PDAP Assessment Project

Current status of PDAP Assessment

- Achievement:80%(April, 2009 for PDAP version 0.3)
  - 20%: MUST make a document: Assessment How-to
- Assessment is a dangerous work to discuss about technical points.
  - Document Assessment may work well
  - Assessment of Technical parts is hard work because
    - Protocol sometimes depends on System
    - Protocol sometimes depends on Data Model
    - Simplicity of protocol sometimes requires complex implementation for environments around PDAP.
    - There will be no differences between design and assessment if drastic changes are required.
- A/I
  - Making a document: “Assessment How-to”

PDAP Assessment Document version 0.4

Release Notes

Version	Release Date	Page	Comment
0.1	Jan. 12, 2009	-	Initial release
0.2	Jan. 20, 2009	16-17	Image service assessment update
0.3	Feb. 20, 2009	18	TARGET_NAME description
0.4	Feb. 21, 2009	19	Wildcard discussion

1. Lack of the version control of PDAP itself

- PDAP protocol should define the method to carry the PDAP protocol version to keep the compatibility for further extension.
  - Example1: Implementation of query for version
    - http://somewhere.pdap.net/?getVersion
  - Example2: Version information is always included in each transactions.

```
<VOTABLE version="1.1">
 <PDAP version="0.3" /PDAP>
 <RESOURCE type="result"> …
```

2. Lack of the repository information in PDAP general output

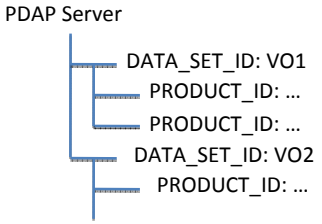
- PDAP server should guarantee that the same input derive the same output unless the product repository change. If the output changes in spite of the same input, the product repository is changed. The date and time of updating repository should be included in the general output to tell client.

Example:

```
<VOTABLE version="1.1">
 <PDAP version="0.3"
      lastUpdate="2009/01/10 19:49:10" /PDAP>
 <RESOURCE type="result"> …
```


3. Lack of the mechanism obtaining mandatory input

- For uses, mandatory input such as DATA_SET_ID is often complex and difficult to search.
- For further applications, the directory-like structure will require the mechanism to obtain these compulsory input.



4. Lack of the mechanism to identify products in PDAP Server

- In most cases, the completely same products are required by reference. PDAP specification should define the product identification.
- Also PDAP should provide the mechanism to query by the ID.

Example:

```
<TABLE>
    <FIELD ID="PDAP.UNIQUE_ID" ucd=…
    <FIELD ID="DATA_SET.DATA_SET_ID" ucd="DATA_SET_ID" utype="…
    <DATA>
        <TABLEDATA>
            <TR>
                <TD>0000-1111-2222-3333-AAAA-BBBB-CCCC-DDDD</TD>
                <TD>SELENE_Moon_TC-1_Level2B</TD>
```


5.General Service Output (VOTable) should include the number of data

- For user queries, a number of products may be matched and PDAP server may stop during transferring data due to timeout. Then PDAP client has no information from the PDAP server. To keep from this situation, the limit of products should be defined and PDAP should support the offset and limit mechanism. For more precise transaction, TRANSACTION_ID will be required.
- For implementation, the number of products is required for memory allocation. Without counting the data fields, the performance will be increased.

Not PDAP, but VOTable Problem

Example:

```
<DATA>
<TABLEDATA OFFSET="0" COUNT="1000">
    <TR> … </TR>
    …
    <TR> … </TR>
</TABLEDATA>
```

6. Image Service extended Output Fields should contain reference frame

- For instance, there are two lunar frames – Principal Axis (PA) and Mean Earth/Polar Axis (ME). Without frame information, user will confuse where the region is specified.
- Considering technical implementation, the name of frame should be the same as other systems, such as SPICE.

7. Definition of the unit

- Image service or spectrum service will require the unit of values, such as deg or rad, eV or nm, etc. Planetary Data Access Protocol Version 0.3 should describe the concrete specification of unit, and its reference such as: *Astronomy and Astrophysics Supplement*, v.**143**, p.23-32.

