BMC Bioinformatics

Research article

Integrating sequence and structural biology with DAS

Andreas Prlić*[1], Thomas A Down[3], Eugene Kulesha[2], Robert D Finn[1], Andreas Kähäri[2] and Tim JP Hubbard[1]

Address: [1]The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK, [2]European Bioinformatics Institute, Hinxton, Cambridge, UK and [3]Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge University, Cambridge, UK

Email: Andreas Prlić* - ap3@sanger.ac.uk; Thomas A Down - thomas.a.down@googlemail.com; Eugene Kulesha - ek@ebi.ac.uk; Robert D Finn - rdf@sanger.ac.uk; Andreas Kähäri - ak@ebi.ac.uk; Tim JP Hubbard - th@sanger.ac.uk

* Corresponding author

Abstract

Background: The Distributed Annotation System (DAS) is a network protocol for exchanging biological data. It is frequently used to share annotations of genomes and protein sequence.

Results: Here we present several extensions to the current DAS 1.5 protocol. These provide new commands to share alignments, three dimensional molecular structure data, add the possibility for registration and discovery of DAS servers, and provide a convention how to provide different types of data plots. We present examples of web sites and applications that use the new extensions. We operate a public registry of DAS sources, which now includes entries for more than 250 distinct sources.

Conclusion: Our DAS extensions are essential for the management of the growing number of services and exchange of diverse biological data sets. In addition the extensions allow new types of applications to be developed and scientific questions to be addressed. The registry of DAS sources is available at http://www.dasregistry.org

Background

New large scale techniques in biology have produced a rapidly increasing range of publically available data. A number of centralized database resources are available that aim to integrate this data, such as Entrez [1], Interpro [2], MSD [3], and Ensembl [4]. Databases that integrate huge datasets are faced with problems regarding scaleability of their storage facilities, how to manage frequent updates and how to exchange the data with the community. Rather than building bigger centralized repositories, an alternative strategy is to encourage individual data providers to offer their results in standard formats, then provide users with tools which can find and communicate with these distributed repositories, and build integrated

views on demand. The *Distributed Annotation System* (DAS) is a widely used system for biological sequence annotation that follows this latter approach [5]. DAS is frequently used for

1. Integration of personal data into major bioinformatics portals like Ensembl [4].

2. Integration of the annotations from external sources into local applications.

3. Access to the most recent versions of biological databases without the need for local installations.

Reference and Annotation servers

DAS is a simple client-server network protocol. DAS clients make requests by fetching a defined URL from a DAS server, and receive simple XML responses. It is a web service, in the sense that all requests are made using the HyperText Transfer Protocol (HTTP). This also allows to use existing web proxy infrastructure for accessing remote servers through firewalls. An important part of the request URL is the DAS command, which specifies the type of information that the client is requesting, and thus the response format which it expects to receive. Key commands in the DAS 1.5 protocol are *sequence*, which returns all or part of a DNA or protein sequence, and *features*, which returns a set of annotated regions on a sequence, such as genes or protein domains. See [6] for the full specification of the DAS 1.5 protocol.

Central to the DAS system is the idea of *reference objects*. These are biological data objects with stable identifiers which are targets for annotation. In the original DAS protocol, the reference objects are always biological sequences: chromosomes, scaffold sequences from genome assemblies or protein sequences. When a DAS client starts, its first action is to connect to an appropriate reference server and retrieve the reference sequence. Once a reference object has been loaded, the DAS client will contact one or more *annotation* servers to obtain the data provided for the reference objects. Typical annotations might include sets of predicted exons on a genome sequence, or matches to a protein domain model on a protein sequence. It is the client's responsibility to collect all relevant annotations and display them in a user-friendly manner. Genomic sequences are generally much longer than protein sequences, in terms of number of nucleotides and amino acids, so it is possible to restrict DAS requests to sub-ranges of the data. Biological sequence data is dynamic: sequences change as new sequence data becomes available. Thus it is important that DAS clients are able to check that the annotation they have received match the available reference sequence. The ideal solution is a version number for each sequence, administered by whatever authority assigns identifiers for sequences that is incremented each time the sequence changes. Not all biological databases provide such a version number though, so the members of the BioSapiens project [7] have agreed to use the MD5 digest of the sequences for both the reference objects and the annotations as part of the version attribute in the DAS responses.

