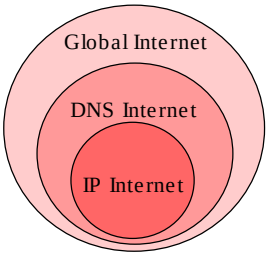
COMPUTER NETWORKS AND NETWORK-BASED BIOINFORMATICS RESOURCES

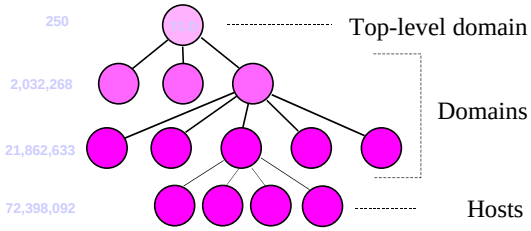
What is Internet



Domain Name System

The ***Domain Name System*** (DNS) is a hierarchical, distributed method of organizing the name space of the Internet. The DNS administratively groups hosts into a hierarchy of authority that allows addressing and other information to be widely distributed and maintained. A big advantage to the DNS is that using it eliminates dependence on a centrally-maintained file that maps host names to addresses.

Domain Name System



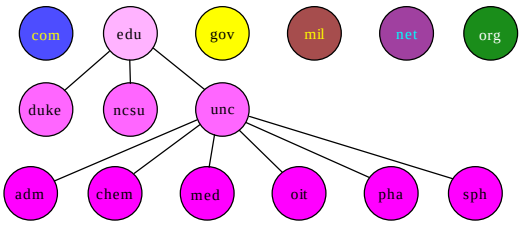
Data from ISC
(January 2000)

Top-level domains (TLD)

Domain in general (except most US networks) is an abbreviation for the country name, eg:

World		USA and "generic"	
us	USA	edu	universities
ca	Canada	com	commercial
fr	France	gov	federal institutions
il	Israel	mil	military
br	Brazil	net	networks
jp	Japan	org	other organisations
cn	China	int	international organisations

Domain Name System



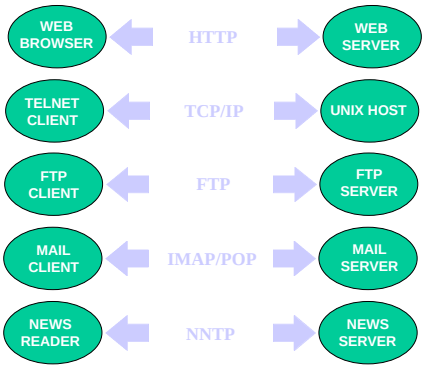
Domain Name

A ***Fully Qualified Domain Name*** (**FQDN**) is a domain name that includes all higher level domains relevant to the entity named. For example, for a host, a FQDN would include the string that identifies the particular host, plus all domains of which the host is a part up to and including the top-level domain :

user@node.subdomain.domain.tld

Client - Server Model

Client - Server Model



Uniform Resource Locator (URL)

protocol://host.domain[:port]/path/filename

file - local system file (file:///C|/DOS/README.1ST)
ftp - an anonymous FTP server (ftp://ftp.pdb.pdb.gov)
http - a World Wide Web server (http://mmlin4.pha.unc.edu/~cmb96)
gopher - a Gopher server (gopher://pdb.pdb.bnl.gov/77/)
news - an NNTP news server (news:bionet.announce)
telnet - a telnet session (telnet://nun.oit.unc.edu)
wais - a WAIS server (wais://sunsite.unc.edu)
mailto - mail interface (mailto:cmb96@mmlin4.pha.unc.edu)

Network applications in science

- Virtual Laboratory
- Virtual Library
- Virtual Conference
- Virtual Classroom

Network collaboration

Real-time data sharing -- exchange of information between remote participants in the project

Resources sharing -- remote access to the instruments and computers

Resources integration -- simultaneous use of remote instruments and computers

Bioinformatics servers

Remote data access -- databse search, cross-links between the databases

Remote computing -- use of server's processing capabilities (sequence alignment, structure prediction, homology modeling)

