DAI WG Report to the CMC

Response to CMC resolution (TBD Number)

# The CMC Resolution:

CMC acknowledges that in today’s  Data Economy, space systems  generating and transporting large amounts of relevant data;  be it science (GAIA is generating 50 GB/day) earth observation (Sentinel-2A generating 1 TB/ h), navigation or telecommunication satellites,  constitute key assets for the Data industry.

In this context, standards supporting proper storage and preservation of data, open access to the space data, merging of data from different sources, analysis and processing big data and all related processes and mechanisms are of paramount importance.

Scientific communities are already for quite some time working nationally and internationally to define adequate standards for their data, be it astronomy, planetary exploration or earth observation.

The question is what should the CCSDS organisation do and how should it position itself in the wake of such space data revolution and  how could the work done over the years to standardise data storage and archiving be leveraged at the service of the new paradigms.

In this context, and following the request of the DAI WG to the CMC for support, CMC would like to encourage the DAI WG to address the following questions within a short paper to be delivered to CMC *by 1 February 2017*:

1.        What is the current DAI roadmap and how does it make a valuable contribution to the data issue, taking into account what has already been produced including the overall worldwide efforts in this area?

2.        Taking into account that there are dedicated international organisations working on archiving and merging scientific data, astronomy data, as well as Earth data, what should be the role of CCSDS in general and that of the DAI WG in particular?

3.        Is the long term data preservation of the housekeeping data properly addressed in these initiatives? Is there a need for developing this part of the data preservation?

4.        In this context, what would be the next project for the DAI WG in the future, should resources be available, and how would this project fit with the current CCSDS reference architecture and work plan?

# Introduction

The Data Archive Ingest (DAI) WG thanks the CMC for the opportunity to respond to their insightful context and questions that were posed in their resolution.

To add context for the broad scope of the DAI WG, we have been focused on long-term preservation of data, such as space mission science results. “Long-term” implies access even after the mission systems have been terminated and the data in the archive must be “self describing” to different communities of users and new systems seeking access.

DAI has not specifically addressed “near-term” telemetry archive access (for example) by MOC systems while the mission is still underway, although our processes and approaches are important practices for both near-term and long-term digital archives. Specifically, it is important to note that long-term preservation requires functions to be performed during the mission, capturing data descriptions, etc. So the DAI WG products, while focused on long-term preservation, apply to the entire life cycle of a space mission, even the early conceptual phases (Phase A). Unfortunately, most agencies’ missions do not adopt our preservation techniques during the early timeframes. We want to promote awareness of this among the CCSDS Space Agencies.

To respond to the CMC, first we will provide some explanations that respond to the context in the resolution’s introductory material. Then we will explicitly answer the questions one-by-one.

We interpret the introductory context material in the CMC resolution to ask the question about what CCSDS and DAI should do to position the CCSDS agencies to function in the “big data” world.

It is well known that the CCSDS DAI WG formulated the OAIS standard for the purposes of digital preservation of the space agencies’ valuable mission data, but the solutions and practices were so valuable that OAIS has been adopted by archival facilities around the world.

The DAI WG has a roadmap (to be described in response to question 1) and plans to refocus on interoperable protocols for digital archive access. This includes a general framework on the user side, an abstraction layer (similar to that developed by the SM&C WG), and archive interfaces that are specific to each discipline or “designated community”. A very successful effort would result in the user interface and the abstraction layer being another general solution that works well for the global problem of digital preservation, resulting in adoption by the same global community that adopted OAIS. But the discipline-specific archive interfaces developed by DAI would be dedicated to the space agencies’ needs for space science, spacecraft data (housekeeping) and perhaps other data types unique to space programs (such as Enterprise Data for developing spacecraft).

This architecture would have the unique feature of enabling cross-discipline research in trusted digital archives. The Space Science researcher would use the archive user interface and abstraction layer and the space science specific “plug-in” to access space science results and spacecraft telemetry. But (s)he would also be able to install a plug-in for geological data archives or weather data archives by installing those “plug-ins”, developed by those communities. Meanwhile the geology or weather researcher would similarly be able to plug in the space science archive plug-in to correlate their discipline data with spacecraft mission results.

The importance of this cannot be overstated. In the future, the greatest technological advances in the world are expected to come from cross-discipline research.

From our extensive contacts with the digital preservation community, we can verify that no other organizations are working towards standards for interoperable protocols to provide this capability to the space agencies or any other archival organization.

