|  |  |
| --- | --- |
| **Ref.:** MOIMS/DAI/202404  Minutes of CCSDS MOIMS Data Archiving Ingest (DAI) Working Group  29 April-3 May 2024  Washington DC |  |
| **EXECUTIVE SUMMARY**  Over this week DAI Working Group has managed to resolve many issues affecting the draft BB for OAIS-IF, and re-organised the way in which the testing will be planned, carried out and summarised in the Yellow Book. | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subject:** Minutes of the DAI Working Group, | | | | | |
| **Meeting Date:** 29 April-3 May 2024 | | | **Place:** Washington DC, USA | | |
| Present: | | | Copies To: | | |
|  |  |  | CNES |  |  |
| UKSA | David Giaretta | DG (v) |  |  |  |
| ESA | Roberta Svanetti  Massimo Schiavon | RS (v)  MS (v) | ESA |  |  |
| NASA | John Garrett  Steve Hughes  Terry Longstreth | JGG  SH  TL | NASA |  |  |
| JAXA |  |  |  |  |  |
| Brazil |  |  | Various | Email list | DAI |
| SANA |  |  |  |  |  |

[**1 Mon 29th April 2024 4**](#_heading=h.30j0zll)

[1.1 OAIS-IF YB discussion 4](#_heading=)

[1.2 OAIS-IF BB Discussions 4](#_heading=h.9zhno78lqt63)

[**2 Tues 30th April 2024 6**](#_heading=h.f0sbvkrvtwc8)

[2.1 OAIS-IF BB Discussions 6](#_heading=h.shmq7z7ja18k)

[**Issues 9**](#_heading=h.spprqt7jo6lk)

# Mon 29th April 2024

## OAIS-IF YB discussion

The ESA contract programmer (MS) will tentatively finish the development of the second iteration of the OAIS-IF prototype in the Summer 2024, depending on the ESA/NASA use cases approval. He will be available to give feedback derived from the first iteration OAIS-IF prototype on BB 9am-10am Washington DC time on Tuesday.

RS has circulated a draft YB reflecting the ESA first prototype of OAIS-IF, with interoperability between two archives belonging to different parts of ESA.

DG proposed that we define the simplest acceptable YB contents and ensure we have that written while MS is available. We can then add other interoperability details as time allows.

Options:

* define curl commands which retrieve JSON. The JSON is verified against the documented general structure of what is expected and is validated against the JSON-schema (screen shot of GUI or of command line).
* check that where an AIP is returned then the Fixity hash is checked against the deserialised data object (or downloaded file from URI)
* the Optional aspects should be tested but note that they do not have to be implemented by either endpoint in real-life implementations.
* write the use cases as templates using “Endpoint 1” and “Endpoint 2”, then the tests could have a number of permutations of “Endpoint 1” and “Endpoint 2” as ESASDPA/ESASDTIPA/PDS(/DG), with separate test results.

Testing could use DG’s implementation, noting that this is not guaranteed to exist long term - BUT that does not matter for CCSDS YB.

Source codes could be made available on GITHub so that others could build implementations and check.

Check other examples of YBs - Tom Gannett said (DG thinks) that many YBs have not been published on CCSDS site.

## OAIS-IF BB Discussions

See BB latest version (v10)

<https://www.dropbox.com/scl/fi/fvzdqeagvq7mbkil41p2z/OAIS-IF-Core-Specification-v10.docx?rlkey=bcpthnlllppxdyrtx971y2u9l&dl=0>

Changes identified:

* need to add an explanation of how Fixity is used i.e. applies to the original data object before serialisation, or file pointed to by URI
* add to Conformity section (1.6) the kind of text in B1.1.2 i.e. that the underlying transfer mechanism maintains integrity of what is transferred e.g. TCP/IP  
  Also add that Authenticition is taken care of by that mechanism (e.g. Basic HTTP or API Key??)
* There was a discussion of BigInteger and the getSize() method for digital objects with large volumes. If the value is negative then this indicated that the size is not available or not known - since BigInteger does not have the equivalent of NaN. SH shared an example of PDS which provides min and max estimates.
* Add to schema as optional OaisIfTable and OaisIfTimeStamp etc
  + ACTION: SH will send text for PDS definition of how time is encoded in String
    - completed:

Name: ASCII\_Date\_Time\_YMD

Class Description: The ASCII\_Date\_Time\_YMD class indicates a date in YMD format and time.

