



Full length article

SAMP, the Simple Application Messaging Protocol: Letting applications talk to each other

M.B. Taylor^{a,*}, T. Boch^b, J. Taylor^{c,1}^a H. H. Wills Physics Laboratory, University of Bristol, UK^b Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, F-67000, Strasbourg, France^c Institute for Astronomy, Royal Observatory, Blackford Hill, University of Edinburgh, UK

ARTICLE INFO

Article history:

Received 31 October 2014

Accepted 21 December 2014

Available online 2 February 2015

keywords:

Interoperability
Message-passing
Publish–subscribe

ABSTRACT

SAMP, the Simple Application Messaging Protocol, is a hub-based communication standard for the exchange of data and control between participating client applications. It has been developed within the context of the Virtual Observatory with the aim of enabling specialised data analysis tools to cooperate as a loosely integrated suite, and is now in use by many and varied desktop and web-based applications dealing with astronomical data. This paper reviews the requirements and design principles that led to SAMP's specification, provides a high-level description of the protocol, and discusses some of its common and possible future usage patterns, with particular attention to those factors that have aided its success in practice.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Astronomical research requires complex and flexible manipulation and processing of various different types of data. Images, spectra, catalogues, time series, coverage maps and other data types need their own special handling, typically provided by specialist tools. Data sets of different types meanwhile are usually related in various ways arising from their physical origin, for instance catalogues are often derived from images and best understood in conjunction with them, and spectra and time series usually originate from specific sky positions or regions which may be represented on images and described by catalogue entries. To extract scientific meaning from the data it is usually necessary to exploit these linkages between data items as well as the internal structure of each.

The working astronomer therefore uses a selection of different software components, each specialising in a particular type of data or manipulation, for different data sets and different tasks, and has to integrate these together in a way that takes account of the relationships of the data items under consideration.

For batch or pipeline-type processing the required tool integration is usually, in terms of data flow, fairly straightforward: the output of one step can be fed to the input of the next as a file, stream

of bytes, or some kind of parameter list, often under the control of a script of some kind.

During the exploratory or interactive phase of data analysis however, this traditional model of tool integration is less satisfactory. Within a given GUI analysis application it is usual to interact with the data using mouse and keyboard gestures to perform actions like selection or navigation with instant visual feedback, in many cases with some kind of internal linkage between different data views. But communicating such actions or their results between different tools tends to be much more cumbersome. A way can often be found to reflect a result generated by one tool in the state of another, for instance by reading sky coordinates reported by one tool and typing or pasting them into another, or saving an intermediate result from one tool to temporary storage and reloading it into another, but it can be fiddly and tedious, especially if similar actions are required repeatedly. This lack of convenience is more than just an annoyance, it can interrupt the flow of the data exploration, reduce the parameter space able to be investigated, and effectively stifle discovery of relationships present in the data.

From this point of view, a single monolithic astronomical data analysis user application providing the best available facilities for interactive presentation, manipulation and analysis of all kinds of astronomical data and their interrelationships seems an attractive prospect. In reality of course, no such one-stop analysis tool exists. The obvious practical difficulties aside, it is not even clear that deviating so far from the Unix philosophy of “Make each program do one thing well” (McIlroy et al., 1978) would be desirable.

* Corresponding author.

E-mail address: m.b.taylor@bristol.ac.uk (M.B. Taylor).¹ Present address: Google, USA.

These considerations have driven the development of a framework for communication between independently-developed software items, written in different languages and running in different processes. Such applications can thus be made to appear to the user as a loosely integrated suite of cooperating tools, providing facilities such as data exchange, linked views and peer-to-peer or client-server remote control. Although communication between interactive desktop tools was the original stimulus for what is now SAMP, the framework is flexible enough to support other usage patterns as well.

Two previous papers on SAMP have been presented in the ADASS conference series: Taylor et al. (2012a) briefly outlines the architecture and explains the Web Profile, and Fitzpatrick et al. (2013) lists some existing client libraries. The current paper discusses the protocol, its communication model, and its current usage in sufficient detail to understand the design decisions taken and their consequences, particularly from the point of view of the usage scenario outlined above. Section 2 traces the evolution of SAMP from its predecessor PLASTIC alongside a comparison with some alternative messaging systems, Section 3 outlines some high-level design principles, Section 4 presents a description of the protocol itself along with some of the thinking behind it, Section 5 considers its use in practice, and Section 6 concludes by reviewing the current status and possible future directions for SAMP, as well as the factors that have encouraged its uptake. For the complete and definitive details of the protocol, the reader is referred to the standard document itself (Taylor et al., 2012b).

2. Context

2.1. History

In the context of the emerging Virtual Observatory in the mid-2000s, the benefit of connecting client-side tools to improve productivity when working with multiple data types became apparent. In fact this problem was not specific to the VO, but the ease with which multiple related data products could be acquired using VO technologies, themselves sometimes requiring the use of separate tools for data discovery and acquisition, amplified the benefits that such tool integration could deliver. Additionally, the new shared funding and communications channels between institutionally and geographically separated software developers that arose from various VO initiatives proved important in practice as a platform for experimentation and agreement in this area.

The external scripting capabilities of tools such as Aladin (Bonnarel et al., 2000), SPLAT-VO (Škoda et al., 2014) and SAOImage ds9 (Joye and Mandel, 2003) already provided the option of tightly coupled master/slave control between pairs of applications, but did not lend themselves to the kind of cooperative interaction envisaged. The developers of Aladin experimented with Java interfaces designed for two-way communications; these delivered some limited integration, but were restricted to applications operating within the same Java virtual machine. Meanwhile the Astro Runtime (Winstanley et al., 2007) developed by AstroGrid was providing to desktop tools a simplified façade for a range of VO services using their choice of communication technology (XML-RPC, REST, Java RMI or JVM call).

From this background, in 2005 discussions between developers of the AstroGrid, Aladin, VisIVO (Comparato et al., 2007) and TOP-CAT (Taylor, 2005) software in the context of the Euro-VO framework and the SC4DEVO workshop series led to the development of a new communication protocol PLASTIC: the Platform for Astronomical Tool InterConnection (Taylor et al., 2007; Boch et al., 2006).

PLASTIC built on the implementation and technology choices present in the Astro Runtime to provide the interaction capabilities required by the participating teams, and prototyped many features that were later inherited by SAMP, including a central hub, publish-subscribe messaging, use of XML-RPC, loosely-defined message semantics, and a pragmatic approach to providing “good-enough” communications. It proved popular with developers and users, and was incorporated into a dozen or so desktop applications, which could thereby be used together effectively in productive and sometimes novel ways.

