International Journal of Astrobiology

cambridge.org/ija

The interstellar communication relay

Brian McConnell

SETI Open Data Archive, San Francisco, CA, USA

Abstract

The paper describes the architecture for a data repository and distribution system to be used in the case of a SETI detection event. This system is conceptually modelled after the Deep Space Network, although the hardware and infrastructure involved are different and substantially less expensive to operate. The system is designed to accommodate a large number of users from a variety of fields who wish to contribute to the analysis and comprehension effort that would follow the detection of an information-bearing signal.

Research Article

Cite this article: McConnell B (2020). The interstellar communication relay. *International Journal of Astrobiology* **19**, 419–422. https://doi.org/10.1017/S1473550420000178

Received: 12 June 2020
Revised: 15 July 2020
Accepted: 17 July 2020
First published online: 26 August 2020

Key words:
Information theory; interstellar communication; meti; seti

Author for correspondence:
Brian McConnell,
E-mail: bsmcconnell@gmail.com

Problem statement

SETI organizations understandably devote the bulk of their effort and funding to detection efforts. Relatively little attention has been given to the activities that would follow the detection of an extraterrestrial signal beyond its initial confirmation and announcement. The protocol for this process, the International Academy of Astronautics (IAA SETI) Declaration of Principles (International Academy of Astronautics, 1989), was first adopted in 1989 and last updated in 2010 (International Academy of Astronautics, 2010). It details the steps leading up to the announcement of a SETI detection event. The Rio Scale, first proposed in 2000 by Almár and Tarter (2000) and updated in 2018 (Forgan *et al.*, 2019), provides additional information about the detection confidence and significance of a candidate extraterrestrial (ET) signal. While these are helpful in defining the steps to be taken up to the public announcement of detection, both are silent about how to distribute information and data to a large audience of scientists and laypeople following the confirmed detection of an information-bearing signal.

SETI organizations may face a number of challenges following a confirmed detection, among them:

(1) Inadequate information infrastructure to meet the demand from people and organizations seeking information about the detection. This demand is likely to be especially intense in the case of an information-bearing signal.
(2) Coordinating the flow of information from multiple observing sites, especially in the case of an information-bearing signal where multiple sites may be transcribing data from the signal(s).
(3) Deliberate attempts to interfere with or disrupt the flow of information from observing sites and organizations, especially in cases where state actors wish to monopolize access to information (Wisian and Traphagan, 2020).
(4) The spread of misinformation by private actors seeking to exploit the contact event for their own agenda or financial gain.

SETI organizations can mitigate these challenges and risks by building a reliable and scalable platform for archiving and distributing news and data for a broad user community and by having this ready in a standby capacity so that it can be pressed into service on short notice.

Design goals & requirements

Given that a SETI detection event is a low probability, high impact event, the system should meet the following criteria.

- It should be able to automatically scale to serve a large number of users (>1 million concurrent users).
- It should use tools and data formats that are familiar to a variety of users with differing skill sets.
- It should be inexpensive to maintain in its resting state, with zero or very low fixed operating costs. Significant usage costs should only be incurred when the system is activated.
- It should be open to multiple observing sites so that additional upstream data providers can be added to the system without much effort.
- Data archived on the system should be stored in a redundant manner, to protect against accidental data loss, malicious attacks, etc.

© The Author(s), 2020. Published by Cambridge University Press

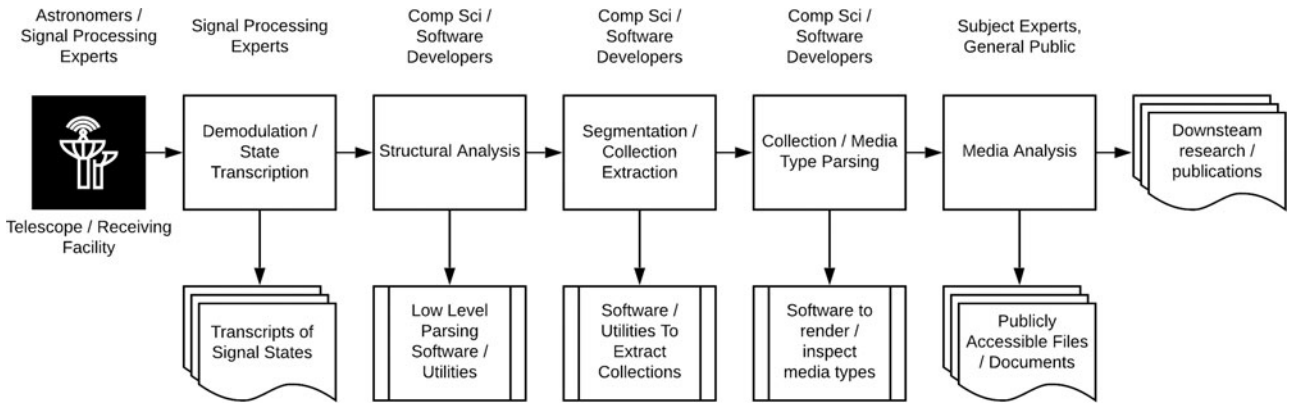


Fig. 1. Processing Pipeline diagram.