Character\_Data\_Type - formation\_rule: YYYY-MM-DDTHH:MM:SS.SSSSSS(Z)

Questions about APIs

* do we have e.g. “api/GA/….” etc? Should we remove the “api/”???
* a proposal of naming convention was shared by DG as follows:

|  |
| --- |
| As discussed, I looked at examples of REST API's and found (again) <https://restfulapi.net/resource-naming/>  I think the API in the BB should be changed to be more consistent.  The article suggests:   * Use nouns to represent resources * Use “/” to indicate hierarchy * Use hyphens to improve readability * Do not use underscores * Use lower case letters in URI * Never use CRUD function names in URI   Here are some ideas:   1. rather than having the URI named "Get", the GET part is in the type of request, as it shows in the Swagger docs. 2. the resources are    1. Sources such as archives or users etc    2. Information which will be available packaged as general IP or AIP   i. With subtypes RepInfo and PDI elements   * 1. Data objects   So we should have something like (NOTE this has been edited following discussion atthe meeting)  [http://www.oais.info:8085/oaisif/v1/switchboard/sources](http://www.oais.info:8085/oaisif/sb/sources) to list switchboard entries  [http://www.oais.info:8765/oaisif/v1/generic-adapter/information](http://www.oais.info:8765/oaisif/ga/information)packages - to list information packages available  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage/{id](http://www.oais.info:8765/oaisif/ga/information?id=nnn)} –a specific info packages  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage/{id}/pdi](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn) –PDI of specific info [IF POSSIBLE]  ALTERNATIVE AS  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage/pdi](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)?id={id} ALTERNATIVE AS  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage-pdi](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)?id={id}  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage/{id}/pdi](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)/reference –PDI of specific info [IF POSSIBLE] etc  [http://www.oais.info:8765/oaisif/v1/](http://www.oais.info:8765/oaisif/ga/information/data?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/information](http://www.oais.info:8765/oaisif/ga/information/data?id=nnn)[package](http://www.oais.info:8765/oaisif/ga/information/pdi?id=nnn)[/{id}/data](http://www.oais.info:8765/oaisif/ga/information/data?id=nnn) – data object of specific info  [http://www.oais.info:8765/oaisif/](http://www.oais.info:8765/oaisif/ga/information/repinfo?id=nnn)[generic-adapter](http://www.oais.info:8765/oaisif/ga/information)[/informationpackage/{id}/repinfo](http://www.oais.info:8765/oaisif/ga/information/repinfo?id=nnn) – repinfo object of specific info  [http://www.oais.info:8083/oaisif/](http://www.oais.info:8083/oaisif/rr/information)representation-info-repository[/information](http://www.oais.info:8083/oaisif/rr/information)packages – list all repinfo in the RRORI  Note we should also VERSION the endpoints so instead of “oaisif/” we should use “oaisif/v1/” for the initial version. |

DISCUSSION:

* + - “sb” → switchboard
    - “ga” → generic-adapter”
    - “rr” → representation-info-repository
  + Alternatives
    - <http://www.oais.info:8765/oaisif/ga/information>packages?details=(size,id,datecreated)
    - <http://www.oais.info:8765/oaisif/ga/information>packages?detailoptions - returns list of strings that can be included in list

Question about query

* what to do when the Data Object is e.g. a ZIP file or encrypted? What should the RepInfo be? E.g. a Bagit file has various files within it. The initial RepInfo is that it is a Bagit file e.g. rfc8493. Where is the RepInfo for the individual files contained in the BagIt.
  + instead of just saying this is rfc8493, could point to text file which says this is a BagIt, and also defines the contents
  + OR could put a description of contents of Bagit in SemanticRepInfo
* If Provenance, Context etc are all in a METS file, should there be a specification of the specific section/page?? Should also include definition of Vocabulary. This should be discussed in the GB.