8. Writing

- Example of START_TIME/STOP_TIME must include 'T' between date and time. In addition following formation rule, the description of millisecond must be 3 digits followed by zeroes if it is written.
  - START_TIME: 1997-03-03T16:18:39.000
  - STOP_TIME: 1997-03-11T19:29:36.000

9. Feedback from the trial implementation of Kaguya PDAP

- REFERENCE_FORMAT, PUBLISHER, CONTRIBUTOR, PUBLISHING_DATE, RIGHTS are mandatory, but these are not well defined.
- The server providing DATA_SET_ID in PDAP system will be useful for users.

Use-case Analysis Feedback

Compulsory Input Definition

- Considering use-case analysis of scientific users, the search words are categorized as followings:
  - Target (Earth, Moon, Mars, Asteroids, etc)
    - TARGET_TYPE
  - Field (mineralogy, geology, aeronomy, volcanology, seismology, etc.)
    - TBD (too difficult ! How to categorized ?)
  - Data Type (image, spectrum, GIS, ancillary data, etc.)
    - TBD
  - Instruments (Visible Camera, Near-Infrared, X-ray, Ultra Violet, etc.)
    - INSTRUMENT_NAME

Suggestion for Design 1.

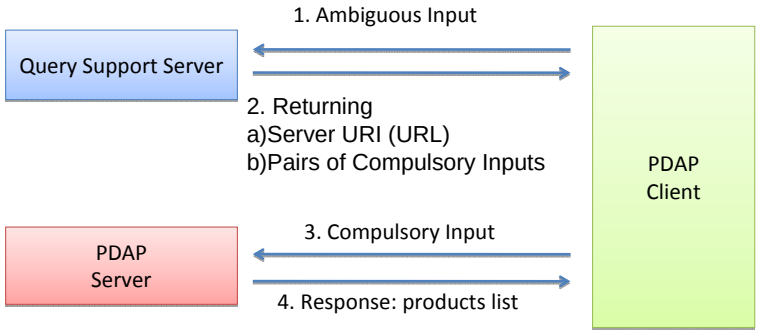
DATA_SET_ID is not essential in General Service

- For implementation of database, user can search products without DATA_SET_ID input. Removing this mandatory input allows users to search by time, or area (lon/lat).
- For example, if user provides keyword 'TARGET=Mars' without DATA_SET_ID, PDAP server can response correct items.

Suggestion for Design 2.

Query Support Server

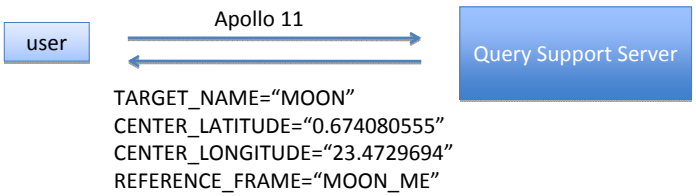
It is useful that the product list or keywords to obtain product list from ambiguous input.



Merit:
1.General Service itself keeps simple service.

Suggestion for Design 2.

Example of Query Support Server



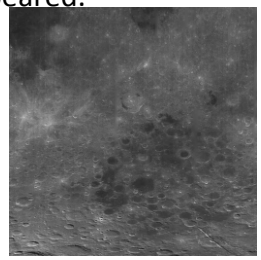
PDAP system can return information it stored in its system, but common information such as name of local area cannot be return.

Query keyword is always confusing users, and the correct pairs of keywords will support users. This is often embedded in HTML, but the synchronization will not be ensured for other systems.

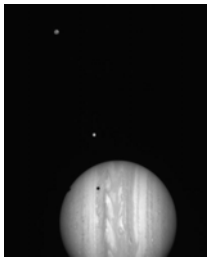
Not only PDAP, but also SIAP/SSAP will be useful if Query Support Server can support astronomical tables.

Image Service1:Definition of CENTER_LONGITUDE and CENTER_LATITUDE

- Some images do not have the center: CENTER_LONGITUDE/CENTER_LATITUDE. The return value should be written in the documents if the center is not appeared.



TARGET: MOON
(lon,lat)=(-35,278)
Taken by Clementine



TARGET: JUPITER
(lon,lat)=(???,???)
Taken by New Horizons

Image Service2: Coordinates of the four corners

- How to define the coordinates of the four corners if the target is completely enclosed ?