Infospace navigation -- pointers to the available resources

Bioinformatics servers

Real-time

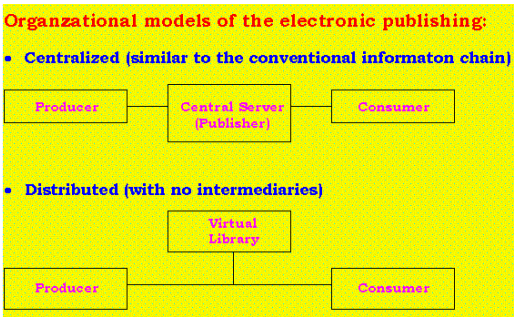
Asynchronous

Digital information cycle

Creation and capture
Storage and management
Rights management
Search and access
Distribution

Electronic publishing

Quality (peer review, retrospective evaluation)
Reliability (stability of serves, control over alterations, proper archiving and mirroring)



Hypertext Functionality in Scientific Literature

Active references
Forwarding references
Dynamic publishing

Ethical, Legal, and Economical Issues of Electronic Publishing

Intellectual property rights
Ownership of information
Information as a commodity

Molecular Databases

Nucleic acid sequences: GenBank, DNA Data Bank of Japan, EMBL Nucleotide Sequence Database
Nucleic acid structures: NDB - Nucleic Acid Database

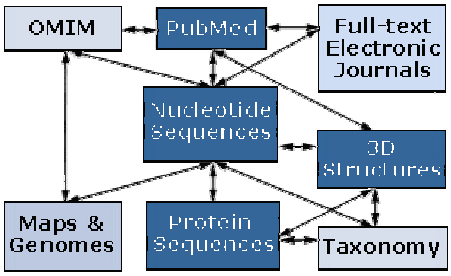
Protein sequences: PIR - Protein Information Resource, SWISS-PROT
Protein structures: PDB - Protein Data Bank , NRL_3D

Physical properties: Biological Macromolecule Crystallization Database, BioMagResBank

Molecular images: SWISS-3DIMAGE, Molecules R US

NCBI integrated search and retrieval system

Entrez



NCBI Databases

- **nr** - All non-redundant GenBank CDS translations+PDB+SwissProt+PIR
- **month** - All new or revised GenBank CDS released in the last 30 days
- **swissprot** - the last major release of the SWISS-PROT protein sequence database (no updates)
- **yeast** - Yeast (Saccharomyces cerevisiae) protein sequences.
- **E. coli** - E. coli genomic CDS translations
- **pdb** - Sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank
- **kabat** - Kabat's database of sequences of immunological interest

Database structure

database	a collection of related structured information about entities
file	a collection of records
record	a set of fields
field	a single characteristic of an entity
character	a symbol used in data field

Example of a Genbank entry

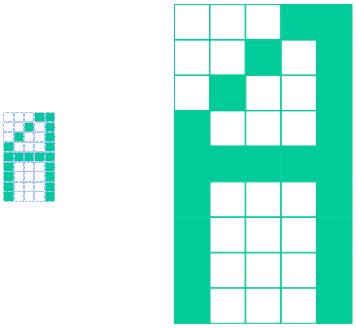
```
LOCUS       VIBHALUXA    3141 bp    DNA             BCT       15-FEB-1996
DEFINITION  V.harveyi luciferase alpha and beta subunit (luxA and luxB) genes,
            complete cds.
ACCESSION   M10961 M13494
NID         g155174
KEYWORDS    luciferase.
SOURCE      Vibrio harveyi DNA.
  ORGANISM  Vibrio harveyi
            Eubacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
            Vibrio.
REFERENCE   1  (bases 1 to 1838)
  AUTHORS   Cohn,D.H., Mileham,A.J., Simon,M.I., Nealson,K.H., Rausch,S.K.,
            Bonam,D. and Baldwin,T.O.
  TITLE     Nucleotide sequence of the luxA gene of Vibrio harveyi and the
            complete amino acid sequence of the alpha subunit of bacterial
            luciferase
  JOURNAL   J. Biol. Chem. 260 (10), 6139-6146 (1985)
  MEDLINE   85207595
REFERENCE   2  (bases 1745 to 3141)
  AUTHORS   Johnston,T.C., Thompson,R.B. and Baldwin,T.O.
  TITLE     Nucleotide sequence of the luxB gene of Vibrio harveyi and the
            complete amino acid sequence of the beta subunit of bacterial
            luciferase
  JOURNAL   J. Biol. Chem. 261 (11), 4805-4811 (1986)
  MEDLINE   86168191
```