Interest in PLASTIC was however largely confined to Europe. Efforts to gain IVOA endorsement and expand the pool of applications that could communicate in this way led after some discussion to the drafting by European and US authors of a successor standard, the Simple Application Messaging Protocol, which was accepted as an IVOA Recommendation in 2009 (SAMP version 1.11). This standard was intentionally similar in many respects to PLASTIC, in order to avoid disrupting patterns of successful cooperation already in use, but the opportunity was taken to amend some decisions that experience had shown to be sub-optimal, and to expand its scope to accommodate other possible usage patterns. Changes made on the basis of lessons learned from PLASTIC included a simplification of the type system, complete language independence (though PLASTIC could be used from any language, certain parts of the protocol were defined with reference to Java), simplification of message targeting, improved security arrangements (security is still rudimentary in SAMP, but opportunities for trivial client spoofing were removed), modification of message names (now both human-readable and wildcard-able rather than opaque URIs), definition of all message parameters and return values as key-value pairs rather than ordered lists, use of fundamentally asynchronous messaging for robustness, restriction rather than proliferation of transport mechanisms, improved error reporting, and better extensibility.

Also new in SAMP was the notion of a Profile to provide formal separation between the abstract messaging model and the transport layer. One reason for its introduction was to enable the possible future use of the protocol for messaging in less “PLASTIC-like” contexts. At the time, requirements for improved performance or security were envisaged; to date extensions in those directions have not been explored, but the Profile mechanism has paid off by supporting the later development of the Web Profile to support browser-based clients alongside desktop ones. The introduction of the Web Profile in SAMP version 1.3 (2012) has been the main change to date since the initial version.

2.2. Other messaging systems

Many general- and special-purpose messaging frameworks exist. It is beyond the scope of this paper to provide a comprehensive review, but we provide here a brief comparison between SAMP and a few of the alternatives.

Several generic messaging frameworks share some features with SAMP, for example AMQP, ZeroMQ, XPA, and D-Bus. To our knowledge, none satisfy SAMP’s key requirements of an easily implementable platform-neutral standard supporting straightforward messaging between a shared community of clients in quite the way required, though some of these systems could be used as transport layers on which future SAMP profiles could be built, in the same way that XML-RPC has been used in the existing profiles. SAMP deliberately restricts some choices related to implementation and usage to reduce the burden on client developers, so it is perhaps not surprising that generic messaging frameworks are not, on their own, appropriate.

One or two messaging systems however merit further mention. WAMP, the Web Application Messaging Protocol,² bears some striking architectural similarities to SAMP including a combination of RPC and publish–subscribe messaging mediated by a central component known in WAMP terminology as the Router. However, it does not address the issue of router discovery, so there is no prescribed way for clients to initiate communication.

The Intent mechanism for inter-process communication that forms part of the Android operating system, while its target environment clearly differs from that of SAMP, shares with it some characteristics in terms of design and usage. For instance message semantics may be defined in a way which is either app-specific (“explicit”) or vague (“implicit”), in the latter case resulting in a user choice at runtime between candidate receiving apps. Furthermore, bulk data transfer is achieved using URIs to refer to an external data source rather than conveying the payload within the message.

An example of a domain-specific messaging framework is the Systems Biology Workbench (Sauro et al., 2003), which is close in spirit to SAMP, enabling platform-independent remote method invocation between components (known as “modules”) in the field of systems biology, based around the SBML data format. One point of difference is that the SBW infrastructure itself orchestrates the loading of modules to provide the required functionality (the Android OS does something similar with Intents); in SAMP this choice of components is left to the user, though components like Applauncher (Lafrasse et al., 2012) are available to layer automatic client startup on top of the basic protocol where desired.

3. Design for interoperability

The overriding objective for the design of SAMP has been to foster interoperability in practice. This requires not just a messaging system with sufficient communication capabilities, but also one which developers of popular analysis tools, and ideally private scripts as well, are actually willing and able to integrate into their software. To achieve this, a number of principles have been followed.

In the first place, it is as far as possible platform independent. The definition of the protocol is not dependent on or biased towards use of particular implementation languages or operating systems.

Second, ease of adoption. Application authors have found basic use of SAMP to require little implementation effort. In practice, availability of easily deployable SAMP client libraries developed within the SAMP community for a number of implementation languages have been an important factor in this. However the communications are, by design, simple enough that basic SAMP use is not hard to achieve given only an XML parser and HTTP access capabilities. Ease of use by end users is equally important, so that those running analysis tools can benefit from the integration capabilities that SAMP provides without needing to perform expert configuration (ideally, any explicit setup at all) or to understand the details of the messaging system.

Third, extensibility and flexibility. Building into the system the capability to use it in ways driven by the requirements of the client tools rather than just those foreseen by the standard authors increases its likely usefulness. The mechanisms for extensibility have been particularly designed to allow the introduction of new features without affecting existing ones, with the aim of reducing compatibility issues.

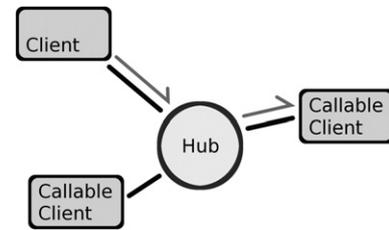


Fig. 1. Schematic of SAMP Hub and clients joined in a star topology. Black lines indicate clients registered with the hub. The red half-arrows indicate the progress of a message from a sending client (which may or may not be callable) to a receiving client (which must be callable), passing through the Hub.

Finally, the approach has been above all pragmatic, favouring the straightforward over the rigorous in cases of conflict. For instance message delivery is not guaranteed, but can be expected to work most of the time. The security model will prevent casual interference, but may be vulnerable to determined attack. Semantics are tagged using short readable strings on the assumption of sensible choices, rather than URIs with guaranteed private namespaces. Performance is easily good enough to handle exchange of short control messages on a timescale commensurate with user actions, but not for sustained throughput of high data volumes. It may however be noted that most of these items could if required be ameliorated by future introduction of a new Profile with different transport characteristics.

These principles and their application, in some cases informed by positive and negative lessons from the experience of PLASTIC, might not be appropriate for all contexts but have led to a messaging infrastructure which ought to be easy for client developers to understand and adopt, and which has in fact been widely taken up.