Upper Left (?,?) Upper Right (?,?)

LowerLeft (?,?) Lower Right (?,?)

Idea1:

UPPER_LEFT_LONGITUDE = CENTER_LONGITUDE – 90

UPPER_LEFT_LATITUDE = CENTER_LATITUDE – 90

…

It is convenient for user queries. The coordinate of hemisphere is searchable.

Must describe the reference point for TARGET_NAME

- The document MUST describe some reference of the candidates for TARGET_NAME. When TARGET_NAME is a planet, an asteroid, or astronomical body defined in IAU, the TARGET_NAME value should be uniformed between PDAP system for interoperability.
- In addition, the document MUST define upper case or lower case to specify (or equivalent to each cases).

See also:

http://www.iau.org/public_press/themes/naming/

http://planetarynames.wr.usgs.gov/append7.html

Discussion of including Wildcard

- Usablility
  - Considering usability PDAP should permit wildcard in input of each parameter. For instance,in a big project, there are too many DATA_SET_IDs, PRODUCT_IDs and they are hard to be expected.
- Security
  - Wildcard is often used to do security attack for relational database system. The implementer should care about the security matters if wildcard is adopted.
- Simplicity
  - Wildcard makes protocol complicated which means the response of PDAP system slower.

Attribute

Project Specific Parameter

- PDS, PSA, and any other format have their own meta-data respectively. Should PDAP support query interface for these parameters ?

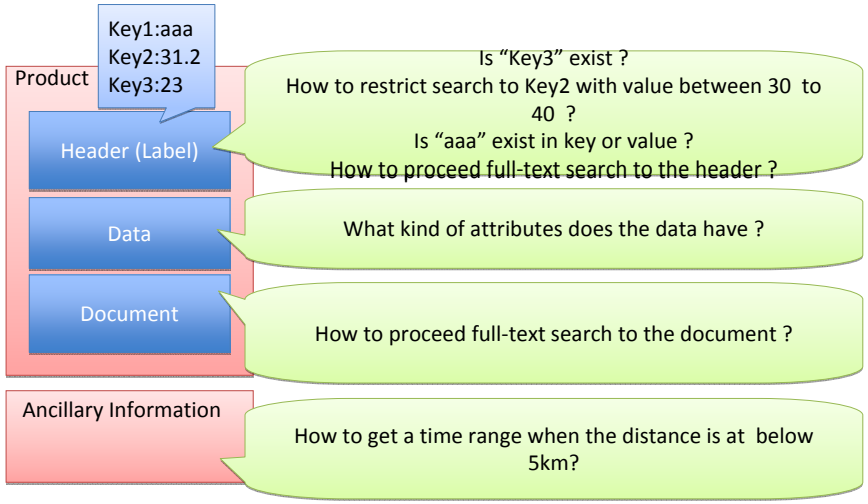
PDAP a la carte

ISAS/JAXA Yukio Yamamoto

Flexibility of Time specification
Not “START_TIME”,”STOP_TIME” But “TIME”

- “TIME” will be useful to specify intermittent time list following other range definition for input parameters.
  - Ex1) from 2005-10-01 00:00:00 to 2005-10-01 23:59:59
    Current:
    &START_TIME=2005-10-01T00:00:00&STOP_TIME=2005-10-01T23:59:59
    New:
    &TIME=2005-10-01T00:00:00/2005-10-01T23:59:59
  - Ex2) from 2005-10-01 00:00:00 to 2005-10-01 23:59:59 or 2005-10-03 00:00:00 to 2005-10-03 23:59:59
    Current:
    CANNOT !
    New:
    &TIME=2005-10-01T00:00:00/2005-10-01T23:59:59,2005-10-03T00:00:00/2005-10-03T23:59:59

Data Structure and Requirements for PDAP



Parameter “PDAP_PARAM_SCOPE”