So as you see, the answer to your general question about the relevance and role of CCSDS in the “big data” world is easily apparent and profound: First enable broad, open access to space mission “big data” archives to both space program insiders and outsiders, while preserving the data for generations to come. And then to share that capability with the world at large to enable digitally-driven cross-discipline technological advancement of mankind, and contribute to solving the “digital dark ages” problem.

Some other benefits of the DAI WG approach are:

* Development with the ISO community spreads the cost of development to other organizations outside the space agencies.
* It expands the global relevance of our spaceflight missions to the broader data industry, and contributes to the best kind of technological advances (cross-discipline).
* It will raise the stature of CCSDS as a great leader in interoperability standards.

Granted that the DAI WG members may not be objective with this next statement, but our assessment is that the work described here is the most important work in CCSDS because over the long term (for future generations) it is the most beneficial to the space agencies and the world at large. And we are perplexed about why space agency support has dropped off so greatly in this area, compared to the broad support that was provided for the Panel Two work early in the CCSDS era.

# Question 1: The DAI Roadmap

What is the current DAI roadmap and how does it make a valuable contribution to the data issue, taking into account what has already been produced including the overall worldwide efforts in this area?

First, some history. DAI has previously placed the greatest emphasis on the core underlying processes ([OAIS](https://public.ccsds.org/Pubs/650x0m2.pdf), [IPELTU](http://cwe.ccsds.org/fm/Lists/Projects/DispForm.aspx?ID=368&Source=http%3A%2F%2Fcwe%2Eccsds%2Eorg%2Ffm%2FLists%2FProjects%2FAllItems%2Easpx&ContentTypeId=0x0100B63160D64FE81342BE42A874DE7E703D)) before addressing interoperable protocols or interface specifications extensively. Even so, some interface specifications were accomplished ([PAIS](https://public.ccsds.org/Pubs/651x1b1.pdf)).

The roadmap for the DAI WG has two components: Data architecture and schedule.

## The DAI vision for architecture and protocols

Notable characteristics of the planned protocol architecture (Figure 1) are:

* User interfaces (Producer and Consumer) at the top.
* An Archive Abstraction Layer in the middle (similar to the MAL from the SM&C WG)
* Bindings on the bottom to interface to specific types of archives.
* Legacy interfaces to current archives are shown on the right side. The new architecture can be an add-on to legacy systems, and operate in parallel.
* The overall framework would apply to all OAIS libraries and archives around the world. And they will help develop it. Under the CCSDS umbrella, we can be assured it will meet the needs of the Space Agencies.
* Unique “bindings” may actually be APIs or protocols. Only the bindings specific to spaceflight mission archives need be developed with the participation of the CCSDS space agencies. Other bindings can be developed elsewhere in the industry. For example, the Web/HTML binding may be developed by the IIPC (International Internet Preservation Consortium) but will be available to space agencies (and their national archives) for their web preservation purposes.
* Four example bindings are illustrated, but we anticipate that many other organizations would develop bindings for their specific data types or communities.
* So the users (producers and consumers) have a standardized interface that is stable, and they have the ability to “plug in” bindings for many different data types to access data in other communities besides their own.
* In a few words, all-inclusive, all-serving, CCSDS-compatible and modular. And supportive of cross-discipline research.



Figure 1Abstract Protocol Architecture

If this brief introduction to the long-range vision of Digital Preservation work in CCSDS is not self-explanatory, we would be happy to again attempt a special topic presentation to the CMC at an upcoming telecon or meeting.

The DAI WG intends to “socialize” this concept with other Digital Preservation organizations around the world, and to hopefully get their support to this as a long-range industry-wide vision. A CMC resolution supporting this long range direction (while not yet approving specific projects) would be a welcome support to this effort.

## The DAI Notional Schedule

In the DAI summary schedule below we cover historic, current and future work items. For the “road map” you can see where we must finish current reviews/projects before we can start on the new architecture/protocol work. The “future” tasks can be mapped to specific features of the architecture. We realize that the mapping is not included in this paper for the sake of brevity, but we can provide that if desired by the CMC. The major points on this schedule are:

* We anticipate a pace of about a document per year, with 3 or 4 documents in work at any given time.
* Some may be able to be combined, which could shorten the delivery cycle.



# Question 2: Roles

Taking into account that there are dedicated international organisations working on archiving and merging scientific data, astronomy data, as well as Earth data, what should be the role of CCSDS in general and that of the DAI WG in particular?