4. Protocol description

SAMP is based on a star topology, and its central component is a Hub through which all communications are passed (Fig. 1). Clients first perform a resource discovery step to locate the Hub, and then register with it, establishing a private communication channel through which subsequent calls to the Hub’s services can be made. These services include accepting metadata about the registering client, providing information about other registered clients, and forwarding messages to those clients. These messages may elicit responses, which may optionally be passed back to the message sender, again via the Hub. All registered clients are able to send messages in this way. Any client may optionally declare itself callable, in which case it is also able to receive messages sent by others. Callability is optional since it is more difficult to achieve in client code, requiring some server-like capacity on top of the ability to invoke Hub services, and simple actions like sending an image or table can be achieved without this requirement. In addition to declaring itself callable, a client wishing to receive messages must explicitly subscribe³ to one or more MTypes (message types). Every message is labelled with an MType, and the Hub will only deliver messages to clients that have declared their interest in the MType in question with an appropriate subscription. When sending messages, clients may either broadcast them to all subscribed clients or target them to a named client, but in the latter case delivery will fail if the target client has not appropriately subscribed. If a client has no further use for SAMP communications (for instance on application exit), it can and should unregister.

³ The term “subscription” derives from the “publish–subscribe” messaging pattern. It may however be more helpful to think of subscription as declaring support for a particular message type.

² <http://wamp.ws/>.

This framework combines the notions of publish–subscribe (pub/sub) and Remote Procedure Call (RPC) messaging. Like publish–subscribe, messages are only delivered to appropriately subscribed recipients, but like RPC the sender may optionally target messages to a selected recipient, and may optionally receive responses from the recipient(s). The targeting mode, response requirement, and message content are all decoupled from each other.

The details of this system are codified in a three-layer architecture:

Abstract API: defines the services provided by the Hub and clients

Profile: maps the Abstract API to specific communication operations, such as bytes on the wire

MTypes: provide semantics for the actual messages exchanged between clients.

Note that SAMP thus defines two distinct sets of Remote Procedure Call (RPC) operations: the functions declared by the Abstract API, concerning the mechanics of client–hub communication and message delivery, and SAMP Messages themselves classified by MType, bearing the application-level content that clients wish to exchange with each other. The syntax and semantics of the former are carefully defined by the SAMP standard, but the form and content of the latter are agreed outside of SAMP itself by cooperating client developers.

Because of the central rôle of the Hub in this pattern, it presents a single point of failure and potential bottleneck. However, SAMP messages are usually short, and in practice performance issues have not generally been apparent.

The following subsections present a more detailed account of these ideas, along with some of the considerations that influenced their design. Sections 4.1–4.3 describe the three architectural layers listed above, and Sections 4.4 and 4.5 describe the underlying type system and how it is used to underpin extensibility in SAMP.

4.1. Abstract API

The Abstract API defines the messaging capabilities of SAMP. It takes the form of a list of a dozen or so function definitions with typed arguments and return values, and well-defined semantics. Most of these functions represent services provided by the Hub, for instance `register` (which returns information required by the client for future communications, typically an identification token) and `notifyAll` (which requests forwarding a given message to all appropriately subscribed clients). The remainder represent services required from callable clients, such as `receiveNotification` (which consumes a given message originating from another client).

The messaging model in principle associates a response with every message, containing at least a completion status flag along with zero or more MType-defined return values. However it is up to the sending client whether a response is required from any given message; in many cases the status flag is the only return value, and in this case a sending application may or may not wish to make the effort to pass this on to its user (for instance “the table you sent was/was not successfully received”). If the sender has no interest in the return value, it can use the “send-and-forget” (*notification*) pattern, with lower cost for sender, recipient and hub.

Message processing is fundamentally asynchronous from the receiver’s point of view, so that message/response times are not limited to the lifetime of an RPC call in the underlying transport mechanism. However, the Hub provides an optional synchronous façade for sending messages when clients expect fast turnaround and wish to avoid the additional complication of asynchronous processing.

4.2. Profile

A particular SAMP Profile is what turns the Abstract API into a set of rules that a client can actually use to communicate with a running Hub, and hence with other clients.

It performs two main jobs: first, it describes how the functions defined by the API are turned into concrete communication operations, by specifying an RPC-capable transport mechanism and rules for mapping the SAMP data types into the parameters and responses used by that mechanism. Second, it defines a hub discovery mechanism, which tells clients how to establish initial communications with the Hub, usually involving some authentication step. Particular profiles may also specify additional profile-specific hub or client services exposed as functions alongside those mandated by the Abstract API.

Initially (SAMP 1.11, 2009) only a single profile was defined, the Standard Profile. This uses XML-RPC⁴ as a transport mechanism, and allows hub discovery by storing the URL of the hub’s XML-RPC server along with a secret randomly generated key in a private “lockfile” in the user’s home directory.

Version 1.3 of the standard (2012) introduced a second, the Web Profile, for use by web-based clients. This is required for applications running within web pages, since the sandboxed environment imposed by the browser makes the Standard Profile inaccessible. It shares use of XML-RPC and some other characteristics with the Standard Profile, but hub discovery has to be done differently, and there are a number of complications to do with security, described in Taylor et al. (2012a) as well as the Standard.

This decoupling between the functionality of the service interface and its incarnation in a specific transport mechanism allows different transports to be introduced without changes to the core protocol or existing clients, and has a number of benefits. In a given SAMP session, a client may use the most appropriate Profile for its SAMP communications and exchange messages seamlessly with other clients using different profiles; a desktop application can exchange messages with a web page just as easily as with another desktop client. This works because clients only ever communicate directly with the Hub and not with each other, while the Hub performs lossless translation between profile-specific network operations and the messaging model defined by the Abstract API.

Future requirements may result in additional Profile definitions, and there is nothing in principle to prevent hub developers from implementing new ones outside the frame of the SAMP standard. However, from an interoperability point of view it is important that all profiles are supported by all common Hub implementations, so that a client can rely on the availability of a chosen profile in a SAMP environment, and for this reason unnecessary proliferation of profiles is discouraged.

4.3. MTypes

An MType (message type) is the description for a message with particular syntax and semantics. It is analogous to a function definition in an API, and consists of a labelling string (sometimes itself also known as the MType) along with a set of zero or more typed and named arguments, a set of zero or more typed and named return values, and some associated semantics indicating what the sender of such a message is trying to convey.

By way of example, a commonly used MType is `image.load.fits`, defined like this:

⁴ XML-RPC is a simple protocol for remote procedure calling based on HTTP and XML. It resembles a very much slimmed-down SOAP. Documentation can be found at <http://www.xmlrpc.com/>.

Name:`image.load.fits`**Semantics:**

Loads a 2-d FITS image

Arguments:`url` (string):

URL of the image to load

`image-id` (string, optional):

Identifier for use in subsequent messages

`name` (string, optional):

Name for labelling loaded image in UI

Return Values:

None.