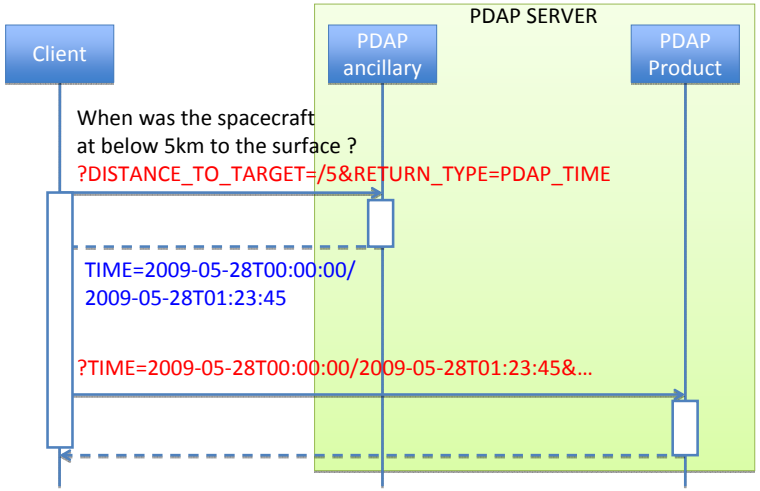
To keep Interoperability, search parameter MUST be standardization, while the name of header labels depend on each projects. PDAP will be available if any labels in product headers are searchable.

PDAP_PARAM_SCOPE=PDAP keeps interoperability between other PDAP servers.
PDAP_PARAM_SCOPE=PRODUCT can search with any keyword in product headers.

ex1)
? PDAP_PARAM_SCOPE=PRODUCT&TEMPERATURE=30.0/50.0

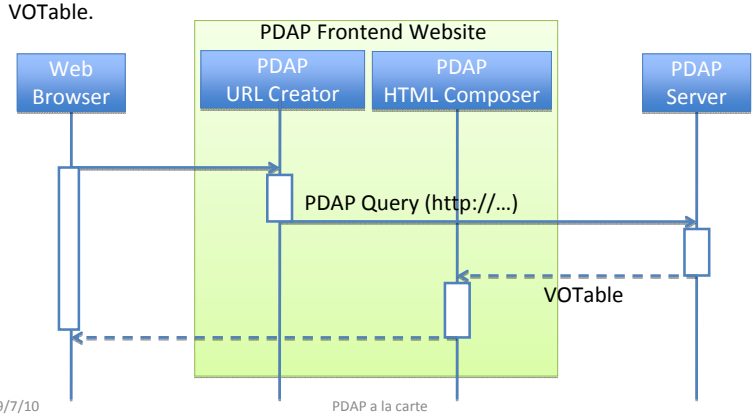
ex2) INST_MODE=A or INST_MODE=B
? PDAP_PARAM_SCOPE=PRODUCT&INST_MODE=“A,B”

PDAP Ancillary and RETURN_TYPE=PDAP_TIME



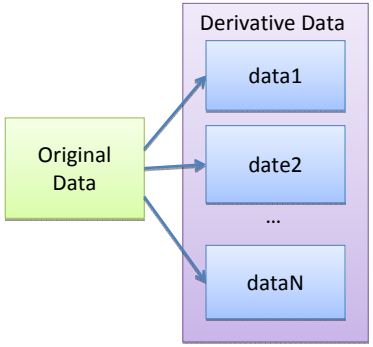
PDAP Optional RETURN_TYPE=HTML is dangerous

RETURN_TYPE=HTML is dangerous because the requirement will be lost due to its convenience. For PDAP users, both RETURN_TYPE=HTML and RETURN_TYPE=VOTALBE can transfer same information, and the differences must be only format. HTML for human interface should be created by VOTable.



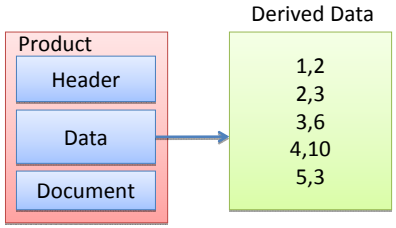
PDAP Optional
Concept of Interface for interoperability

- Currently PDAP Response (VOtable) format depends on the RESOURCE_CLASS parameter. Concept of INTERFACE defines a group of derived data and a response group. RESOURCE_CLASS controls PDAP INPUT and OUTPUT based on original data, while INTERFACE defined PDAP OUTPUT first, then derived data from original data is ensured the INTERAFCE-required format.