Results

New DAS commands

The original DAS specification was rather genome and sequence centric. DAS mediated data exchange and visualisation today is heavily used in the popular genome browsers like Ensembl ([4]), Wormbase [8], and GBrowse

[9]. Several protein sequence DAS clients are available like *Dasty* [10], ProView [11] or the CBS DAS Viewer [12]. The multiple sequence alignment viewer Jalview has recently been extended to support accessing features through DAS [13]. CARGO is a flexible system for visualization and analysis of biological information which among several different sources of data can use DAS [14].

To generalize the system and enable the development of new applications, we have extended the specification and added several new commands. The new DAS commands are:

1. *alignment* – request an alignment or mapping between two or more reference objects. The alignment format is quite general, and can specify pairwise and multiple alignments between protein sequences, structures and chromosomes.

2. *structure* – request coordinate data for three dimensional reference objects such as protein structures.

3. *sources* – request and query meta information about available DAS sources.

Implementation

The guiding principle for the design of the new DAS commands was to provide a *reference* object for three dimensional information, similar to the sequence command. The alignment command should be rather generic and support multiple alignments of objects of the same data type. In connection with the *coordinate systems* introduced below, it can also provide mappings between objects of different data types, like mappings between protein sequence and structure residue positions or alignments of chromosomes. Similar to the original DAS commands, data should be pre-computed. For serving structure alignments, the actual calculation of the superimposition should be done on the server side, so the client can be independent from the actual implementations of structure alignment algorithms and support any of the number of available methods, for example [15-18].

We have extended the two DAS server implementations ProServer, which is written in Perl [19] and its Java counterpart, Dazzle [20] to support the new commands. For a detailed documentation on the specification of the new DAS commands please see [21].

Publishing and discovery of DAS sources

DAS is now a widely-accepted communication protocol. With a growing number of people offering data via DAS, it has become increasingly difficult to keep track of the different servers and the functionality and data they offer. Originally there was no specification for a registration

mechanism and if a user wanted to add a new DAS source to a client, they were expected to discover the server's URL outside of the DAS system, perhaps by reading web pages or papers, then manually enter the URL into their DAS client. Moreover, with the extensions presented here DAS becomes applicable to new data types, which adds another level of complexity. We believe a registration system is essential to support the continued development and usage of DAS.

In order to address this problem we have developed a DAS extension for communicating with registry servers, the *sources* command. Registry entries include the base URL for communicating with a server, a human-readable description of the data it provides, and also some machine-readable meta information which helps clients to automatically decide which sources might be relevant to a given reference object. See Figure 1 for an example how a DAS source is being described with this meta information.

Multiple Data Types
DAS is used to annotate many different kinds of biological entities. Genome sequences, gene loci, protein sequences, and molecular structures are currently the most common targets for DAS annotation. The DAS registration system therefore needs to be able to deal with the different data types. A DAS client on the other side usually will only support a certain set of these types, for example, a particular view of the Human genome will not necessarily be able to display data that has been annotated on the Chimpanzee genome. To allow data integration, the DAS clients must be able to find all the DAS annotation sources which can be used with a particular reference object. One of the goals of the DAS registration service is to provide a description of these sets of data types. It is called *coordinate system*, and can also be thought of as a "namespace" for reference objects.

There are several parts to the definition of a coordinate system:

• The *authority* or name of a data set. This usually is the name of an institution or organization that defines identifiers or accession codes. For example, UniProt is an authority to assign protein sequence accession codes.

• The *type* of object refers to the physical entity being used. Currently supported are *Chromosome, Clone, Contig, Gene_ID, NT_Contig, Protein Sequence, Protein Structure*, and *Scaffold*.

• The *organism* from which the data was derived. This field can be empty in the case of multi-species databases like UniProt, but is important for genome assemblies.

• Some resources, especially genome assemblies, have a versioning scheme which applies to the entire collection of sequences rather than to a single sequence. In this case, the optional *version* field can represent this, and allows data sources annotating several different versions of a given genome to co-exist in the same registration server.

Implementation of a DAS Registry
Based on the concepts of the *sources* meta description of DAS sources, we have established a central DAS registry which implements this protocol and fulfils the following roles:

1. It allows discovery of available DAS sources via a web page, or as machine-readable XML which can be used directly by DAS client programs.

2. It automatically validates registered DAS sources to ensure that they return well formed DAS XML.

3. It periodically tests DAS sources and notifies their administrators if they are unavailable.

4. It can group the registered DAS sources according to the coordinate systems of their data.

5. It can also communicate bi-directionally with DAS clients and activate or highlight DAS sources in clients.

The DAS registration server is available at [22]. Currently it contains a listing of more than 250 DAS servers provided by about 40 groups in 13 countries. To obtain a listing of meta information about all registered DAS servers, the new *sources* DAS command is provided at [23]. The same command can be used to batch-upload DAS servers, for example Ensembl provides a listing of DAS sources via [24]. The registry automatically obtains this listing and updates the meta information for these servers.