The name is a short hierarchical string composed of atoms separated by the “.” character. As well as identifying to a recipient the type of an incoming message, it is used by clients to *subscribe* to messages, that is to indicate to the Hub which messages they are prepared to receive. For the purpose of subscription a limited wildcarding syntax is available, so by using the MType patterns `image.load.fits`, `image.*` or `*.*` a client may declare interest in only the above message, or all image-related messages, or all messages, respectively.

In general, a callable client will only subscribe to those MTypes on which it can meaningfully act, so for instance an image analysis tool typically would subscribe to `image.load.fits`, but not to `spectrum.load.ssa-generic`. A client that has an image FITS file to send can then either query the Hub for those clients subscribed to the image load message and offer its user the choice of which one to target, or request the Hub to *broadcast* the image load message to all (and only) the image-capable clients.

4.4. Type system

Supporting the function list defined by the Abstract API and the parameters and return values specified by MTypes is a type system defining the types of value permitted, as well as rules for encoding various structured objects using these types: message objects themselves, success and failure message responses, application metadata, and MType subscription lists.

This system contains only three types: *string*, *list* and *map*. A string is a sequence of 7-bit ASCII printable characters, a list is an ordered sequence of strings, lists or maps, and a map is an unordered set of associations of string keys with values of type string, list or map. Structured objects are specified by the use of well-known keys in maps, there is no special representation for null values, and non-string scalar types must be serialised as strings. (Obvious) conventions are suggested for serialising integer, floating point and boolean values into string form, but these suggestions are provided for the convenience of MType definitions that wish to exchange such values without reinventing the wheel, and are not a normative part of the protocol.

This restricted type system has been deliberately chosen to introduce minimal dependency of messaging behaviour on the details of non-core parts of the delivery system, in particular profile-specific transport mechanisms and language-specific client libraries. This both reduces the restrictions on what languages and transport layers may be used with SAMP, and ensures that values transmitted will not be modified during processing by parts of the messaging system outside of client control.

The type system is rich enough to represent complex structured data where required, but note it is not intended for use with binary data, and transmission of bulk data or large payloads in general is discouraged within SAMP messages in favour of passing URLs around instead, meaning that client and Hub implementations can work on the assumption of short message payloads.

This convention of out-of-band bulk data transfer does place an additional burden on sending clients however, since to transmit a bulk data item (such as a table or image) not already available from an existing URL it is necessary to make it so available, for instance by writing bytes to a temporary file or serving them from an embedded HTTP server. It can also present complications if the sending and receiving client are not able to see the same URLs, for instance due to different security contexts; in this case, additional Hub services may be required to assist with data transfer between domains (accordingly, the Web Profile provides services to assist with cross-domain data exchange).

Note also that the string type does not natively accommodate Unicode text, including XML. The restriction to 7-bit ASCII is driven by the requirement for use from non-Unicode-capable environments such as Fortran, IDL and some shell scripting languages. This has not caused known problems to date, but inability to handle Unicode text without additional encoding could prove awkward in some cases, and it may be necessary to revisit this restriction in a future revision of the standard.

4.5. Extensible vocabularies

Extensibility is built into this system via the notion of an *extensible vocabulary* used when representing structured objects.

Structured objects are represented as maps with well-known keys, but the rule is that additional keys are always permitted, and that hubs and clients must ignore any keys they do not understand, propagating them to downstream consumers where applicable (compare the NDF extension architecture described in [Jenness et al. \(in press\)](#)). A corollary is that such non-well-known keys must be defined in such a way that ignoring them will result in reasonable behaviour. The Abstract API tends to prefer *maps* (unordered name/value pairs) over ordered parameter value lists, which makes this extensibility pervasive throughout the messaging system, applying for instance to client metadata and subscription declarations, message transmission information, and MType-specified message parameter lists and return values.

For instance, a client sending a message must pass it to the Hub as a map with two required keys: `samp.mtype` giving the MType label and `samp.params` giving the MType-specified parameter list (itself a map). But a client may optionally insert additional non-standard key/value pairs into that map, for instance using a non-standard key *priority* to associate a particular priority level with the message. If the Hub happens to support this non-standard feature, it is able to treat the message specially in view of this declaration; in any case it will propagate the message to recipient clients with the additional entry present, so if one of those supports this feature then it may use the value in processing. The same rule applies for instance to the MType-determined message parameter list; an MType like `image.load.fits` has a required parameter `url`, but a sending client may add a non-standard parameter like `colormap` (or `ds9.colormap`) alongside the well-known ones for the benefit of any client that happens to support it. Clients can therefore piggy-back experimental or application-specific instructions on top of generic messages to achieve more detailed control where available, falling back to the baseline functionality if it is not. Using this extensibility pattern, new or enhanced features of particular MTypes or of the protocol itself can be prototyped very easily, requiring no changes to the SAMP standard or infrastructure implementation beyond those components actually using the non-standard features, and imposing no negative impact on existing messaging operations. If they are found to be useful, they may be adopted in the future as (most likely optional) well-known keys alongside the original ones.

Some associated namespacing rules apply. Well-known keys defined by the SAMP standard are in the reserved `samp` namespace, meaning they begin with the string “`samp.`”. When

introducing non-standard keys it is not permitted to use this namespace, but any other syntactically legal string is allowed. The special namespace `x-samp` is available for keys proposed for future incorporation into the standard, and hubs and clients should treat keys which differ only in the substitution of `samp` for `x-samp` as identical, to ease standardisation of prototype features. In the case of MType parameters and return values, which are mostly not defined by SAMP itself, there is no reserved namespace.

5. Use in practice

The protocol described above is capable of supporting a wide range of different messaging patterns. For use in a particular scenario, a number of practical considerations must also be worked out. This section discusses how the framework has been applied to date to support the original goal of helping to integrate data analysis tools used by astronomers.

Section 5.1 explains how hub provision is managed, Section 5.2 describes some common patterns of message semantics, and Section 5.3 addresses the social mechanisms by which these are agreed on by the SAMP community. Section 5.4 reviews the existing landscape of SAMP infrastructure software and SAMP-aware tools, and Section 5.5 provides some concrete examples of it in action.

5.1. Hub provision

SAMP's star topology means that a Hub (in most cases, exactly one Hub) must be running for any messaging to take place. Ideally, an independent Hub process would be started as part of user session setup to ensure its constant availability. While this is quite possible and appropriate in some scenarios, even the minimal configuration required to establish it (a hub startup line in a session startup file) requires the kind of explicit user action which cannot always be relied upon. Simply put, if the functionality does not show up in the user interface with zero user effort, most users will never discover it.

