

DAS Current Situation and Future Developments

Jonathan Warren

DAS coordinator for the Sanger
Institute

www.biodas.org

www.dasregistry.org

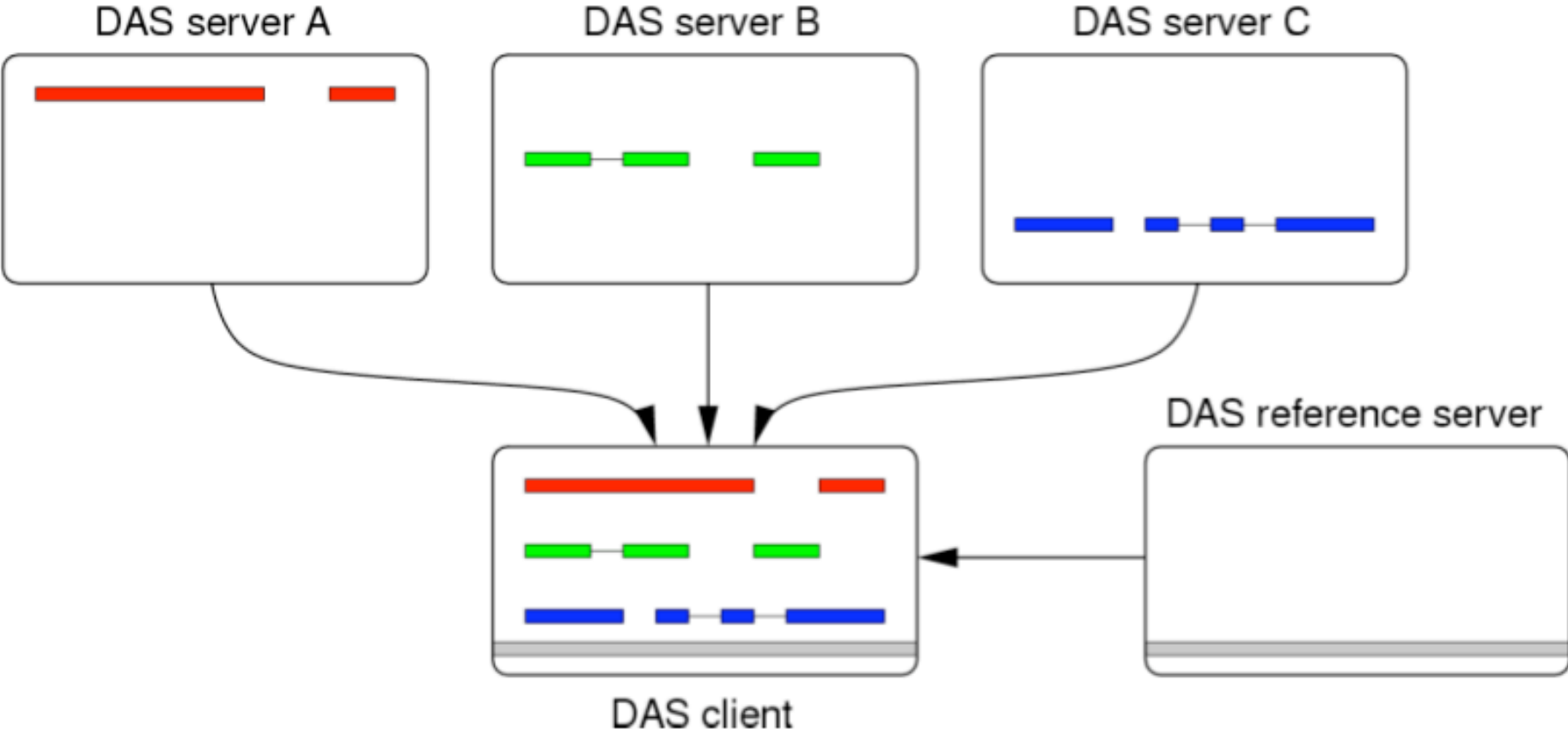
Contents:

- What is DAS?
- What is the history and current situation?
- What are we doing at the moment and where are we going?

What is DAS?

- Distributed Annotation System
- Based on HTTP and XML
- User perspective
 - Client
 - Choose coordinate system
 - Connects to one registry for DAS server list
 - Request a region of interest from the reference and many annotations from the DAS servers.

What is DAS?



What is DAS?

