nasmyth K.</authorlist>
      <title>APC(Cdc20) promotes exit from mitosis by .....</title>
      <dbxref db="MEDLINE" dbkey="20110935" />
      <journal>Nature</journal>
      <location firstpage="203" lastpage="207" volume="402" />
      <year>1999</year>
    </publication>
  </publist>
  <memberlist>
    <dbxref db="PREFILE" dbkey="PS50218" name="FIZZY_DOMAIN" />
    <dbxref db="PRODOM" dbkey="PD004563" name="PD004563" />
  </memberlist>
</interpro>
```

```

<!ELEMENT interpro (name, type, examplelist, memberlist, publist, parlist*,  

chlist*, seclist*, abstract)>

<!ELEMENT name (#PCDATA)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT abstract (#PCDATA|cite|dbxref|sub|sup|p|li|i|ol|reaction|pre)*>
<!ELEMENT examplelist (example*)>
<!ELEMENT example (#PCDATA|protein|dbxref|cite)*>
<!ELEMENT publist (publication*)>
<!ELEMENT memberlist (dbxref*)>
<!ELEMENT protein (#PCDATA|protein)*>

<!ATTLIST interpro
  id          ID      #REQUIRED>
<!ATTLIST dbxref
  db          CDATA  #IMPLIED
  dbkey       CDATA  #IMPLIED
  name        CDATA  #IMPLIED>
<!ATTLIST protein
  sprt_ac    CDATA  #REQUIRED
  status      (?|T|P|F|N) #IMPLIED
  start       CDATA  #IMPLIED
  end         CDATA  #IMPLIED>

```

Web Services

- XML based network services
- Allow software applications to access data “intelligently”: identification of contents, interpretation of semantics information
- Standards available for their identification (UDDI), description (WSDL) and composition (WSFL)
- Metadata needed

WSDL: the description

Web Services Description Language (WSDL)

- Standard for the description of Web Services
- Define localization, access ways and detailed description
- Abstract functionalities, practical details
- WSDL Binding: implementation for SOAP, HTTP, MIME

WSDL: XEMBL.wsdl.xml

```
<definitions name="XEMBL" targetNamespace="http://www.ebi.ac.uk/XEMBL
  xmlns:tns="http://www.ebi.ac.uk/XEMBL
  xmlns:xsd="http://www.w3.org/2001/XMLSchema
  xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/
  xmlns:soapenc="http://schemas.xmlsoap.org/soap/encoding/
  xmlns="http://schemas.xmlsoap.org/wsdl/
<documentation>Documentation of this Web Service, together with a
sample client and links to Bsm1 (Labbook, Inc.) and AGAVE
(DoubleTwist, Inc.) can be found at the European Bioinformatics
Institute http://www.ebi.ac.uk/xembl/</documentation>
+ <message name="getNucSeqRequest" xmlns:tns="http://www.ebi.ac.uk/XEMBL">
+ <message name="getNucSeqResponse">
+ <portType name="XEMBLPortType">
+ <binding name="XEMBLServiceBinding" type="tns:XEMBLPortType">
+ <service name="XEMBLService">
</definitions>
```

WSDL: XEMBL.wsdl.xml

<service> </service>

```
<service name="XEMBLService">
  <documentation>Returns full information on EMBL Nucleotide Sequences
  formatted as Bsml XML or Agave XML. I.e. returns sequence itself,
  cross-references, taxonomy, literature, full feature information,
  etc.</documentation>
  <port name="XEMBLPort" binding="tns:XEMBLServiceBinding">
    <soap:address
      location="http://www.ebi.ac.uk:80/cgi-bin/xembl/XEMBL-SOAP.pl" />
  </port>
</service>
```

WSDL: XEMBL.wsdl.xml <binding> </binding>

```
<binding name="XEMBLServiceBinding" type="tns:XEMBLPortType">
  <soap:binding style="rpc"
                transport="http://schemas.xmlsoap.org/soap/http" />
  <operation name="getNucSeq">
    <soap:operation soapAction="http://www.ebi.ac.uk/XEMBL#getNucSeq" />
    <input>
      <soap:body use="encoded" namespace="http://www.ebi.ac.uk/XEMBL
                  encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
    </input>
    <output>
      <soap:body use="encoded" namespace="http://www.ebi.ac.uk/XEMBL
                  encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
    </output>
  </operation>
</binding>
```

WSDL: XEMBL.wsdl.xml <portType> </portType>

```
<portType name="XEMBLPortType">
  <operation name="getNucSeq">
    <input message="tns:getNucSeqRequest" name="getNucSeq" />
    <output message="tns:getNucSeqResponse" name="getNucSeqResponse" />
  </operation>
</portType>
```

WSDL: XEMBL.wsdl.xml <message></message>

```
<message name="getNucSeqRequest" xmlns:tns="http://www.ebi.ac.uk/XEMBL">
  <part name="format" type="xsd:string">
    <documentation>Input parameter that indicates the result format that
      should be returned. Legit values: Bsml or sciobj.
      Defaults to Bsml if format not recognised.
    </documentation>
  </part>
  <part name="ids" type="xsd:string">
    <documentation>A space delimited list of international Nucleotide
      Sequence accession numbers (IDs).
      For example: "HSERPG U83300 AC000057".
      Minimum number of IDs is 1.
    </documentation>
  </part>
</message>

<message name="getNucSeqResponse">
  <part name="result" type="xsd:string">
    <documentation>An XML formatted result in either Bsml or AGAVE
      format.</documentation>
  </part>
</message>
```

WSFL: the composition

Web Services Flow Language (WSFL)

- Standard for the description of sets of web services
- Flow models: specify how to use a set of WSs to achieve a defined goal
- Global models: describe interactions among the WSs and the overall behaviour of the set
- Recursive composition: each model can be seen as a WS itself and can be recursively inserted in other models

UDDI: the identification

Universal Description, Discovery & Integration UDDI

- It is independent from HW and SW
- Allows the creation of registries of WSS
- Based on W3C and IETF standards
- It also uses HTTP and DNS protocols

Some acknowledgements.....

A. Doyle (ECACC)
B. Dutertre (CERDIC)
J. Franklin (ASFRA)
D. Fritze (DSMZ)
F. Guissart (BCCM)
M. Kracht (DSMZ)
F. Malusa (IST)
D. Marra (IST)

L. Réchaussat (INSERM)
D. Smith (CABI)
E. Stackebrandt (DSMZ)
J. Stalpers (CBS)
G. Stegehuis (CBS)
M. Vanhoucke (BCCM)
B. Vaughan (HGMP-RC)