The DAS registry is backed by a MySQL relational database, and can be accessed in several different ways. As well as the DAS-style *sources* interface, a HTML front-end is provided which allows manual interaction with the registry [22]. A SOAP style webservice can be accessed from [25].

Two tools that the DAS registry offers are validation and auto-activation of DAS sources in DAS clients.

Validation
There are a number of different DAS server programs available. Frequently used server libraries include Dazzle [20], ProServer [26], LDAS [27] or GBrowse [28]. Due to the simplicity of DAS, sometimes individually implemented CGI scripts are used as well, thus it is important to ensure that a correct response is given regardless of the

```
<SOURCE uri="DS_146"
    title="ensp_pdb_mapping"
    doc_href="http://www.efamily.org.uk/software/dasclients/spice/ensembl.shtml"
    description="If you click on a feature of this DAS track, SPICE can be launched
        from ENSEMBL. - Mapping of ENSP protein sequences to
        protein structures. Based on the Compara ENSP - UniProt and
        the MSD - UniProt to PDB mapping. ">

    <MAINTAINER email="ap3@sanger.ac.uk" />

    <VERSION uri="latest" created="2005-06-03">

        <COORDINATES uri="http://www.dasregistry.org/coordsys/CS_DS16"
                source="Protein Sequence"
                authority="Ensembl"
                test_range="ENSP00000345942"
        >Ensembl,Protein Sequence</COORDINATES>

        <CAPABILITY  type="das1:features"
                query_uri="http://das.sanger.ac.uk/das/ensppdbmapping/features" />

        <PROP name="label" value="ENSEMBL" />
        <PROP name="label" value="eFamily" />
        <PROP name="leaseTime" value="2007-06-30" />

    </VERSION>

</SOURCE>
```

Figure 1
Example of the new DAS sources response. It provides meta description for DAS servers. The meaning of the different elements are: SOURCE: This DAS source provides mappings to protein structures for Ensembl proteins. It has a unique ID in the DAS registration server (DS_146). The doc_href provides a link to a web page with more documentation on this data. MAINTAINER: contact this person if there is a problem with the server, or if you have any questions about it. VERSION: this DAS source has been registered at the DAS registry in 2005 and provides the latest available data. COORDINATES: The data is served as annotations for the set of protein sequences provided by Ensembl. An example accession code is provided which can be used to check if the server is giving a valid response. CAPABILITIES: The DAS/1 features command is being supported by this server. The query uri field provides the location of the DAS source. PROP: Different labels can be assigned. The leaseTime is the last time when this server has been validated successfully.

implementation. To do this the DAS registry employs a validation mechanism such that only servers providing well formed DAS-XML can get registered. For this validation a *test code* is required, which is an accession code that identifies a reference object for which features are provided. To ensure that the listing within the DAS registry is current, the servers are periodically validated to ensure that the server is still available and the XML response is well formed. Successful attempts are logged, and a graphical summary of the availability of a DAS source is provided via the registry's web interface. If the DAS source can not be validated for more than two days, a *watchdog* can optionally inform the server's administrator by email. If a server is down for a longer period of time, the server

administrator can be contacted to inquire about the status of the server. If the server remains unavailable for an extended period, it will be removed from the listing.