It is quite true that scientific communities have for quite some time been working nationally and internationally to define adequate standards for their data, be it astronomy, planetary exploration or earth observation. However, all of the ones that we know about are defining formats and ontologies that are discipline-specific. They are working only in their “designated communities” (in OAIS terms) and they are not addressing the proper storage and preservation of data in trustworthy digital repositories. These domain specific standards complement the CCSDS standards about digital preservation.In particular OAIS defines “Representation Information” as the information needed to understand and use the data being preserved. The domain specific formats and ontologies are specific examples of Representation Information.

A survey was conducted using the MOIMS-DAI extensive contacts in these many areas.

Some specific results of our survey are:

* OGC; The list of OGC standards is available at http://www.opengeospatial.org/standards. While there are OGC standards unique to earth science topics, and that touch on digital preservation for the vertical applications of that domain. However, there is not work being done for data types, and there is not a general standard that supports digital preservation (the main focus of DAI). For example, their APIs are specific to geospatial data. We have some common membership in both the DAI WG and the OGC, so there is not a risk of losing sight of both efforts.
* ISO TC211/WG7 also produces Geographic Information standards. They have incorporated OAIS into much of their work, and our assessment is that our work is complimentary.
* IIPC; The IIPC conference and the Preservation Working Group in April 2016 was attended by DAI representatives. During the conference the community expressed interest in developing tools, but when the question was raised about developing standards, the response from the community was negative. No activity has begun.
* OMG; While the OMG has listed “Archives” on their list of future efforts, they have not begun any work, and there is no schedule for starting. The topic for them is simply spacecraft telemetry archives (not other space mission related archives) and they are not addressing long-term digital preservation methods. This info is from a phone call to the co-chair of the OMG Space Domain Task Force.
* RDA has many groups, two of the most relevant are the [Active Data Management Plans Interest Group](https://www.rd-alliance.org/groups/active-data-management-plans.html) and the [Preservation e-Infrastructure Interest Group](https://www.rd-alliance.org/groups/preservation-e-infrastructure-ig.html). Both these groups have common members between each other and with MOIMS-DAI. This ensures the consistency of the work of the groups. The roadmap of complementary work here is that RDA groups will help in the review and implementation of the CCSDS standards. Other RDA groups are developing specific Representation Information.
* ISO TC46 and the BnF (National Library of France) draft for interoperable archives (but not for digital preservation, because it does not address OAIS archives with AIPs?)
* LTDP; The goals of the Long Term Data Preservation framework (LTDP <https://earth.esa.int/web/gscb/ltdp>) are

To preserve the European, and Canadian, EO space data sets for an unlimited time-span.

* + Ensure and facilitate the accessibility and usability of the preserved data sets respecting the individual entities applicable data policies.
	+ Through the adoption of a cooperative and harmonized collective approach among the data owners (LTDP Framework) based on the application of European LTDP Common Guidelines and sustained through cooperative (multi-source) long term funding schemes.
	+ Ensure, to the maximum extent, the coherency with the preservation of other non-space based environmental data and international policies..

The LTPD provides a number of guidelines in order to harmonize the approach, including a checklist for the types of information which should be preserved. However the work is very much based on OAIS rather than seeking to replace it, for example “Mission Documentation shall include Representation Information, Packaging Information and Preservation Descriptive Information according to OAIS information model [RD-6].

.

* PIN, nestor and DPC are national membership organisations in France, Germany and the UK respectively which provide for a for discussing digital preservation. None are developing standards and in fact all have made significant numbers of suggestions for the update of OAIS.
* IPDA?
* Others… (?)

 Those are the principal entities working in the Digital Preservation Arena that might possibly have been embarking on work that could overlap the DAI. However indications are in each case that they are not seeking to develop interoperable standards that specifically promote interoperability between trusted OAIS repositories between agencies and across disciplines. Therefore, within the data industry, the DAI work supports these unique roles for CCSDS:

* Preserve the space agencies’ digital data assets for future generations. (With a general solution that also meets global needs of many nations and industries.)
* Enable access across disciplines; Space researchers accessing other disciplines and other discipline researchers accessing the science and mission benefits of our space programs. (Again, with a general solution that also meets global needs of many nations and industries.)

# Question 3: Housekeeping

Is the long term data preservation of the housekeeping data properly addressed in these initiatives? Is there a need for developing this part of the data preservation?