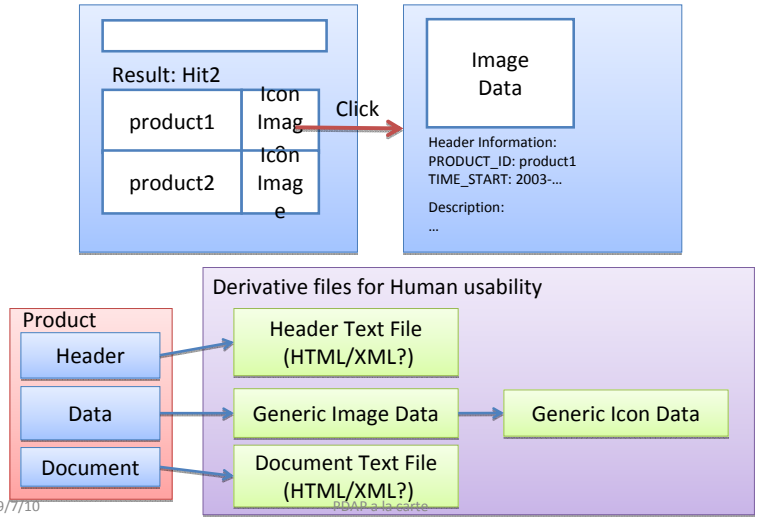
VOTable Response:
<…
<FIELD ID="INTERFACE:DATA1" …/>
<FIELD ID="INTERFACE:DATA2" …/>
…
<FIELD ID="INTERFACE:DATAN" …/>
…
<TD>
<![CDATA[http://…/data1.dat]>
<![CDATA[http://…/data2.dat]>
…
<![CDATA[http://…/dataN.dat]>

PDAP Optional
Interface example1: XY-PLOT Interface



VOTable Response:
<…
<FIELD ID="XY-PLOT:TITLE" />
<FIELD ID="XY-PLOT:X-UNIT" />
<FIELD ID="XY-PLOT:Y-UNIT" />
<FIELD ID="XY-PLOT:X-LABEL" />
<FIELD ID="XY-PLOT:Y-LABEL" />
<FIELD ID="XY-PLOT:data" />
…
<TD>X-ray Spectrum</TD>
<TD>keV</TD>
<TD>count/sec<TD>
<TD>Energy</TD>
<TD>Intensity</TD>
<TD><![CDATA[http://…/dataN.dat]></TD>

PDAP Optional
Interface example2: Browser Interface



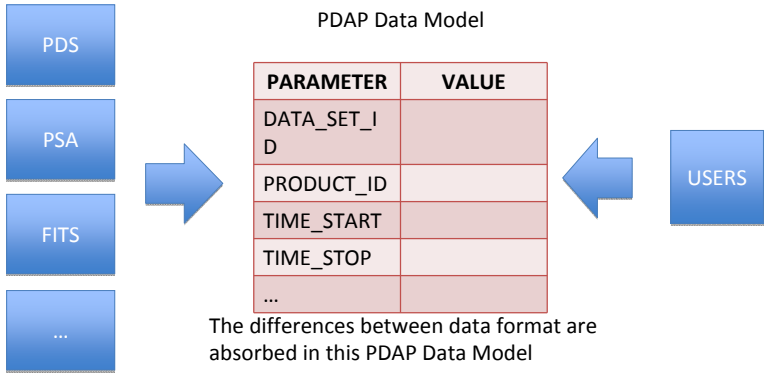
PDAP Optional
Interface example3: Other Group Interface

- Astronomy I/F
  - VSOP (SIAP, SSAP, VOTable)
  - SDSS (SQL I/F)

Discussion Point

- Balance between Generic and Specific
  - PDAP is a kind of web service interface
  - Generic Service is already considered and there are some candidates:
    - SOAP, SMPP, RESTful, Google Web Service
  - To expand PDAP specification, PDAP become a Generic Service in the end.

Required IPDA Data Model Definition (adoption of PDS Labels ?)



The differences between data format are absorbed in this PDAP Data Model

Grazie !