- The system should provide scientists and the public with an authoritative source of information and data, both to facilitate analysis and also to mitigate the negative impact of bad actors claiming to have special access to information.

Until recently it would have been expensive to build and maintain such a system due to the high fixed costs of operating a data centre, provisioning broadband internet service and paying systems administration staff to maintain specialized hardware, especially in the context of the limited operating budgets of SETI programs. These have since been virtualized in the form of cloud computing services and applications (Amazon Web Services, Google App Engine, Github, etc). It is now possible to build facilities such as this at orders of magnitude less cost, with very low fixed costs.

Use cases & user communities

The Interstellar Communications Relay will serve a number of user communities, among them:

- Communications experts who may or may not be affiliated with SETI. These may include digital signal processing experts, computer scientists and other professionals who want to contribute to signal confirmation or low-level analysis efforts.
- Amateur and professional scientists and researchers who wish to contribute to the analysis and comprehension of information extracted from an ET transmission.
- Journalists, broadcast media and the public seeking access to derived data products from the system.

These user communities can be broken down into several tiers which map to different infrastructure requirements.

The processing pipeline

The interstellar communication relay (ICR) will create the infrastructure for a decentralized data processing pipeline which serves a number of use cases and user communities as shown below (Fig. 1).

Tier 0 – signal processing (Onsite/direct connectivity)

Tier 0 users will require to access raw data feeds, either in real-time or for retrospective data mining and analysis. A variety of constraints, such as limited connectivity to and from receiving sites, make a cloud-based system impractical for this use case.

For these users, a 'BYOCC' (bring your own compute cluster) approach where the hosting sites provide standardized facilities and interconnects will be more feasible. This will allow for third party gear to be connected to a high-speed LAN or read-only disk array at the observing site(s) with minimal effort. Not much is required to support this except to agree on conventions for physical connectivity, disk access, etc.

User attributes

- Highly technical, self-sufficient once trained
- Requires high-speed access to a large corpus of data
- Has the financial means to build their own equipment and travel to observing sites
- Likely to be a university or corporate team, possibly with state backing

Tier 1 – signal processing (Remote/cloud-based)

For researchers that want to work with archival data from a remote location, the system will provide them with the ability to request data via a public web API. The requester would receive a blob of data or an hypertext transfer protocol (HTTP) error in response. This would enable smaller research teams to build systems that can retrieve data as it is posted to the system and to build automated processing pipelines to suit their needs. The Breakthrough Listen team recently released such a data set which provides a good model for other SETI organizations and observing sites to follow (Lebofsky *et al.*, 2019).

User attributes

- Highly technical, self-sufficient once trained.
- May or may not be well funded, may or may not be affiliated with a university or corporate team, though the former is more likely.
- Does not need access to the primary data 'firehose', more likely to be doing specialized analysis using reduced data products.

Tier 2 – demodulation/state transcription

In the case of an information-bearing signal, the demodulated data would be made available to anyone who wants to contribute to analysis and comprehension efforts. This data set will likely be smaller than raw and reduced signal data sets, probably on the order of a few bits to a few kilobits per second, although higher data rates are possible depending on the transmitter's equipment and energy budget (Shostak, 2010). While the data set will likely be smaller, the number

https://doi.org/10.1017/S1473550420000178 Published online by Cambridge University Press