# Tues 30th April 2024

## OAIS-IF BB Discussions

OAIS-IF - MS’s shared observations based on earlier versions of BB (Nov 2023) and associated discussions.

The following was sent by MS.-

|  |
| --- |
| Here is a summary of the observations presented in today's meeting.  JSON payload Inconsistencies:   * Almost all field names use PascalCase except DataObject.size, InformationPackage.PDI.Provenance.DataObject.size and InformationPackage.version * The field InformationPackage.PDI.Fixity.RepresentationInformation.IdentifierType has a value of "URL" all the others use the more generic type "URI" * In the GIP example InformationPackage.InformationObject.RepInfo.OtherRI.OrGroup field is a JSON Object but in AIP example InformationPackage.InformationObject.RepresentationInformation.AndGroup.OtherRI.OrGroup is a JSON Array * In the GIP example, field InformationPackage.InformationObject.RepInfo instead of InformationPackage.InformationObject.RepresentationInformation   General suggestions:   * Use [RFC 7807 - Problem Details for HTTP APIs (ietf.org)](https://datatracker.ietf.org/doc/html/rfc7807) to represent errors * Avoid HATEOAS, stick to maturity level 2 of [Richardson Maturity Model](https://en.wikipedia.org/wiki/Richardson_Maturity_Model#Level_3:_Hypermedia_controls) * Endpoint paths should return a single type of data. For instance use an endpoint for collection of objects (ex. /pdis) and another for accessing a specific object (ex. /pdis/123 where 123 is the ID of the desired PDI). The first returns a list of PDI objects, the second a PDI object * It's ok to provide JSON Schema spec for JSON payloads, but the OpenAPI spec is also needed   Attached (pasted below):   * SpecificAdapterController.java - example of using OpenAPI and Spring MVC annotations to produce APIs and OpenAPI specification   package esa.ccsds.oaisif.proto.controller.rest;  import esa.ccsds.oaisif.proto.controller.rest.model.ArchivalInformationPackage;  import esa.ccsds.oaisif.proto.controller.rest.model.DataObject;  import esa.ccsds.oaisif.proto.controller.rest.model.Identifier;  import esa.ccsds.oaisif.proto.controller.rest.model.InformationPackageWrapper;  import esa.ccsds.oaisif.proto.core.InformationObject;  import esa.ccsds.oaisif.proto.core.PreservationDescriptionInformation;  import io.swagger.v3.oas.annotations.media.Content;  import io.swagger.v3.oas.annotations.media.Schema;  import io.swagger.v3.oas.annotations.responses.ApiResponse;  import io.swagger.v3.oas.annotations.responses.ApiResponses;  import jakarta.servlet.http.HttpServletRequest;  import org.springframework.http.ProblemDetail;  import org.springframework.http.ResponseEntity;  import org.springframework.web.bind.annotation.PathVariable;  import org.springframework.web.bind.annotation.RequestMapping;  import java.util.List;  import static org.springframework.web.bind.annotation.RequestMethod.GET;  import static org.springframework.web.bind.annotation.RequestMethod.POST;  @ApiResponses(value = {  @ApiResponse(  responseCode = "400",  content = @Content(mediaType = "application/json", schema = @Schema(implementation = ProblemDetail.class))),  @ApiResponse(  responseCode = "500",  content = @Content(mediaType = "application/json", schema = @Schema(implementation = ProblemDetail.class))),  @ApiResponse(  responseCode = "404",  description = "Object not found"  )  })  interface SpecificAdapterController {  @RequestMapping(path = "aips", method = GET)  ResponseEntity<List<InformationPackageWrapper<ArchivalInformationPackage>>> listAIPs();  @RequestMapping(path = "aips/{id}", method = GET)  ResponseEntity<InformationPackageWrapper<ArchivalInformationPackage>> getAIP(@PathVariable String doid);  @RequestMapping(path = "pdis", method = GET)  ResponseEntity<List<PreservationDescriptionInformation>> listPDIs();  @RequestMapping(path = "pdis/{id}", method = GET)  ResponseEntity<PreservationDescriptionInformation> getPDI(@PathVariable String doid);  @RequestMapping(path = "ios", method = GET)  ResponseEntity<List<InformationObject>> listIOs(); // @ParameterObject Pageable pageable);  @RequestMapping(path = "ios/{id}", method = GET)  ResponseEntity<InformationObject> getIO(@PathVariable String doid);  @RequestMapping(path = "dos/{id}", method = GET)  ResponseEntity<DataObject> getDO(@PathVariable String id);  @RequestMapping(path = "dos/download/\*\*", method = GET)  ResponseEntity<?> download(HttpServletRequest request);  @RequestMapping(path = "dos/upload", method = POST)  ResponseEntity<Identifier> upload(HttpServletRequest request);  }   * application.yml - Spring Boot properties I use to configure Spring and SpringDoc   + springdoc:   swagger-ui:  tagsSorter: alpha  api-docs:  version: OPENAPI\_3\_1  spring:  jackson:  property-naming-strategy: com.fasterxml.jackson.databind.PropertyNamingStrategies.UpperCamelCaseStrategy  default-property-inclusion: non\_null |