- Clients are “intelligent”
- Servers should be spec compliant
 - Common request format
 - Common xml format returned

History

- 2001 The Distributed Annotation System
Robin D Dowell, Rodney M Jokerst, Allen Day, Sean R Eddy and Lincoln Stein
- DAS1.53E + DAS2 running in parallel.
- At workshop in March this year DAS1.6E is expected to provide the functionality that DAS2 users desired => DAS2 is dead?
- 1.6 spec has new features and is a consolidation of the way DAS is being used.
- 1.6E has extensions being developed.

Some DAS 1.5/1.6 Commands:

- Sources
- Features
- Sequence
- types
- Stylesheet
- Structure
- Alignment
- Interaction

Current Situation- addressing issues:

Represent features with more than two
levels 1.6

- Reliably relate feature types to a more
structured ontology 1.6

Addressing issues cont:

- Easily identify when two DAS servers are using the same coordinate system (doable with help of Sanger DAS registry)
- Have a standard way to create and edit DAS features 1.6E

Addressing issues cont:

- **Verify whether a DAS server is compliant with the specification.**
 - Critical for improving interoperability between clients and servers developed by different groups. The Registry

What are we doing?

- Increasing validation capability of the registry for 1.53E and upcoming 1.6E spec
- RelaxNG
 - <http://www.dasregistry.org/validation/sources.rng>

sorted by id

[1 - 10 | 625] next > last >> | show all

pos	id	clients	nickname	status	capabilities	coordinateSystem	description	project
1	DS_109		uniprot aristotle			UniParc, Protein Sequence IPI, Protein Sequence UniProt, Protein Sequence	show > go to site	
2	DS_110		dssp			PDBresnum, Protein Structure	show >	
3	DS_111		cath			PDBresnum, Protein Structure	show > go to site	
4	DS_112		structure			PDBresnum, Protein Structure	show >	
5	DS_113		alig_pdb_sp			UniProt, Protein Sequence PDBresnum, Protein Structure	show >	
6	DS_114		signalp			UniProt, Protein Sequence	show > go to site	
7	DS_115		netphos			UniProt, Protein Sequence	show > go to site	
8	DS_116		netoglyc			UniProt, Protein Sequence	show > go to site	
9	DS_117		tmhmm			UniProt, Protein Sequence	show > go to site	
10	DS_118		prop			UniProt, Protein Sequence	show > go to site	