It is common practice therefore, though by no means a requirement, for some SAMP-aware tools to come with an embedded hub. In this scenario, when a SAMP-aware tool starts up, it first checks for a running hub. If one exists, it registers with it; if not, and if it has the capability, it starts its own embedded Hub, and registers with that. Note that a client running an embedded Hub communicates with it in just the same way as with an independent one, it has no privileged access. Non-hub-capable clients may choose to check for a running Hub and connect on startup, on explicit user request, or when periodic polling indicates that one has become available. The effect is that usually when two or more SAMP-aware tools are running, a Hub will be present and those tools will find themselves connected to it, enabling messaging. Sometimes the application hosting the embedded hub will be shut down during a session taking the Hub down with it, and in that case another application may notice the fact and start one up, at which point some or all previously registered clients may notice the new hub and re-register with it.

This somewhat haphazard model of hub provision does not form a robust platform for high-reliability messaging, but, in accordance with SAMP's pragmatic approach, operates well enough most of the time, with a minimum of user effort; usually, it "just works". Note that where explicit control of an independent hub process can be arranged, for example as part of a managed user environment, more robust connectivity will result; an example is the Herschel Interactive Processing Environment (Balm, 2012).

5.2. MType semantics

A messaging framework only serves any purpose if there exists a vocabulary of messages understood by the applications which are going to exchange them. In SAMP terms that means establishing a collection of more or less well-known MTypes (Section 4.3). Choosing the right semantics for this collection is crucial to the utility and character of the messaging system in practice.

The most obvious approach for providing message-based control of an application is to identify (at least some of) the capabilities it offers and define a messaging interface with parameters and return values exposing those capabilities. An image display application might expose a set of MTypes allowing image load into a new window, zoom configuration, colour map choice, WCS display and so on. This allows other applications to control its behaviour in detail and is suitable for tight integration of a known set of tools with a good understanding of each other's capabilities, for instance to execute a pre-orchestrated sequence of processing steps.

However, this approach is less effective in less predictable environments. The controlling client needs to understand the capabilities of its partner client in order to control it. But if the set of tools in use at runtime is chosen by a user from an open-ended set rather than mandated by a developer, the identity of the partner client or clients is not known in advance. In general, different applications even of similar types have different capability sets and internal data models, and these cannot readily be encompassed by any single general abstraction. Different image display tools may support different data formats, may or may not support multiple loaded windows or images, may specify zooms in different ways, may offer different selections of colour maps, may provide WCS display with different options or not at all and so on, and the burden on a client wishing to control a range of different recipients quickly becomes large. Even if an application developer is prepared to study the messaging APIs offered by existing available tools and implement logic managing message dispatch for each case, the resulting code will not cope with applications unknown to the developer, for instance ones yet to be written.

For uncontrolled environments in which the user selects the range of cooperating tools at runtime therefore, a "loose integration" model has turned out to be more successful. This approach focuses on a messaging interface consisting of a fairly small number of MTypes with semantics that are non-client-specific and rather vague. The semantics of the most-used messages generally boils down to "Here is an X", where X may be some resource type such as a table, image, spectrum, sky position, coverage region, bibcode etc., or sometimes a reference to an X from a previous message, for instance a row selection relating to an earlier-sent table. The implication of such an MType is that the receiver should do something appropriate with the X in question: load, display, highlight, or otherwise perform some action which makes sense given the receiving application's capabilities. Callable SAMP clients should therefore advertise themselves (by subscribing to the appropriate MTypes) as X-capable tools only if they are in a position to do something sensible with an X should they be presented with one. Such an advertisement serves as a hint to potential X-senders, though it does not constitute a guarantee of any particular behaviour. This framework typically manifests itself in a client user interface as an option, for an X currently known by that client, either to *broadcast* it to all X-capable clients, or to target it to an entry selected by the user from a dynamically-discovered list of X-capable clients (see Fig. 3 for an example). For this kind of usage, the presence of a human in the loop to direct messages between clients as required by a particular workflow is an important part of the system, but the decisions required from the user are generally simple ones, e.g. which of a small list of clients to contact.

For clients to interoperate as reliably as possible in this scenario, it is not sufficient just to agree on the notion of a table or

an image for exchange, it is also necessary to specify the exact data exchange format. In the case of tabular data, a variety of possible exchange formats is in common use: FITS binary and ASCII tables, VOTable, Comma-Separated Values and a host of others including many ASCII-based variants. Different choices are convenient in different usage contexts, suggesting the need for a variety of distinct format-specific MTypes. However, a proliferation of alternative exchange formats, though superficially convenient, erodes interoperability. If multiple exchange formats capable of serialising the same thing are available, the sender has to choose which to send, and the receiver may or may not be able to receive it. Well-behaved recipients should include conversion code for as many formats as possible, and well-behaved senders should send data in a format dependent on what is supported by the intended recipient. For applications willing to expend a lot of effort on interoperability the work required at both ends increases rapidly with the number of available formats, while less conscientious implementations may find themselves unable to exchange data of essentially compatible types, or the community of SAMP clients may fragment into format-specific sub-communities unable to communicate globally. As much as possible therefore, it is desirable to restrict the options to a single well-defined exchange format for each basic data type.

This can be a difficult balance to get right. In the case of images, astronomy is fortunate that FITS serves as the *lingua franca*, and the only commonly-used MType is `image.load.fits`. For tabular data, clients are strongly encouraged to use `table.load.votable` even if it means translating to/from some other format; however other `table.load.*` variants are in use for specialist purposes, for instance for the CDF format,⁵ which though tabular is not readily translated to VOTable without loss of information, and which tends to be used in communities not familiar with VOTable. In the case of spectra, for various reasons related to the form in which spectral data is typically obtained and the typical capabilities of spectrum-capable clients, the relevant MType is `spectrum.load.ssa-generic`, which permits any format to be used for the spectral data, with additional parameters specifying the format actually in use.

SAMP is capable of supporting both tight and loose integration, and both are in use, but for coupling interactive data analysis tools the loose integration model has proven the most productive, and able to support ways of working that have not been possible using other available messaging systems.

5.3. MType definition process

A suitable collection of MTypes must not only exist but be known by potential message senders – which means the authors of the relevant software – in order for useful messaging to take place.

In the case of application-specific MTypes, the documentation of available MTypes and their definitions is clearly best handled as part of the documentation of the application itself. These typically provide functionality that only makes sense for a given tool, and make use of a suitably specific namespace, for instance `script.aladin.send`, which allows external applications to control Aladin by sending commands in its scripting language.

Developers are also free to define their own MTypes for use privately or in some closed group with locally agreed conventions for documentation, perhaps to support some tight-coupling-like usage.

However, for well-known MTypes intended for unrestricted use, for instance of the loose-coupling variety described above, some public process is required to establish and publicise their

definitions, so that client developers can both become aware of the conventions currently in use by other tools, and contribute their requirements for new or modified functionality.