After review of the latest version of the BB, MS informed us that OpenAPI is aiming to provide some of the JSON Schema functionality, but it is acceptable to use JSON Schema in the BB.

<https://oaisif-esa-proto.185.121.38.206.sslip.io/swagger-ui/index.html#/>

[oaisif-esa-proto.185.121.38.206.sslip.io/v3/api-docs](https://oaisif-esa-proto.185.121.38.206.sslip.io/v3/api-docs) shows JSON for API including error codes (from the code). This can be processed to take just the GET endpoints to create text to put in the BB.

Example code :

@RequestMapping(path = "aips/{id}", method = GET)

ResponseEntity<InformationPackageWrapper<ArchivalInformationPackage>> getAIP(@PathVariable String doid);

[oais.info:8765/api/GA/GetConfig](http://www.oais.info:8765/api/GA/GetConfig) lists properties. Add a new subelement under each property which is a URI pointing to more details e.g.

{

"id": 1713976929396,

"propertyName": "MYQUERYMETHOD",

"propertyValue": "SQL"

}

turns into

{

"id": 1713976929396,

"propertyName": "MYQUERYMETHOD",

"propertyValue": "SQL"

“propertyURI”: “http……..” - to provide details of the query method.

}

## Plan for completion of BB

1. Make corrections to naming of variables as described by MS
2. Include section to describe endpoints - see above
3. Complete JSON Schema to include OAISIfTable etc
4. Add section with OpenAPI documentation for GETs. This will be generated from <http://www.oais.info:8765/v3/api-docs>, extracting the GET parts. This may require updating the s/w to match then generating the file.

## Issues

* Where to document Identifier type etc? Maybe in GB or is it needed in BB?
* Queries.

# Wednesday 1st May 2024

Discussion of Representation Information required for OAIS-IF took place. Some of this discussion could be included in the GB.

The topics included how to express RepInfo for “container objects” e.g. ZIP file or BAGIT files, noting that many common file types such as Word and PDF can also be used as containers.

Besides the format (zip, bag, word etc) one certainly also needs Semantic RepInfo.

The contents of the container could be described in that Semantic RI, or there may be information inside the container which provides the RepINfo - we would have to identify how to get to it. Clearly there are ways to do it but it is not obvious how to standardise such a description.

The good thing is that it is not the job of the OAIS-IF BB to provide the answers to this question.