Auto-activation
Users can browse and search through the DAS registry website. If they find an interesting DAS source it can be conveniently be enabled or highlighted in their favorite DAS client, if this client supports this. The listing of each DAS server in the registry's web interface contains specially formatted links to the clients. When selected, these links not only launch the DAS client, but highlight or turn on that particular DAS source, which we have termed auto-activation. At present, the clients that support the
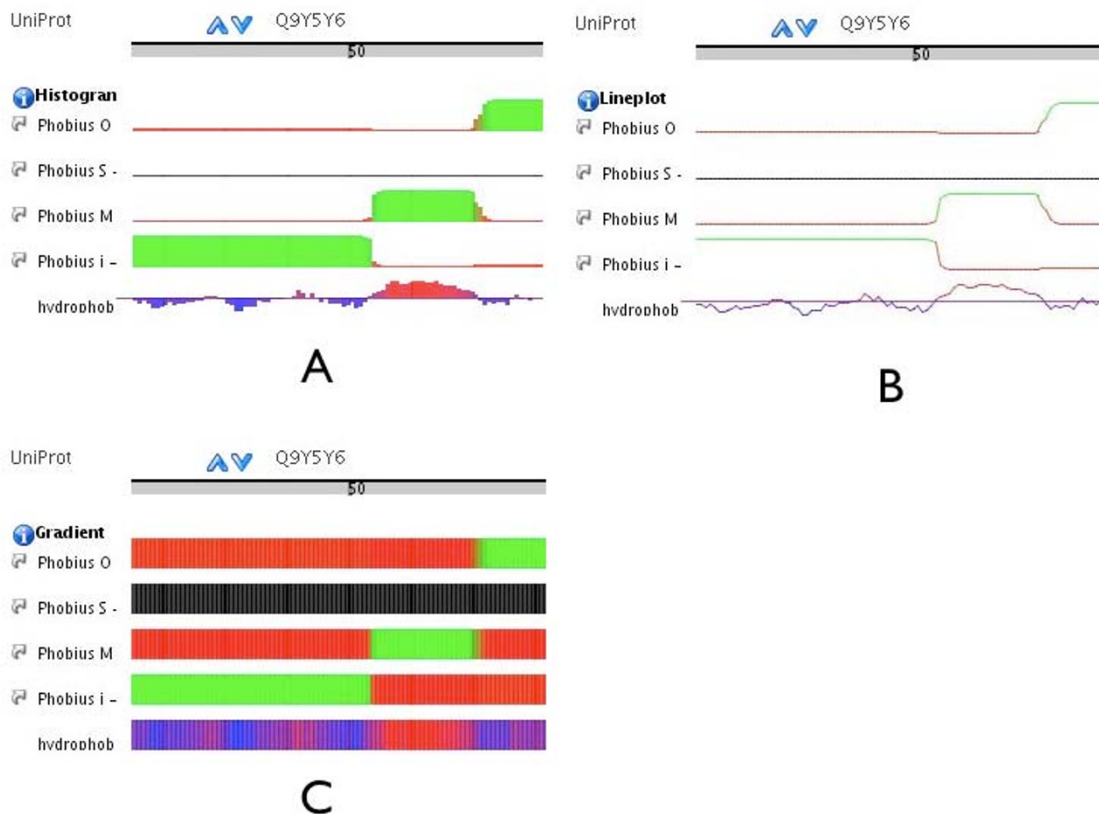
Figure 2
We are providing a convention how to configure the display of different data-plots via the DAS stylesheet command. Here an example how this is used for plotting the scores of the signal peptide and transmembrane topology predictor Phobius [48], as well as hydropathicity data for the Suppressor of tumorigenicity protein 14 (Q9Y5Y6). The DAS features used in the three displays are exactly the same. The configuration is obtained from the stylesheet responses. The example can be also viewed via [49]. A) Using the stylesheet for histogram display. B) Using the stylesheet for lineplot display. C) Using the stylesheet for gradient display.

registry directed auto-activation are Ensembl [4], SPICE [29], and Dasty [10]. The registry also provides a *send to friend* mechanism to share auto-activation links by email.

Positional and non-positional features and plots of data
The DAS protocol provides the *features* command to share annotations that have a *start* and *stop* position. This convention does not deal well with special cases where annotation can not be assigned to a defined region, or where the feature has a size of one. For data that is *non-positional*, such as literature references, the working convention has been adopted of providing 0 as the start and stop positions.

Some data sets, such as hydropathicity plots or signal peptide prediction scores, can be transmitted via DAS by pro-

viding a different score for each feature position. The DAS 1.5 specification provides the *stylesheet* command which allows to specify how a DAS client should visualize data. We developed a convention how to configure different types of plots for Ensembl and SPICE [30] and implemented support for this in the two DAS clients (see Figure 2).

Applications
The new DAS extensions presented here are actively being used by several web sites and applications. This section highlights a couple of these as examples.