One possibility is to decide on a fixed list to form part of the SAMP standard. A small number of “administrative” MTypes, concerned with the messaging infrastructure, for instance `samp.hub.event.register` which informs existing clients when a new client has registered, have been written in to the standard in this way. All of these are in the reserved `samp.` namespace. However, for astronomy-specific MTypes this option was rejected, partly in order to avoid the introduction of astronomy-specific details into a standard which is otherwise not tied to a particular domain, and partly because the rather heavyweight IVOA process for standard review (Hanisch et al., 2010), in which draft to acceptance rarely takes less than 12 months, would impede introduction and updating of MTypes as required by implementation experience and new application demands. Another option is periodic publication of MType definitions in an IVOA Note. Such Notes may be issued at will without formal review, but no straightforward updating mechanism is in use, and this option was still felt to be undesirably cumbersome.

Instead, a wiki page⁶ was set up on the IVOA web site listing currently agreed MTypes. An informal understanding was adopted in which application developers are encouraged to discuss requirements for new MTypes or modifications to existing ones either privately or on the associated mailing list,⁷ and if consensus is reached, to edit the wiki page accordingly. This was intended as a provisional measure to be reviewed and modified as required, but, six years later, the need for a more formal process has not been apparent, and there are no current plans to modify this arrangement.

At time of writing, a dozen MTypes are listed on the wiki, concerned with exchanging tables, row selections, FITS images, spectra, sky coordinates, VO Resource identifiers (Demleitner et al., 2014), MOC sky coverage maps (Boch et al., 2014), bibcodes and one or two other items. The list has been fairly stable, though new entries and new optional parameters are sometimes added as required.

5.4. Existing software

Since its first version in 2008, a wide range of SAMP-enabled infrastructure and application software has become available.

Of infrastructure software that is actively maintained at the time of writing, interchangeable Hub implementations exist in Java and Python, and client toolkits in Java, Python, C and Javascript. Validation, debugging and development support tools are also available. Historical, partial or experimental SAMP functionality has also appeared in other languages including Perl, C Sharp and IDL.

Applications using SAMP number in the dozens, and include GUI analysis tools for images, catalogues, spectra, SEDs, time series and interferometry data, observation tools, outreach applications, command-line and graphical data access and manipulation suites, interactive processing environments, web archives either exposing simple results pages or offering sophisticated browser functionality, throwaway user scripts, and more.

SAMP infrastructure libraries are surveyed in Fitzpatrick et al. (2013), though a notable more recent development is the inclusion of the Python hub and client implementation in the Astropy package (Astropy Collaboration et al., 2013) since its version 0.4.

⁶ At time of writing, this wiki page can be found near <http://www.ivoa.net/samp/>. If the process for MType publication changes in the future, that URL should still indicate where to look for a list.

⁷ apps-samp@ivoa.net.

⁵ <http://cdf.gsfc.nasa.gov/>.

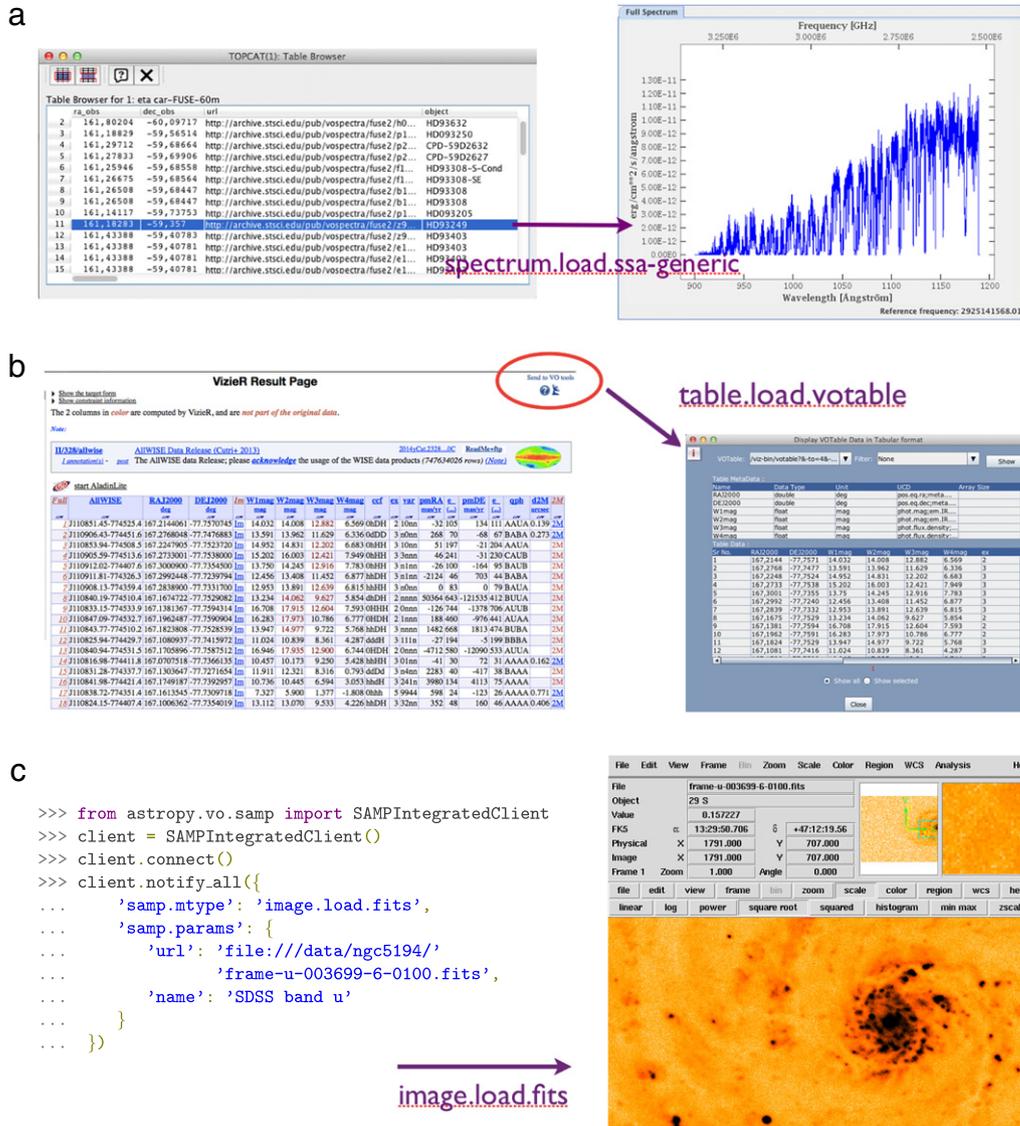


Fig. 2. Transmission of data items from one SAMP client to another. (a) A table has been acquired using Simple Spectral Access Protocol by the TOPCAT table analysis tool, in which each row references an external spectrum by URL. When the user selects a row of this table, the spectrum is sent to the CASSIS spectrum tool. (b) The source catalogue resulting from a VizieR query, displayed in a web page, is sent to the VOPlot application. (c) A local FITS image is loaded into the ds9 image viewer from the AstroPy command line.