The test cases do not have to include full Representation Information. The BAGIT discussion was started by the observation that ESA used BAGIT files in one of the archives.

## ISO 16363 RIDS

All the RIDs for ISO 16363 and ISO 16919 were closed (see below) so the appropriate changes should be made in the docs before submission to the CCSDS Technical Editor.

* <http://review.oais.info/show_bug.cgi?id=390> DE-01 Unintended repetition of words and parts of a sentence
  + change made and RID closed.
* <http://review.oais.info/show_bug.cgi?id=391> DE-02 Delete footnotes - links invalid
  + footnotes deleted and RID closed.
* <http://review.oais.info/show_bug.cgi?id=392> DE-03 Add a current publication about DIN 31644 and the certification at Nestor to this standard.
  + reference added and RID closed.
* <http://review.oais.info/show_bug.cgi?id=389> Duplication between 3.1.2.1 and the new metric 4.3.5
  + 4.3.5 was re-written and RID closed.
* <http://review.oais.info/show_bug.cgi?id=349> Clarify metric 4.4.1 - Ensure that Software is an acceptable, though not ideal, form of RepInfo
  + Metric was clarified and RID closed.

## Climate change discussion

ISO has circulated the following:

|  |
| --- |
| 2. Addition of Climate Change considerations to the Management Systems Standards  2.1 Decision  In support of the ISO London Declaration on Climate Change, ISO passed a resolution that will result in two new statements of text being added to a number of existing management systems standards, and will be included in all new standards under development/revision, to address the need to consider the effect of Climate Change on the ability to achieve the intended results of the management system. The changes will be introduced initially as Amendments to these published standards.  The changes (two new statements) will be incorporated into the new text of the Harmonized Structure (Appendix 2 of the Annex SL in the ISO/IEC Directives Part 1 Consolidated ISO Supplement) as follows.  The amendments to the standards will be published on February 23rd 2024. |
| 4.1 Understanding the organization and its context.  The organization shall determine external and internal issues that are relevant to its purpose and that affect its ability to achieve the intended result(s) of its XXX management system.  Added: The organization shall determine whether climate change is a relevant issue.  4.2 Understanding the needs and expectations of interested parties.  The organization shall determine:  • the interested parties that are relevant to the XXX management system.  • the relevant requirements of these interested parties.  • which of these requirements will be addressed through the XXX management system.  Added: NOTE: Relevant interested parties can have requirements related to climate change. |

The document referred to is at [iso.org/sites/directives/current/consolidated/index.html#\_Toc165305044](https://www.iso.org/sites/directives/current/consolidated/index.html#_Toc165305044)

Discussion, for decision tomorrow if possible:

* Add these notes to ISO 16363, metric 5.1.1
* Additionally add to relevant sections of ISO 16919.

## ACTIONS

* DG: check if there is a table for DE comments to ISO 16363. If so, complete it as was done for other ISO comments.
* DG:Update ISO 16363/ISO 16919 following RID resolution.
* JGG/DG: DAI resolution: Request MC to publish the following as MAGENTA books:
  + CCSDS 653.0-R-1: Information Preparation to Enable Long Term Use (IPELTU) (ISO 23507)
  + CCSDS 650.0-P-2.1: Reference Model for an Open Archival Information System (OAIS) (ISO 14721)
  + CCSDS 652.0-P-1.1: Audit and Certification of Trustworthy Digital Repositories (ISO 16363)
  + CCSDS 652.1-P-2: Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories (ISO 16919).

We were informed by the MOIMS AD that he needed:

|  |
| --- |
| the final version of the word files with all the RIDs from the Agency Reviews implemented (ideally as track changes) and the RID settlements.  With that I can issue the resolutions for publication. |

DG has informed the ISO contact that we will have to wait for the final version from the CCSDS Technical Editor.

# Thursday 2nd May 2024

Reviewed and updated the report to MOIMS plenary, including resolutions to MC for the Magenta books.