[1 - 10 | 625] next > last >> | show all

<http://www.dasregistry.org/das/sources>

http://www.dasregistry.org/services/das:das_directory?wsdl

```
Source of: http://www.dasregistry.org/das/sources
<?xml version='1.0' encoding='UTF-8' ?>
<?xml-stylesheet type="text/xsl" href="das.xsl"?>
<SOURCES>
<SOURCE uri="DS_109" title="uniprot aristotle" doc_href="http://www.ebi.ac.uk/uniprot-das/" description="This datasource (aristotle) is a legacy datasource that comprises the new 'uniprot', 'ipi' and
<MAINTAINER email="rantunes@ebi.ac.uk" />
<VERSION uri="DS_109" created="2005-03-21T16:26:03+0000">
<COORDINATES uri="http://www.dasregistry.org/dasregistry/coordsys/CS_DS93" source="Protein Sequence" authority="UniParc" test_range="UPI00000017EA">UniParc,Protein Sequence</COORDINATES>
<COORDINATES uri="http://www.dasregistry.org/dasregistry/coordsys/CS_DS35" source="Protein Sequence" authority="IPI" test_range="IPI00000021">IPI,Protein Sequence</COORDINATES>
<COORDINATES uri="http://www.dasregistry.org/dasregistry/coordsys/CS_DS6" source="Protein Sequence" authority="UniProt" test_range="P00280">UniProt,Protein Sequence</COORDINATES>
<CAPABILITY type="das1:sequence" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/sequence" />
<CAPABILITY type="das1:types" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/types" />
<CAPABILITY type="das1:features" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features" />
<CAPABILITY type="das1:entry_points" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/entry_points" />
<CAPABILITY type="das1:stylesheet" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/stylesheet" />
<PROP name="label" value="Predicted" />
<PROP name="label" value="Manually curated" />
<PROP name="label" value="ENSEMBL" />
<PROP name="leaseTime" value="2009-07-30T13:00:06+0000" />
<PROP name="projectHome" value="http://www.biosapiens.info" />
<PROP name="projectIcon" value="http://www.dasregistry.org/ProjectIcon?id=74" />
<PROP name="projectDesc" value="BioSapiens is a Network of Excellence, funded by the European Union's 6th Framework Programme, and made up of bioinformatics researchers from 25 institutions based in
The objective of the BioSapiens is to provide a large" />
<PROP name="projectName" value="BioSapiens" />
<PROP name="valid" value="sequence" />
<PROP name="valid" value="types" />
<PROP name="valid" value="features" />
<PROP name="valid" value="entry_points" />
<PROP name="valid" value="stylesheet" />
</VERSION>
</SOURCE>
<SOURCE uri="DS_409" title="uniprot" doc_href="http://www.ebi.ac.uk/uniprot-das/" description="DAS 1 reference server for sequence and features from UniProt. UniProt (Universal Protein Resource) is t
<MAINTAINER email="rantunes@ebi.ac.uk" />
<VERSION uri="DS_409" created="2007-08-15T15:43:55+0000">
<COORDINATES uri="http://www.dasregistry.org/dasregistry/coordsys/CS_DS6" source="Protein Sequence" authority="UniProt" test_range="P00280">UniProt,Protein Sequence</COORDINATES>
<CAPABILITY type="das1:sequence" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot/sequence" />
<CAPABILITY type="das1:types" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot/types" />
<CAPABILITY type="das1:features" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot/features" />
<CAPABILITY type="das1:entry_points" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot/entry_points" />
<CAPABILITY type="das1:stylesheet" query_uri="http://www.ebi.ac.uk/das-srv/uniprot/das/uniprot/stylesheet" />
<PROP name="label" value="Predicted" />
<PROP name="label" value="Manually curated" />
<PROP name="label" value="eProtein" />
<PROP name="label" value="BioSapiens" />
<PROP name="leaseTime" value="2009-07-30T13:00:06+0000" />
<PROP name="projectHome" value="http://www.biosapiens.info" />
<PROP name="projectIcon" value="http://www.dasregistry.org/ProjectIcon?id=74" />
<PROP name="projectDesc" value="BioSapiens is a Network of Excellence, funded by the European Union's 6th Framework Programme, and made up of bioinformatics researchers from 25 institutions based in
The objective of the BioSapiens is to provide a large" />
<PROP name="projectName" value="BioSapiens" />
<PROP name="valid" value="sequence" />
<PROP name="valid" value="types" />
<PROP name="valid" value="features" />
<PROP name="valid" value="entry_points" />
<PROP name="valid" value="stylesheet" />
</VERSION>
</SOURCE>
<SOURCE uri="DS_110" title="dssp" description="Information from DSSP">
<MAINTAINER email="m.hekkelman@cmbi.ru.nl" />
<VERSION uri="DS_110" created="2005-03-21T16:27:01+0000">
<COORDINATES uri="http://www.dasregistry.org/dasregistry/coordsys/CS_DS7" source="Protein Structure" authority="PDBresnum" test_range="5pti">PDBresnum,Protein Structure</COORDINATES>
<CAPABILITY type="das1:features" query_uri="http://cmbi4.cmbi.ru.nl/das/dssp/features" />
<PROP name="label" value="BioSapiens" />
<PROP name="leaseTime" value="2008-06-11T07:00:02+0000" />
<PROP name="projectHome" value="http://www.biosapiens.info" />
<PROP name="projectIcon" value="http://www.dasregistry.org/ProjectIcon?id=74" />
```

Currently?

- More validation (headers and feature by id).
- Capability of bulk uploading/mirroring DAS sources to Registry (sources cmd).
 - Adding all of ensembl genomes (bacteria and viruses) as DAS sources and to the registry.
- Completing the 1.6 spec - hierarchies, nextFeature.
- Updating client libraries and servers to work with both 1.53 and 1.6 spec

Future continued:

- New user interface to the registry for faster searching using Lucene - also limited version available from Sanger and EBI sites.
- Greater support for ontologies-give me all the sources that provide genes?

Libraries:

- PERL
 - Proserver, LDAS - servers
 - Bio::Das::Lite - client library
- Java
 - Dazzle, MyDAS - servers
 - Dasobert - client library

Some DAS providers:

- * Affymetrix
- * BioSapiens servers
- * Ensembl server
- * KEGG DAS
- * Sanger DAS server
- * EBI Genomic DAS server
- * EBI Protein DAS server
- * Uniprot DAS server
- * TIGR's listing of servers
- * UCSC server

Some clients:

- * Ensembl
- * Spice
- * Dasty
- * Pfam
- * STRAP
- * DASher

Useful links:

- www.biodas.org
- www.dasregistry.org

Acknowledgments:

- Andy Jenkinson (EBI)
- James Smith
- Andreas Prlic
- Gregg Helt
- Lincoln Stein
- Tim Hubbard
- The DAS community