Pfam
The *Pfam* database [31] now makes a large proportion of its data available via DAS. In addition to the classical DAS-
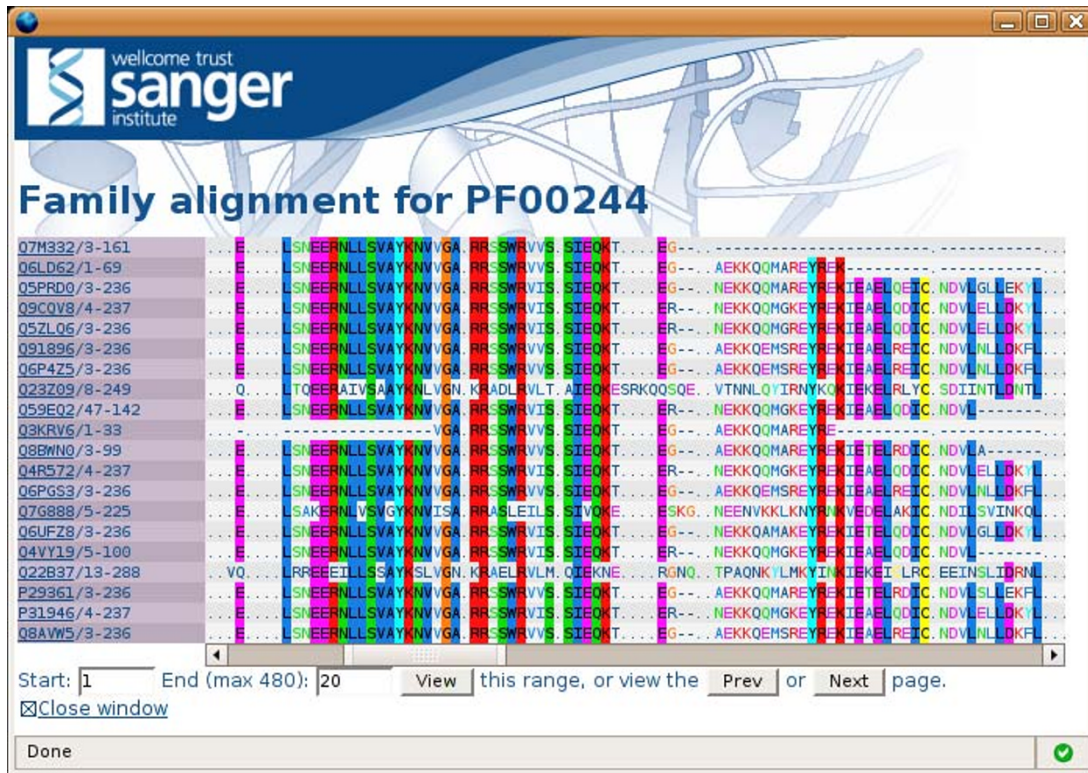
Figure 3
The Pfam DAS alignment viewer. Here, the first 20 sequences from the 14-3-3 [Pfam:PF00244] full alignment are requested from the pfamFullAlign DAS alignment server and displayed. The alignment is coloured according to the consensus of the whole alignment, which is provided as a DAS feature. The client allows simple paging through an alignment using the navigation buttons below the alignment.

feature command that is used to provide domain annotations of individual protein sequences, the Pfam multiple sequence alignments are also made available via DAS (using the alignment extension described above). One of the options in the specification allows the retrieval of one or more ranges from the requested query alignment. This option is of particular use when trying to view the very big Pfam alignments. For example, the Pkinase [Pfam:PF00069] *full* alignment contains 19113 sequences (Pfam release 21.0) making it, at best, very slow to visualise the alignment when loaded as a single file. By requesting ranges, the alignment visualisation is much faster. However, most alignment visualisation tools utilise some kind of colouring scheme to highlight regions of conservation based on the properties of the whole alignment. Therefore, to enable a region of an alignment to be placed in context of the whole alignment, Pfam also provides the alignment consensus via a DAS-*feature* request. The data from these two requests can be combined such that

requested alignment regions can be coloured according to similarity (see Figure 3).

Mapping of Genome Features on Protein structures

We have set up alignment DAS server that provide mappings from the Ensembl peptide sequences to Uniprot, and from Uniprot to PDB. The DAS client *SPICE*, is a browser of protein structures, sequences, and their annotations [29]. Based on the alignment data sources we extended SPICE, which now can easily show the location of genomic features, like SNPs or Intron/Exon boundaries mapped onto the three dimensional structure.