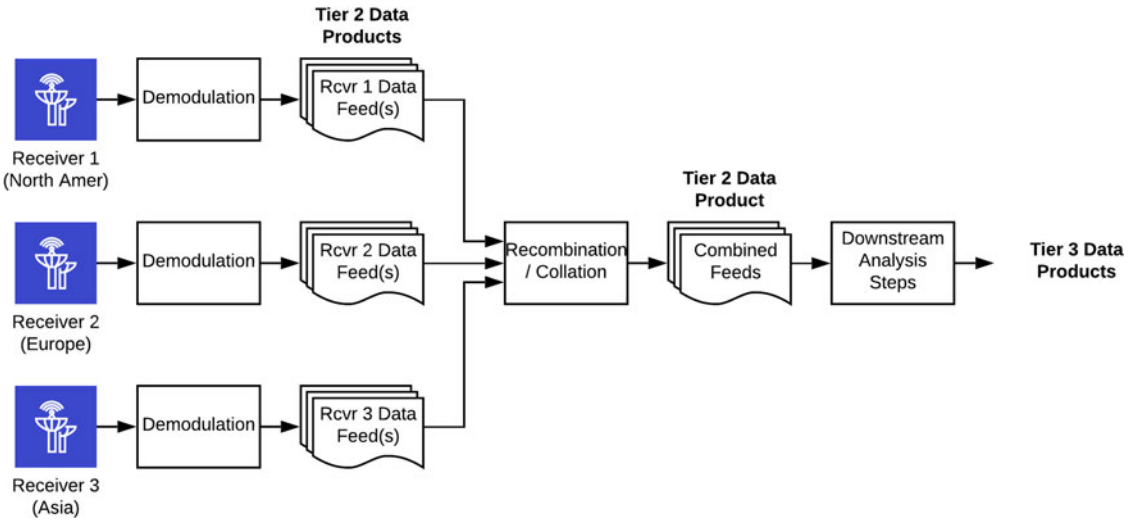


Fig. 2. Processing steps for a geographically dispersed multi-receiver system, where each receiver transcribes demodulated data from one or more subchannels, which can then be recombined and collated in a subsequent processing step.

of consumers may be quite large, as anyone with computing skills may wish to test hypotheses about what the data represents.

The data streams would be stored in a repository that is familiar to software developers and computing professionals. Something like Github is ideal for this sort of use case, as it provides a history of commits and changes, so observers can have confidence that the repo has not been altered retroactively. The repo would be configured so that only authorized sites would be able to commit updates. Github is an off the shelf solution, with proven scalability, so it should be on the shortlist of services to consider for this function (Fig. 2).

The primary data product from this service would be an interchange data format that is easy to parse and manipulate in a variety of programming languages. JSON is a good example as it is easily parsed via built-in libraries and allows for flexibility in terms of interleaving data, metadata and loosely defined schemas. Note that no attempt would be made to decipher this data except to transcribe symbol states into an easily parsed format and collate data from multiple receivers at different geographical locations. Further decipherment and analysis would be conducted in downstream pipeline steps.

User attributes

- Wide range of technical skills. Most would be proficient computer users, though not necessarily computer scientists.
- Wide range of disciplines, with people from many fields checking out data to have a try at interpreting it.
- No requirement for ultra-fast internet connectivity or storage since the data payloads involved are likely to be small initially.

Tier 3 – derived data products

In the case of an information-bearing signal, it is possible that some sections of it will be decoded relatively easily. One can imagine a signal that alternates between simply encoded images and compressed, structurally rich information that is more difficult to comprehend.

Images are an interesting case as uncompressed bitmaps are easily represented via an array of numbers. Planetary images are especially easy to recognize as they typically feature a spheroid object against a null or black background.

As analysts learn how to decode and render different segments of a data feed, the system would emit derived data sets. In the cases of images, an algorithm would extract and transform them into a widely used image format such as the PNG format, as well as scientific formats such as FITS. These too would be uploaded to a git repository, from which other parties could extract and repackage them to suit their needs.

It is important to note that people will generate these derived data products whether or not SETI organizations do so. Because it is easy to manipulate images and other media types, it is probably not a good idea to cede this step in the process to others.

User attributes

- Non-IT subject experts, such as archaeologists, biologists, etc
- Casual users and the general public
- Press and media

Operating costs and modes of operation

Standby mode/Normal operation

The system would normally function as a shared website and news hub for participating SETI organizations. This would be promoted as a news aggregation service that any person or organization following SETI and technosignature research could follow.

The cost of operating this service will be minimal. Cloud-based computing services such as Amazon Webservices (AWS) and Google Compute are billed on the basis of a transaction and/or data volume and also offer generous free tiers of service that are sufficient for light to moderate traffic. These costs can be further reduced through the use of content delivery networks which offload most traffic to caching services distributed throughout the Internet. Ongoing operating costs are estimated at a few hundred dollars per month, not including the cost of the editorial staff to maintain content on the site, costs that SETI organizations can easily bear.

Post detection mode

In the event of a confirmed detection, the system would serve as a central point of contact for news about the detection, as well as for data obtained from the signal. The system would likely

https://doi.org/10.1017/S1473550420000178 Published online by Cambridge University Press

experience high demand during this period, especially in the weeks following the detection.

We can estimate the operating costs as a function of the signal's data transmission rate. Cloud computing services charge on the order of $0.10 USD per gigabyte of data transfer as of this writing. A high estimate of the system's monthly operating cost based on published rate cards can be calculated as follows:

$$C_{(\$/\mathrm{mo})} \approx \$3.24 \times 10^{-5} \times n_{\mathrm{users}} \times R_{(\mathrm{bits/sec})}$$

Consider an example where roughly one million people directly participate in the analysis and comprehension effort and the signal's data rate is on the order of 1 kilobit per second. This translates into an annual operating cost of about $ 400 000. The cost of delivering information to indirect consumers and the general public can be offloaded to other science and media organizations, so the ICR would only need to bear the cost of delivering scientific data products to amateur and professional researchers. This figure is itself a high estimate, as it is likely that some or all of this cost can be offset through philanthropic grants and corporate sponsorships due to the high visibility and societal value of this system.