See <https://cwe.ccsds.org/moims/docs/MOIMS-DAI/Meeting%20Materials%20and%20Minutes/2024-04-Washington-DC/Report%20to%20MOIMS%20DAI%20202404-closing.pptx?d=w7027ae45800c42e9aadaa835e1e38a82>

Reviewed and updated CWE Framework [Projects - All Approved, Pending, and Completed Projects (ccsds.org)](https://cwe.ccsds.org/fm/Lists/Projects/AllOpenChartersWithAllProjects.aspx).

ACTION: JGG will finish the updates to CWE Framework.

ISO contact has informed us (16 April 2024):

|  |
| --- |
| Further to our ongoing discussions about our joint work, we are pleased to report that a new Strategic Partnerships Unit has recently been created in the ISO Central Secretariat.  I have engaged their support so that we can create a plan for moving forward with all aspects of our cooperation.  We will come back to you shortly with few proposed dates to have a call. |

## Further discussions of OAIS-IF JSON serialisation

JGG raised a number of issues about the use of Identifiers in the JSON which is transferred.

The Identifier must be usable to download the bits of the Data Object to which this Identifier points. Note that this MAY require authentication, which is not specified in the BB.

Another consideration is that, as is well known, URLs (as a type of URI) or indeed any other type of Identifier String, may not be resolvable forever. It may be useful to have an optional parameter providing the date up to which the URL is guaranteed to be resolvable, but beyond which no guarantees are provided.

This would be useful in the case of transfer of “AIP”s to a successor archive from an archive which is going to stop operating beyond a specific date. If the optional parameter is not present then no assumption should be made about the longevity of the resolvability of the Identifier, other than what would normally be made for any Identifier. Such a parameter could be:

|  |
| --- |
| {  "Identifier": "<http://opm.example.com/java-sw-system>",  "IdentifierType": "URI"  “ResolvableUntil”: “2024-05-03T12:13:14.5678Z”  } |

“Identifier” and “IdentifierType” are required but “ResolvableUntil” is optional.

The JSON-Schema is available at <http://www.oais.info/oais-if/json-schema/infopackage.schema.json>, with example AIP <http://www.oais.info/oais-if/json-schema/aip-example.json> and an example IP is <http://www.oais.info/oais-if/json-schema/ip-example.json>, all of which will be updated.

To support this the following was proposed for the BB:

|  |
| --- |
| 3.4 IDENTIFIER INTERFACE  The Identifier interface provides a unique (resolvable) name (character string, plus Identifier type e.g. URI, URN, local) to an entity to identify and provide its location from which the bits can be downloaded.  This is not necessarily a globally unique ID because it is used for some ephemeral things like messages.  WE SHOULD ADD:  **The optional parameter ResolvableUntil defines ……** |
| The following was also a possibility but was rejected:  *OR*  *The length of time for which the identifier is resolvable is not defined within this interface*. |

There followed vigorous discussion of the JSON serialisation and detailed examination of the JSON Schema proposed.

At the end of the day, it was tentatively accepted that there were no major errors.

# Friday 3rd May 2024

Discussion with Charles Sheehe (GRC-LCN0) about what DAI is doing and how it might be used for his needs in terms of privacy of information for LunaNet.

DAI described a number of possibilities:

* DAI uses the concept of Information i.e. Data plus RepInfo
* The Data Object could be encrypted and the RepInfo would tell one that the data should be decrypted using a secret key. Once decrypted additional RepInfo will tell one how to understand/use the information.
* In an AIP one could place Access Rights information i.e. who should be allowed to access the data.
* The OAIS-IF BB shows how these should be encoded in JSON, but one could use XFDU if a binary, compressed file if bandwidth is limited e.g. LunaNet.

## Review of JSON serialisation.ACTIONS

Massimo (MS) to review the REST endpoints - especially the blue vs red text as alternatives for selecting packages e.g.

* baseurl/{id}
* baseurl?id={id}

and send feedback on Monday.