Multiple Structure Alignments

The alignment commands can be utilized to export pairwise and multiple protein structure alignments, a feature which has been used by the SISYPHUS database. SISYPHUS is a collection of proteins with non-trivial structural relationships and their alignments [32]. It provides a large
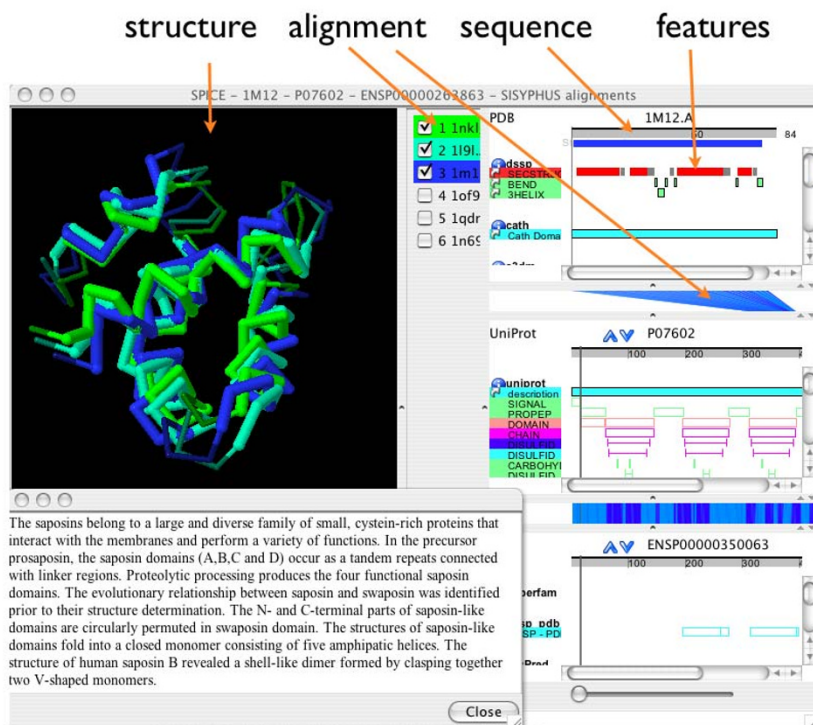
Figure 4
SPICE displays a multiple structure alignment of the Saposins, as provided by the SISYPHUS database. It is possible to choose which of the structures should be superimposed and sequence features can be mapped on the three dimensional alignment. The top of the image shows the different DAS commands that are being used in order to obtain the integration of the data.

amount of the data via DAS. The SISYPHUS multiple structure alignments can now be visualized using SPICE which has been recently extended to view multiple protein structure alignments (see Figure 4, [33]).

Another application of the new DAS structure and alignment commands was the CASP experiment [34], which has the goal to independently assess protein structure predictions. To do so, information about structures that have not been released to the public is passed to the structure prediction teams. These teams then provide models of the "unknown" structure.

To make evaluation and assessment of the predictions easier and more open, alignments of the predictions against the target were made available using DAS and visualized using SPICE. The different structures can be compared with each other and it is possible to switch between sequence dependent and independent alignment servers [35,36].

Discussion
DAS is an important system for exchanging and viewing biological data. The simple, but powerful original design of the DAS specification allowed scope for extension of the protocol. Here we presented several extensions to the core DAS specification that allow to manage, publish and discover the ever growing number of DAS sources and to deal with alignments and protein structure data, hence allowing new types of applications built on DAS. We envisage further extensions to the DAS protocol in the future. A new extension for electron microscopy data has recently been published by Macías et al. [37]. DAS has also received attention beyond the biological community and now even serves as a model in Astrophysics [38]. DAS has many similarities to webservices that are provided using the Simple Object Access Protocol (SOAP). SOAP based services have been made available by a number of institutions, e.g. [39-42]. They use XML requests and responses for the transport of information. In contrast to DAS there is no default definition how the transported

information should look like, although there are approaches to provide standards for some data types [43]. BioMoby is a system for interoperability between data providers, analysis services providers, and client implementations which is using SOAP as an envelope [44].

In contrast to SOAP based web services, DAS follows the paradigm of REST, the Representational State Transfer [45]. The basic difference is that DAS does not have an XML envelope for its messages, but uses well defined URLs for the DAS commands. The DAS responses are transported as XML documents. SOAP is frequently used to request remote calculation of data. In contrast the data provided with DAS is most of the time pre-calculated or can be computed quickly. Together with the Taverna project [46] we agreed on the addition of a "come back later" response for DAS which will allow to perform longer lasting computations triggered by DAS requests. This can be used to establish a bridge between DAS and Taverna workflows. A prototype implementation for this is available via the Dasobert Java library.

A concept sometimes used in connection with SOAP based webservice is the *Service Oriented Architecture* (SOA). A SOA consists of three components. A service provider, a consumer and a service registry. This concept can also be applied to DAS, with DAS servers being the service providers and DAS clients the consumers. By providing a registration server, we have extended DAS to become a full SOA.

Communication with the registry has been integrated into several web sites and applications. Ensembl, Pfam, Dasty, Jalview and SPICE now all can talk to the registry and users can access the registered DAS servers through different configuration interfaces.

Currently the version 2.0 of DAS specification is emerging [47]. We are actively participating in this development process. The *sources* specification for describing DAS services has been adopted as part of the core DAS/2 specification. DAS/2 is being defined with extensibility in mind and will allow the addition of protocols for sharing structure and alignment information.