Legal framework and organization

The sponsoring organization would be governed by a board composed of representatives from recognized SETI and astronomical organizations. An advisory board composed of subject matter experts would manage technical/architectural decisions, such as access rights, data products and formats, etc.

Software and information products generated by the system would be published under an open-source license that allows for their free redistribution and reuse. Participating SETI organizations would additionally commit to publishing their data products under an equivalent open source license. The license selected should not preclude users from developing commercial products or services from these data products such as games and educational software, but should preclude any one organization from asserting exclusive rights to the underlying data used to create these properties. A number of open source and 'copyleft' licensing schemes, such as the Creative Commons license [8] should satisfy this requirement.

Next steps and actions

IAA statement of principle

Just as it defined protocols for the announcement of a SETI detection, the International Academy of Astronautics could define a set of post-detection actions for SETI organizations to endorse and follow, as an addendum to the existing detection protocol. Among the recommended actions to be considered:

(1) Data extracted from an information-bearing message should be published as it is received, and should be published in data formats that are accessible to a broad user community.
(2) Data should be published under a public domain or copyleft licensing scheme that prevents any one entity from establishing exclusive access rights to the underlying data extracted from the transmission but does allow for non-exclusive commercial applications derived from this data.
(3) Raw and reduced signal data should be preserved to the best extent possible, to facilitate future analysis, e.g. to search for lower power sidebands or alternate modulation methods that may not be noticed during early analysis.

Advisory board composition

The advisory board should be composed of subject matter experts from a variety of backgrounds. This board will be responsible for deciding on hosting platforms, data products and application programming interface (API) design, with the goal of supporting a varied user base with differing skill sets.

Agreement on technical standards and data products

Once formed, the advisory board will define standards for hardware interfaces (for Tier 0 users), API specifications (for Tier 1 and 2 users) and data products to be used in Tier 2 analysis.

Observing sites that wish to plan for Tier 0 access will primarily need to provide rack space for computing equipment, along with high-speed Ethernet LAN connectivity and standard power interconnects.

The Tier 1 interface to the system should implement a common application programming interface based on the REST design pattern. The public data set published by the Breakthrough Listen team provides a good template to follow.

Tier 2 data products are yet to be defined. The data format used to record transcribed signal states should be widely accessible and extensible. This will enable users to process data in the programming language of their choice and will enable data providers to extend the metadata interleaved with transcribed data as needed. A JSON or XML-based format will be useful here.

The Tier 3 data products cannot be defined until more is known about the data stream's contents, but it should be possible to anticipate some likely data types such as images and prepare tooling forthem.

Financial support. The system will have low operating costs during its resting state; however, post-detection operating costs may be significant. Sponsoring organizations should agree to a funding structure that accounts for these costs, or obtain commitments from sponsors or funding agencies to cover these costs should a detection event occur.

References

Almár I and Tarter J (2000) The discovery of ETI as a high-consequence, low-probability event. Paper #IAA-00–IAA9.2.01, 51*st* *International Astronautical Congress*, Rio de Janeiro.

Forgan D, Wright J, Tarter J, Korpela E, Siemion A, Almár I and Piotelat E (2019) Rio 2.0: revising the Rio scale for SETI detections. *International Journal of Astrobiology* **18**, 336–344.

International Academy of Astronautics (1989) Declaration of Principles Concerning Activities Following the Detection of Extraterrestrial Intelligence.

International Academy of Astronautics (2010) Declaration of Principles Concerning the Conduct of the Search for Extraterrestrial Intelligence. *SETI Permanent Study Group of the International Academy of Astronautics*.

Lebofsky M, Croft S, Siemion A, Price D, Enriquez J, Isaacson H, MacMahon D, Anderson D, Brzycki B, Cobb J, Czech D, DeBoer D, DeMarines J, Drew J, Foster G, Gajjar V, Gizani N, Hellbourg G, Korpela E, Lacki B, Sheikh S, Werthimer D, Worden P, Yu A and Zhang Y (2019) The breakthrough listen search for intelligent life: public data, formats, reduction, and archiving. *The Astronomical Society of the Pacific* **131**, 1006. 10.1088/1538-3873/ab3e82.

Shostak S (2010) Limits on interstellar messages. In Vakoch DA (ed). *Communication with Extraterrestrial Intelligence*. Albany, NY, USA: SUNY Press, pp. 357–378.

Wisian KW and Traphagan, JW (2020) The search for extraterrestrial intelligence: a realpolitik consideration. *Space Policy*. doi: 10.1016/j.spacepol.2020.101377.

https://doi.org/10.1017/S1473550420000178 Published online by Cambridge University Press