A partial list of some other tools with SAMP functionality may also be found near <http://www.ivoa.net/samp>.

While initially developed and used mostly within stellar and galactic astronomy, use is now becoming common in related fields such as planetary science (Erard et al., 2014) and space physics (Génot et al., 2014). It is probably now the case that most astronomical applications that can benefit from interaction with other such tools include at least a basic SAMP capability. It is harder to ascertain to what extent this functionality is used in practice, but the enthusiasm of application developers to incorporate SAMP is presumably indicative of its utility.

5.5. Examples of use

Current SAMP usage is most prevalent along the lines of the scenario outlined in the Introduction, allowing desktop and web-based clients to cooperate as a loosely integrated suite by employing the small number of data-exchange MTYPES in common use. Fig. 2 shows some examples of the basic case where one client transmits a data item (spectrum, table or image) to another. Fig. 3

illustrates a more sophisticated interaction in which two applications exchange data and control in both directions to provide linked views exploiting the capabilities of each.

SAMP has also been employed in other ways however, for instance to provide a private layer for RPC functionality required internally by Iris (Laurino et al., 2014) and to experiment with visualisation using on-demand data generation in Astro-WISE (Budelmeijer and Valentijn, 2013).

6. Conclusions

SAMP provides a flexible and easy to use messaging framework, deployed in much current astronomical software, which supports various models of inter-tool communication.

The most productive of these models to date has been loose coupling between a user-selected set of independently developed interactive data acquisition and analysis applications, to deliver functionality approaching that of an integrated suite. This model is built on SAMP's combination of publish-subscribe messaging,

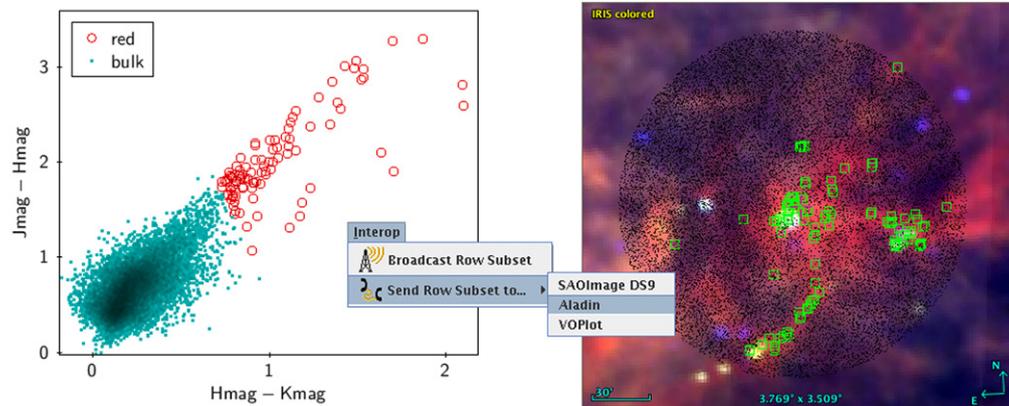


Fig. 3. Linked view of data in TOPCAT and Aladin showing red objects in the vicinity of $V^* V410$ Tau. A full list of sources in the region has been loaded into Aladin (right), then transferred to TOPCAT (left), using SAMP MType `table.load.votable`. The user plots a colour–colour diagram in TOPCAT and selects the reddest objects graphically, causing them to be displayed as red circles, then passes the selection back to Aladin (MType `table.select.rowList`) where they are shown as green squares; the inset menu shows TOPCAT’s user interface for this step. Clicking on one of the points in either application can then highlight the corresponding point in the other (MType `table.highlight.row`). This example illustrates how SAMP enables seamless exploration of data using a combination of parameter space and physical space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

vague message semantics, and ease of adoption by both developers and users leading to widespread uptake.

SAMP’s flexibility means that it is capable of supporting other communication models, some in more marginal use now and some which may be explored further in the future. Introducing new Profiles, different MType libraries or alternative hub provision arrangements could render the same infrastructure suitable for contexts with different requirements for reliability, security or scalability. Another possible scenario is inter-host messaging to support collaborative work; this option has been under consideration throughout SAMP’s history and is possible using existing profiles, though in current configurations it is somewhat cumbersome to set up and has so far not received much user attention. Despite its development history, there is nothing in the protocol specific to either the Virtual Observatory or astronomy, so use in other problem domains is quite feasible, though the authors are not aware of effort currently being deployed in this direction.

SAMP’s design has been informed by the requirements and experimentation of the SAMP developer community, largely within the context of the Virtual Observatory, including positive and negative lessons learned from its predecessor PLASTIC. Some aspects of this design that have proved particularly successful include the decoupling of architectural concerns into API, transport mechanism and semantics, the lightweight, bottom-up process for agreement of semantics, and the built-in extensibility provided by pervasive use of extensible vocabularies.

Together, these fall under the heading of standardising only those things which need to be defined at a given stage, and leaving the option of filling in the details until a time and in a context when the requirements will be clearer. The need for the Web Profile was not foreseen when the first version of the standard was published, but the transport/API decoupling meant it could be retrofitted with no disruption to existing client code. The fact that MType semantics are excluded from the standard itself means that these can be defined and iteratively adjusted with experience of a working transport infrastructure, rather than specifying them up front by committee decision as part of the protocol design, only to find them ill-adapted to tool deployment in practice.

Other factors important to its success have been the small number of MTypes actually in common use, enabled by the convention of vague message semantics and standardisation on data formats, and the unobtrusive embedding of SAMP into existing applications meaning that the functionality is available without requiring any special setup or understanding from the user. The IVOA and other cross-institutional forums associated with the Virtual Observatory

movement have also been of considerable importance in enabling and encouraging the necessary communication between application developers, though much software from outside the VO community is now also involved.

SAMP is not a magic bullet. In typical current use the level of integration it offers between independently developed tools falls short of what would be available from a monolithic application, its pragmatic approach to communications can lead to patchy reliability, and its security model would not be appropriate for use with commercially sensitive data. However its ease of use and widespread uptake have delivered in practice an improved environment for desktop data analysis, allowing working astronomers to get more done.