The principles of data preservation that have been pioneered by the DAI WG are equally applicable to storage of all data types. The OAIS processes, the Producer Archive Interface Specification, the future plan for all Digital Preservation protocols all apply equally to spacecraft science data and housekeeping data. As well they apply to space program enterprise data (CAD, requirements, engineering documentation) and even more broadly to all other data types. The well-developed techniques and anticipated protocols are expected to apply equally to science data, housekeeping data and all other data types that require preservation.

It is remotely possible that when DAI seeks to develop the bindings/plug-ins/APIs/protocols as part of the long-range architecture that perhaps a separate protocol will be optimized for housekeeping data. Generally, it is anticipated that all telemetry types would have a common protocol. The area where science data and housekeeping data may diverge is in the metadata sets. Some metadata would be common, but some would be specific to the nature of housekeeping data. But that distinctive housekeeping work, if necessary, is also expected to be under the DAI WG umbrella.

# Question 4: Next Steps

In this context, what would be the next project for the DAI WG in the future, should resources be available, and how would this project fit with the current CCSDS reference architecture and work plan?

Following our roadmap as described in the response to question 1, the next steps are:

1. Finish the [OAIS](http://cwe.ccsds.org/fm/Lists/Projects/DispForm.aspx?ID=581&Source=http%3A%2F%2Fcwe%2Eccsds%2Eorg%2Ffm%2FLists%2FProjects%2FAllItems%2Easpx&ContentTypeId=0x0100B63160D64FE81342BE42A874DE7E703D) and associated simultaneous [Audit/Certification](http://cwe.ccsds.org/fm/Lists/Projects/DispForm.aspx?ID=582&Source=http%3A%2F%2Fcwe%2Eccsds%2Eorg%2Ffm%2FLists%2FProjects%2FAllItems%2Easpx&ContentTypeId=0x0100B63160D64FE81342BE42A874DE7E703D) reviews that are underway.
2. Finish a PAIS Tutorial and the [IPELTU](http://cwe.ccsds.org/fm/Lists/Projects/DispForm.aspx?ID=368&Source=http%3A%2F%2Fcwe%2Eccsds%2Eorg%2Ffm%2FLists%2FProjects%2FAllItems%2Easpx&ContentTypeId=0x0100B63160D64FE81342BE42A874DE7E703D) documents that are underway.
3. Begin on the Digital Archive Architecture Description Document (ADD). Development of an ADD is a widely recognized CCSDS practice to support a broad long-range plan of protocol development in other working groups (DTN, etc.) and the DAI WG objective is no different. (Note the title may change in future DAI WG deliberations, but the purpose will remain essentially the same.)
4. Complete the “Producer Specs” with the Producer Archive Interface Protocol either as a standalone document or as an update to the PAIS standard.
5. Develop the Archive Abstraction Layer (the equivalent of the Message Abstraction Layer developed by the SM&C WG)
6. Begin on the “Consumer Specs” branch of the document tree that is the equivalent to the “Producer Specs” that have already been published.
7. Develop the bindings/plug-ins/APIs/protocols that are specific to the spaceflight industry (telemetry archive protocols and anything else that is needed).

Also, in the timeframe of the development of the Digital Archive ADD (or perhaps earlier) we expect to begin “socializing” the long-range DAI WG plan with other organizations in the Digital Preservation arena. This will be to garner broad support for the overall development plan and to insure that there are not conflicts or overlap with work in other organizations. It should, as well, encourage resources from other organizations to help with the archive framework efforts, and to encourage other disciplines to develop their own bindings/plug-ins/APIs/protocols that are compatible with the framework and which enable cross-discipline research and operations of trustworthy digital archives.

# Concerns from the DAI WG

The DAI WG has some high-level management issues that require CMC attention

## CMC Is Encouraged to improve Space Agency Support to DAI

As in all CCSDS WGs, there are two aspects of support that is sought from the CCSDS Space Agencies: Adoption and Development. In the case of the DAI WG, we feel that the real work of the DAI WG, ***digital preservation,*** is not well understood by the space agencies’ participants that have a stronger focus on space communications than on mission applications. Hence the space agencies have faltered in their support to the DAI WG.

ADOPTION: First, we encourage the CCSDS space agencies need to improve their agencies’ adoption of the published DAI standards. It’s not good enough to say “we’re close”. Interoperability is like digital data. It’s 100% or it doesn’t work. It is not like lossy compression where close may be good enough.