Example of a Genbank entry

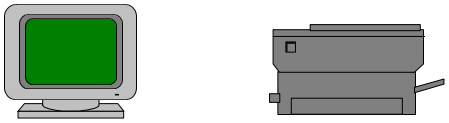
```
FEATURES             Location/Qualifiers
     gene            707..1774
                     /gene="luxA"
     CDS             707..1774
                     /gene="luxA"
                     /codon_start=1
                     /product="luciferase alpha subunit"
                     /db_xref="PID:g155175"
                     /transl_table=11
                     /translation="MKFGNFLLTYQPPELSQTEVMKRLVNLGKASEGCGFDTVWLLEH
                     HFTEFGLLGNPYVAAAHLLGATETLNVGTAAIVLPTAHPVRQAEDVNLLDQMSKGRFR
                     FGICRGLYDKDFRVFGTDMDNSRALMDCWYDLMKEGFNEGYIAADNEHIKFPKIQLNP
                     SAYTQGGAPVYVVAESASTTEWAAERGLPMILSWIINTHEKKAQLDLYNEVATEHGYD
                     VTKIDHCLSYITSVDHDSNRAKDICRNFLGHWYDSYVNATKIFDDSDQTKGYDFNKGQ
                     WRDFVLKGHKDTNRRIDYSYEINPVGTPEECIAIIQQDIDATGIDNICCGFEANGSEE
                     EIIASMKLFQSDVMPYLKEKQ"
BASE COUNT      883 a    665 c    741 g    852 t
ORIGIN      1 bp upstream of EcoRI site.
        1 gaattcacca tgacgacggg caaaaatagt ttgtgcactg tttatcactg gctgcagacc
       61 aagggcacac aaaacattgg cttgattgcg gcaagtctct cagctcgtgt cgcctatgaa
      121 gttatctctg atctggagct gtcttttctg attactgcgg ttggtgtggt gaacttgcgt
      181 gacacactag aaaaagcgct tggttttgat tacctcagtt tgcctatcga tgagctacca
          ....
```

Molecular Visualization

Image representation



Standard (common) resolution

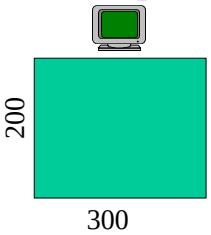


72 dpi

600 dpi

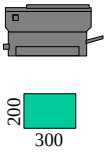
Resolution effect on the image size

72 dpi



2.8 x 4.2 in

600 dpi



0.3 x 0.5 in

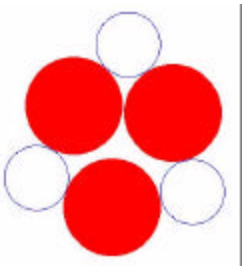
Common color resolution

- Standard VGA: 4 bits per pixel, 16 colors.
- Super VGA: 8 bits per pixel, 256 colors.
- Higher resolution: 16 bits, 32,000 colors;
  24 bits, 16 million colors

Common image formats

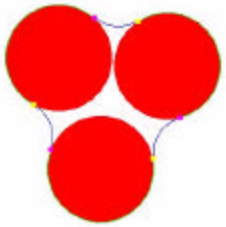
- GIF - Graphic Interchange Format
  8 bits/pixel (256 or fewer colors)
- JPEG - Joint Photographic Experts Group
  24 bits/pixel (16 million colors)

Connolly Surface (Theory)



probe sphere is placed tangent to each pair of atoms

Connolly Surface (Theory)

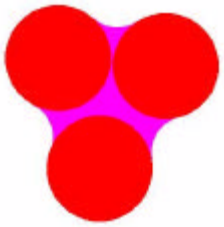


concave arcs are connected by convex arcs

Connolly Surface (Theory)

the union of the convex arcs and concave arcs
produces a continuous, smooth contour

Connolly Surface (Theory)



the region inaccessible to a probe sphere is called
the **solvent-excluded volume**. It consists of two parts:

- the van der Waals volumes of the atoms
- the interstitial volumes between the atoms