Acknowledgements

The development and adoption of SAMP has been a collaborative effort between participating developers within the Virtual Observatory community, particularly the Applications Working Group of the IVOA. Initial work on PLASTIC and SAMP was supported by the EU FP6 project VOTech and FP7 project AIDA. The SC4DEVO workshop series was funded by the UK e-Science Core Programme. MBT has additionally been supported in development of the SAMP protocol and software by several grants from the UK’s PPARC and STFC research councils, and in writing this paper by STFC grant ST/J001414/1. Development of the SAMP Web Profile was supported by Microsoft Research. All of this support is gratefully acknowledged.

References

- Astropy Collaboration, Robitaille, T.P., Tollerud, E.J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A.M., Kerzendorf, W.E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M.M., Nair, P.H., Unther, H.M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J.E.H., Singer, L., Fox, R., Weaver, B.A., Zabalza, V., Edwards, Z.L., Azalee Bostroem, K., Burke, D.J., Casey, A.R., Crawford, S.M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P.L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., Streicher, O., 2013. Astropy: A community Python package for astronomy. *Astron. Astrophys.* 558, A33. doi:10.1051/0004-6361/201322068, arXiv:1307.6212.
- Balm, P., 2012. Herschel interactive processing environment (HIPE): Open to the World and the Future. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), *Astronomical Data Analysis Software and Systems XXI*, p. 733.
- Boch, T., Comparato, M., Taylor, J., Taylor, M., Winstanley, N., 2006. PLASTIC - A Protocol for Desktop Application Interoperability. Technical Report. IVOA Note.
- Boch, T., Donaldson, T., Durand, D., Fernique, P., O’Mullane, W., Reinecke, M., Taylor, M.B., 2014. MOC—HEALPix Multi-Order Coverage Map. Technical Report. IVOA Recommendation.

- Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M., Bartlett, J.G., 2000. The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. *Astron. Astrophys. Suppl. Ser.* 143. doi:10.1051/aas:2000331.
- Buddelmeijer, H., Valentijn, E.A., 2013. Query driven visualization of astronomical catalogs. *Exp. Astron.* 35, 283–300. doi:10.1007/s10686-011-9263-0, arXiv:1110.2294.
- Comparato, M., Becciani, U., Costa, A., Larsson, B., Garilli, B., Gheller, C., Taylor, J., 2007. Visualization, exploration, and data analysis of complex astrophysical data. *Publ. Astron. Soc. Pac.* 119, 898–913. doi:10.1086/521375, arXiv:0707.2474.
- Demleitner, M., Greene, G., Le Sidaner, P., Plante, R.L., 2014. The virtual observatory registry. *Astron. Comput.* doi:10.1016/j.ascom.2014.07.001.
- Erard, S., Cecconi, B., Le Sidaner, P., Berthier, J., Henry, F., Chauvin, C., André, N., Génot, V., Jacquy, C., Gangloff, M., Bourrel, N., Schmitt, B., Capria, M.T., Chanteur, G., 2014. Planetary science virtual observatory architecture. *Astron. Comput.* doi:10.1016/j.ascom.2014.07.005.
- Fitzpatrick, M., Laurino, O., Paioro, L., Taylor, M.B., 2013. Application interoperability with samp. In: Friedel, D., Freeman, M., Plante, R. (Eds.), *ADASS XXII. ASP, San Francisco*, p. 395.
- Génot, V., André, N., Cecconi, B., Bouchemit, M., Budnik, E., Bourrel, N., Gangloff, M., Dufourg, N., Hess, S., Modolo, R., Renard, B., Lormant, N., Beigbeder, L., Popescu, D., Toniutti, J.P., 2014. Joining the yellow hub: uses of the simple application messaging protocol in space physics analysis tools. *Astron. Comput.* doi:10.1016/j.ascom.2014.07.007.
- Hanisch, R.J., Arviset, C., Genova, F., Rino, B., 2010. IVOA Document Standards, Version 1.2. Technical Report. IVOA Recommendation.
- Jenness, T., et al., 2014. Learning from 25 years of the extensible N-Dimensional Data Format. *Astron. Comput.* (in press) doi:10.1016/j.ascom.2014.11.001.
- Joye, W.A., Mandel, E., 2003. New Features of SAOImage DS9. In: Payne, H.E., Jedrzejewski, R.L., Hook, R.N. (Eds.), *Astronomical Data Analysis Software and Systems XII*, p. 489.
- Lafrasse, S., Bourges, L., Mella, G., 2012. SAMP App Launcher: An On-Demand VO Application Starter by JMMC. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), *Astronomical Data Analysis Software and Systems XXI*, p. 379.
- Laurino, O., Budynkiewicz, J., D'Abrusco, R., Bonaventura, N., Busko, I., Cresitello-Dittmar, M., Doe, S.M., Ebert, R., Evans, J.D., Norris, P., Pevunova, O., Refsdal, B., Thomas, B., Thompson, R., 2014. Iris: An extensible application for building and analyzing spectral energy distributions. *Astron. Comput.* doi:10.1016/j.ascom.2014.07.004.
- McIlroy, M.D., Pinson, E.N., Tague, B.A., 1978. *Unix time-sharing system: Foreword.* *Bell Syst. Tech. J.* 57, 1899.
- Sauro, H.M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J., Kitano, H., 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* 7, 355–372.
- Taylor, J.D., Boch, T., Comparato, M., Taylor, M., Winstanley, N., Mann, R.G., 2007. Binding Applications Together with PLASTIC. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), *Astronomical Data Analysis Software and Systems XVI*, p. 511.
- Taylor, M.B., 2005. TOPCAT & STIL: Starlink Table/VOTable Processing Software. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), *Astronomical Data Analysis Software and Systems XIV*, p. 29.
- Taylor, M.B., Boch, T., Fay, J., Fitzpatrick, M., Paioro, L., 2012a. Samp: Application messaging for desktop and web applications. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), *ADASS XXI. ASP, San Francisco*, p. 279.
- Taylor, M.B., Boch, T., Fitzpatrick, M., Allan, A., Paioro, L., Taylor, J., Fay, J., 2012b. Simple Application Messaging Protocol, Version 1.3. Technical Report. IVOA Recommendation. arXiv:1110.0528.
- Škoda, P., Draper, P.W., Castro Neves, M., Andrešič, D., Jenness, T., 2014. Spectroscopic analysis in the virtual observatory environment with SPLAT-VO. *Astron. Comput.* doi:10.1016/j.ascom.2014.06.001.
- Winstanley, N., Taylor, J.D., Taylor, M.B., Noddle, K., Gonzalez-Solares, E., Lindroos, J., 2007. Astro Runtime: An API to the Virtual Observatory. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), *Astronomical Data Analysis Software and Systems XVI*, p. 571.