Data Integration using the Distributed Annotation System (DAS)

Andreas Prlić, Ewan Birney, Tony Cox, Thomas A. Down, Rob Finn, Stefan Gräf, David Jackson, Andreas Kähäri, Eugene Kulesha, Roger Pettett, James Smith, Jim Stalker, Tim J. P. Hubbard
• what is DAS
• what do we do with it
• DAS registration server
• latest developments
Integration of personal data into bioinf. resources
Integration of annotations from external sources into local applications
• online access to most recent data versions
  - no need for local installations
DAS, how it works

Dowell, Jokerst, Allen, Eddy, Stein
BMC Bioninformatics 2001

http:// request

XML response

get sequence

get features
a few principles...

• Clients are “intelligent” (few)
• Servers are simple and easy to set up (many)
• (most of) data is precalculated
• libraries for server and client
• multiple programming languages
http://www.ensembl.org
• > 20 vertebrates / model organism

• 5 mill. page impressions / week
• 100 mirrors/internal installations worldwide
• open source
• used for other species as well
• MySQL
• 5-10 G / species + 100 G multi species data
Add your own uses Registry
Linking protein structure to e! Peptide view
See exon structure mapped onto 3D

Click
interact with Menu & RASMOL

RASMOL commands

Zoom
1a17. 2.45A XRAY description: SERINE/THREONINE PROTEIN PHOSPHATASE 5

1s95.A 1.6A XRAY description: SERINE/THREONINE PROTEIN PHOSPHATASE 5

1s95.B 1.6A XRAY description: SERINE/THREONINE PROTEIN PHOSPHATASE 5
DAS commands

Structure

Features

Alignment

Sequence
Meta information about DAS servers

DAS registry

SPICE
# The DAS registration server

<table>
<thead>
<tr>
<th>DAS</th>
<th>Source Type</th>
<th>Source URL</th>
<th>Source Type Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS_110</td>
<td>dssp</td>
<td>cmbl4.cmi.ru.nl/das/dssp/</td>
<td>features PDBresnum, Protein Structure</td>
</tr>
<tr>
<td>DS_111</td>
<td>cath</td>
<td>cathwww.biochem ... 0/das/cath_pdb/</td>
<td>features PDBresnum, Protein Structure</td>
</tr>
<tr>
<td>DS_112</td>
<td>structure</td>
<td>das.sanger.ac.uk/das/structure/</td>
<td>structure PDBresnum, Protein Structure</td>
</tr>
<tr>
<td>DS_113</td>
<td>alig_pdb_sp</td>
<td>das.sanger.ac.uk/das/msdpdbsp/</td>
<td>alignment UniProt, Protein Sequence</td>
</tr>
<tr>
<td>DS_114</td>
<td>signalp</td>
<td>genome.cbs.dtu.dk:9000/das/signalp/</td>
<td>types features UniProt, Protein Sequence</td>
</tr>
<tr>
<td>DS_115</td>
<td>netphos</td>
<td>genome.cbs.dtu.dk:9000/das/netphos/</td>
<td>types features UniProt, Protein Sequence</td>
</tr>
<tr>
<td>DS_116</td>
<td>netoglyc</td>
<td>genome.cbs.dtu.dk:9000/das/netoglyc/</td>
<td>types features UniProt, Protein Sequence</td>
</tr>
<tr>
<td>DS_117</td>
<td>tmhmm</td>
<td>genome.cbs.dtu.dk:9000</td>
<td></td>
</tr>
</tbody>
</table>
DAS registration server

- allows to “publish” DAS servers & share with community
- communicates with clients
- regularly checks servers, sends notification
What is the glue?

- “Coordinate Systems”
- Authority
- Type of data
- Version
- Organism (optional)
Clients and Coordinate Systems

- Ensembl - most of the views can display DAS sources from multiple CS
- SPICE - PDB, UniProt, Ensp
- Dasty - UniProt
DAS registration server

Service Registry

Find → Publish

Service Requestor  DAS  Service Provider

e.g. Ensembl, SPICE

a DAS source

the DAS - SOA
111 DAS sources
26 institutions
12 countries

+ others
DAS - issues

• inconsistent implementations
• no consistent use annotation types
• error handling
• searches not possible - in DAS/1
• open sharing of data - low security
Structural alignments for proteins with non-trivial relationships

Sisyphus: in greek: crafty

Authors:
Antonina Andreeva, Andreas Prlic, Tim Hubbard, Alexey Murzin

http://sisyphus.mrc-cpe.cam.ac.uk
enter RASMOL like command...
• Alignment DAS:

• rotation matrices, shift vectors

• range information (optional)
Acknowledgments

• T. Down, T. Hubbard
• Web Team, E. Kulesha, R. Pettett, T. Cox
• eFamily Project
• S. Gräf, A. Kahari, BioSapiens
• A. Murzin, A. Andreeva
• R. Finn, H. Hotz, A. Ahmed
• Jmol, Biojava, MSD, everybody who sets up DAS servers