DEVELOPMENT: The space agencies are encouraged to improve their support to the DAI team in developing the processes and protocols for future interoperable capabilities. The science data and engineering data investment of our agencies’ spaceflight missions are costly to acquire. The final “put to bed” costs of long-term digital preservation are minor in comparison. Besides those obvious benefits of long-range digital preservation within an agency for all data types (from science to enterprise data), the interoperability capabilities that will result will also benefit each agencies’ researchers that need to cross disciplines and access science results across a broad spectrum of science disciplines. (Given access permissions, of course.)

Considering the large storehouse of science data from decades of past space missions, what could be more important?

## CMC Position on Broader Participation for the Broader Good:

CCSDS programs such as DAI and DTN that address the needs of mankind beyond the space agencies is a very good thing for CCSDS. The outside-applicability of the DAI work is a parallel to CCSDS DTN which is now evolving to an IETF standard for the broader Internet for the public at large.

It is true that many other organizations that are not space-related will benefit from the CCSDS DAI work; National Archives, Libraries, etc. We think this is positive for the global recogniztion of CCSDS stature, as well as for the pragmatic goal of getting work to benefit the space agencies by using “outside resources”.

We think these standards that become applicable outside of the space community are a great thing for CCSDS, but we would like the CMC to confirm their agreement. If the CMC thinks this is not positive, the DAI WG would like to understand the CMC rationale.

## CMC Expression of Support for the ISO Community

DAI WG needs an expression of commitment from the CMC because the external (ISO) community has expressed concern that the CCSDS space agencies will not fully support their needs. These concerns were expressed as part of the ongoing OAIS review, as well as during conferences and on online blogs. The DAI WG feels that this concern would be eliminated by a resolution from the CMC expressing that CCSDS fully intends to embrace ISO participation in CCSDS reviews, and we intend for the process to be inclusive and transparent.

Compared to other CCSDS WGs, DAI has a large ISO (non-space-agency) participation. If these ISO participants submit their comments only during the ISO review, it would cause extensive delays and rework. DAI really needs ISO participation to happen early, during the normal CCSDS agency review. That’s the reason for this DAI process that is “different” from other CCSDS WGs, and the reason we need this resolution.

During the Fall of 2016, the DAI WG attempted to get this message to the CMC, but it did not make it through the CESG process. We would also like to suggest that the CMC should review the “filter” that the CESG invokes for WGs trying to elevate these external management issues (not technical management), and consider some guidance to the CESG that such external relations topics (and other topics?) should not be filtered out by the CESG.

## DAI Working Group efficiency

DAI WG has been inaccurately accused of inefficiencies in CESG discussions. We would like to put this topic to rest by showing the schedule of past accomplishments (integrated with the current future schedule). We believe that this illustrates a continuous strong pace of productivity when compared to other CCSDS WGs. Also, it should be noted that this is a large output using some external resources. More than any other CCSDS WG, this work uses resources outside space agencies’ budgets, hence it is very efficient for the space agencies.

## DAI Roadmap Details

DAI WG has a new long-range notional plan, and we seek CMC notional approval (understanding that project approvals will come much later). The architecture is illustrated by the block diagram in the response to question 1 (the roadmap). DAI requests that the CMC respond with general agreement that this is a good baseline approach.

## Proposed CMC resolutions:

The CMC recognizes that for the work of the Digital Archive Ingest Working Group, it is important for the ISO participation to be encouraged during the CCSDS Agency Review phase, in order to avoid extensive delays and rework in later ISO review phases. The CMC also recognizes that it is critically important to support this work for the Digital Preservation community much broader than the nominal participation of the CCSDS space agencies. The CMC therefore resolves to express support for the inclusion of the ISO community, and resolves to commit to an inclusive and transparent review process for the future DAI WG projects in standards for Digital Preservation.

Action to the CMC: Provide names of contacts in the National Agencies’ Archives for all of the nations represented by CMC agencies. This will allow the DAI WG efforts to insure that benefit flows not only to the CMC space agencies, but to other organizations that are important to the governments of the CMC space agencies.

DAI WG internal notes:

Things that are not yet incorporated (or may not need to be?)

* The SM&C working group has had “near-term” archive work on it’s future work plan for some time, but it has not started.
* DAI and SM&C have jointly discussed the working group roles and the work plan for both near-term and long-term archive standards during the Rome meeting.