Conclusion
The Distributed Annotation System is becoming widely accepted demonstrated by the fact that support for it is being built into more and more web sites and applications. Here we have presented several extensions that help the DAS system to scale, add support for new data types, and allow the development of new applications. Using these new DAS commands it has become possible to seamlessly walk from the genome level to the protein

structure level and to share sequence and structure alignments.

Availability and Requirements
• Project home page: http://www.dasregistry.org

• Operating system(s): Platform independent

• Programming language: Java, Perl

• Other requirements: Internet Browser

• License: LGPL

• Any restrictions to use by non-academics: none

Authors' contributions
AP implemented SPICE and the DAS registration server, extended the Dazzle server and wrote the paper. TD contributed several libraries, was involved in the design and planning of this project, is main author of the Dazzle server and helped with writing the paper. EK extended the Ensembl code base with new DAS functionality. RF is responsible for the Pfam website, for extending ProServer to support the new DAS commands and contributed some paragraphs to the paper. AP, TD, RF, TH designed the DAS extensions. AP, TD, RF, EK, AK set up and maintain various DAS servers. TH provided substantial advice and guidance during all phases of the project. All authors read and approved the final manuscript.

References
1. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2007, **35(Database):**D5-12.
2. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJA, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **New developments in the InterPro database.** *Nucleic Acids Res* 2007:D224-D228.
3. Tagari M, Tate J, Swaminathan GJ, Newman R, Naim A, Vranken W, Kapopoulou A, Hussain A, Fillon J, Henrick K, Velankar S: **E-MSD: improving data deposition and structure quality.** *Nucleic Acids Res* 2006:D287-D290.
4. Hubbard TJP, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D,

Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A, Birney E: **Ensembl 2007.** *Nucleic Acids Res* 2007, **35(Database issue):**D610-7.

5.   Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L: **The Distributed Annotation System.** *BMC Bioinformatics* 2001, **2:**7-7.
6.   **The Open-Bio DAS homepage** [http://www.biodas.org]
7.   **The BioSapiens project homepage** [http://www.biosapiens.info]
8.   Stein L, Sternberg P, Durbin R, Thierry-Mieg J, Spieth J: **WormBase: network access to the genome and biology of Caenorhabditis elegans.** *Nucleic Acids Res* 2000, **28:**82-86.
9.   Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: **The generic genome browser: a building block for a model organism system database.** *Genome Res* 2002, **12(10):**1599-1610.
10. Jones P, Vinod N, Down T, Hackmann A, Kahari A, Kretschmann E, Quinn A, Wieser D, Hermjakob H, Apweiler R: **Dasty and UniProt DAS: a perfect pair for protein feature visualization.** *Bioinformatics* 2005, **21(14):**3198-9.
11. **The ProView DAS client** [http://www.sanger.ac.uk/proview/]
12. Olason PI: **Integrating protein annotation resources through the Distributed Annotation System.** *Nucleic Acids Res* 2005:W468-70.
13. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20(3):**426-427.
14. Cases I, Pisano DG, Andres E, Carro A, Fernández JM, Gómez-López G, Rodriguez JM, Vera JF, Valencia A, Rojas AM: **CARGO: a web portal to integrate customized biological information.** *Nucleic Acids Res* 2007.
15. Holm L, Sander C: **Protein structure comparison by alignment of distance matrices.** *J Mol Biol* 1993, **233:**123-138.
16. Shindyalov IN, Bourne PE: **Protein structure alignment by incremental combinatorial extension (CE) of the optimal path.** *Protein Eng* 1998, **11:**739-747.
17. Lackner P, Koppensteiner WA, Sippl MJ, Domingues FS: **ProSup: a refined tool for protein structure alignment.** *Protein Eng* 2000, **13:**745-752.
18. Zemla A: **LGA: A method for finding 3D similarities in protein structures.** *Nucleic Acids Res* 2003, **31(13):**3370-3374.
19. Finn RD, Stalker JW, Jackson DK, Kulesha E, Clements J, Pettett R: **ProServer: A simple, extensible Perl DAS server.** *Bioinformatics* 2007, **23(12):**1568-1570.
20. **The Dazzle DAS server library (Java)** [http://www.derkholm.net/thomas/dazzle/]
21. **Documentation of the new DAS extensions** [http://www.dasregistry.org/help_extensions.jsp]
22. **The DAS registration server** [http://www.dasregistry.org]
23. **The new DAS-sources command at the DAS registration server** [http://www.dasregistry.org/das1/sources]
24. **The location of the SOURCES command for Ensembl DAS sources** [http://www.ensembl.org/das/sources]
25. **The location of the WSDL for the DAS registry SOAP web services** [http://www.dasregistry.org/services/das:das_directory?wsdl]
26. **The ProServer DAS server library (Perl)** [http://www.sanger.ac.uk/proserver/]
27. **The LDAS – DAS server library (Perl)** [http://biodas.org/download/ldas/]
28. **The Generic Genome Browser (GBrowse)** [http://www.gmod.org/]
29. Prlić A, Down TA, Hubbard TJP: **Adding Some SPICE to DAS.** *Bioinformatics* 2005, **21(Suppl 2):**ii40-ii41.
30. **How to configure data plots via DAS stylesheets** [http://www.dasregistry.org/extension_stylesheet_histogram.jsp]
31. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A: **Pfam: clans, web tools and services.** *Nucleic Acids Res* 2006:D247-D251.
32. Andreeva A, Prlic A, Hubbard TJP, Murzin AG: **SISYPHUS-structural alignments for proteins with non-trivial relationships.** *Nucleic Acids Res* 2006, **35(Database issue):**D253-9.
33. **The SISYPHUS website** [http://sisyphus.mrc-cpe.cam.ac.uk]

34. Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A: **Critical assessment of methods of protein structure prediction (CASP)-round 6.** *Proteins* 2005, **61(Suppl 7):**3-7.
35. Kryshtafovych A, Prlic A, Dmytriv Z, Daniluk P, Milostan M, Eyrich V, Hubbard T, Fidelis K: **New tools and expanded data analysis capabilities at the protein structure prediction center.** *Proteins* 2007.
36. **CASP-7 results** [http://www.predictioncenter.org/casp/casp7/public/cgi-bin/results.cgi]
37. Macías JR, Jiménez-Lozano N, Carazo JM: **Integrating electron microscopy information into existing Distributed Annotation Systems.** *J Struct Biol* 2007, **158(2):**205-213.
38. Bose R, Mann R, Prina-Ricotti D: **AstroDAS: Sharing Assertions across Astronomy Catalogues through Distributed Annotation.** *Proceedings of the International Provenance and Annotation Workshop (IPAW'06), Chicago, IL. USA Lecture Notes in Computer Science (LNCS)* 2006, **4145:**193-202.
39. Pillai S, Silventoinen V, Kallio K, Senger M, Sobhany S, Tate J, Velankar S, Golovin A, Henrick K, Rice P, Stoehr P, Lopez R: **SOAP-based services provided by the European Bioinformatics Institute.** *Nucleic Acids Res* 2005:W25-W28.
40. Sugawara H, Miyazaki S: **Biological SOAP servers and web services provided by the public sequence data bank.** *Nucleic Acids Res* 2003, **31(13):**3836-3839.
41. Coessens B, Thijs G, Aerts S, Marchal K, Smet FD, Engelen K, Glenisson P, Moreau Y, Mathys J, Moor BD: **INCLUSive: A web portal and service registry for microarray and regulatory sequence analysis.** *Nucleic Acids Res* 2003, **31(13):**3468-3470.
42. Fernández JM, Hoffmann R, Valencia A: **iHOP web services.** *Nucleic Acids Res* 2007, **35(Web server issue):**W21-6.
43. Seibel PN, Krüger J, Hartmeier S, Schwarzer K, Löwenthal K, Mersch H, Dandekar T, Giegerich R: **XML schemas for common bioinformatic data types and their application in workflow systems.** *BMC Bioinformatics* 2006, **7:**490.
44. Wilkinson MD, Links M: **BioMOBY: an open source biological web services proposal.** *Brief Bioinform* 2002, **3(4):**331-341.
45. Fielding RT: **Architectural styles and the design of network-based software architectures.** In *PhD thesis* UNIVERSITY OF CALIFORNIA, IRVINE; 2000. [Chair-Richard N. Taylor]
46. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T: **Taverna: a tool for building and running workflows of services.** *Nucleic Acids Res* 2006, **34:**W729-W732.
47. **The DAS/2 specification** [http://biodas.org/documents/das2/das2_protocol.html]
48. Käll L, Krogh A, Sonnhammer ELL: **An HMM posterior decoder for sequence feature prediction that includes homology information.** *Bioinformatics* 2005, **21(Suppl 1):**i251-i257.
49. **SPICE – data plot examples** [http://www.efamily.org.uk/software/dasclients/spice/histogram.shtml]