Anne-Marie Gontier et al.: PIVEX: a Proposal for a Platform Independant VLBI Exchange Format, IVS 2002 General Meeting Proceedings, p.248–254
http://ivscc.gsfc.nasa.gov/publications/gm2002/gontier

PIVEX: a Proposal for a Platform Independant VLBI Exchange Format

Anne-Marie Gontier, Martine Feissel

Observatoire de Paris, SYRTE
Contact author: Anne-Marie Gontier, e-mail: Anne-Marie.gontier@obspm.fr

Abstract

Following a recommendation of the 2nd IVS Analysis Workshop, a working group was set up to develop a VLBI exchange format independent of platforms and operating systems. The aim of this format is to ease the international extension of VLBI analysis for astrometry and geodesy. The working group members were selected in such a way that their knowledge should collectively encompass all data environments involved. We present, in this paper, the PIVEX format that was discussed and tested and the future insertion of the PIVEX files into the VLBI data flow.

1. The Working Group

The 2nd IVS Analysis Workshop (February 2001) created a Working Group to "organize the discussions and activities concerning the development of a new VLBI database structure independent of platform and operating system". The membership of the WG was based on voluntary participation. It encompasses various data environments involved, namely:

- Correlators: MarkIV, Japanese, Canadian.
- Analysis software: CALC, ERA, GEOSAT, GLOBK, GLORIA, MODEST, OCCAM, STEELBREEZE.
- Data Centers.

The members list is as follows:

Martine Feissel (Chair),	Najat Essaifi,	Anne-Marie Gontier,
Calvin Klatt,	Chopo Ma,	Leonid Petrov,
Kazuhiro Takashima,	Oleg Titov.	

Ed Himwhich and Dan McMillan were involved in the intensive discussions on the PIVEX file structure and information content, held during a three weeks stay of the first author in the VLBI group at GSFC. Further more, the following individuals were kept informed of the work progress:

P.H. Andersen,	J. Boehm,	S. Bolotin,	P. Charlot,	A. Fey,
Y. Koyama,	Z. Malkin,	A. Niell,	A. Nothnagel,	G. Elgered,
H. Schuh,	O. Sovers,	V. Tesmer,	N. Vandenberg.	

2. Environment

Historically, various data treatment levels downstream of the correlator were quite naturally implemented jointly in the Calc-Solve operation. The use of the VLBI data base files is currently integrated into the way GSFC does information transfer between programs, archiving and cataloging. The files consequently contain more than the raw fringe output and calibration data, the minimum information required for analysis. Today, with the multiplication of analysis software packages, it seems again natural to isolate a level of data exchange that will help newcomers to join, without the obligation to extract for themselves the information they need, out of informations dedicated to another software package, implying other modellings, other aprioris, etc. The minimum amount of information needed for analysis depends on what each analysis center wishes to do independently. Generally, for example, analysis centers using the current data base files have not redone the weather calibration from the station logs.

In the current scheme, a series of operations internal to Calc-Solve are performed prior to the posting of the databases on the IVS Data Centers. The external user must then select the part of information he needs (less than 1/3 of the total) to import the observations into his own computing environment. In the scheme the WG is aiming at, the Data Centers files contain only the useful information and can be imported by the analysts using a simple interface.

The current VLBI format consist of a database handler (called Mark3-dbh), two binary (CALC database) and two ascii (NGS) files. The database handler allow us to load in memory one experiment at one and the same time which does not make the global analyses easier. For the same experiment, both in binary and ascii format, the X- and S-band data are stored in two files independantly which leads to unnecessary redundancy. The binary and the ascii files differs by their content and structure. The binary file contains the observation informations as well as models and partial derivatives computed by the CALC software. The ascii format is fixed and documented but some observational informations, like the number of ambiguities, are missing. A proposed astro-geodetic VLBI format is described hereafter.

3. Astro-Geodetic VLBI Format

This project had already been treated in the two following instances, that served as a starting point for the WG discussions.

- Specifications for a "geo-VLBI format", proposed by L. Petrov [1].
- The Gloria database structure, that was discussed internationally before its final adoption in 1999 [2].

The goals of the astro-geodetic VLBI format may be summarized as follow.

- Develop a flexible and efficient format for keeping and sharing astro-geodetic VLBI data. The criteria for optimization, in this context, are the minimization of the size of a database, the minimization of redundancy and the minimization of the time needed for reading/writing the database.
- Develop a share-ware library (Geo Vlbi Handler: GVH) which would support basic operations with datafiles. This library should run under several platforms (ideally under any platform).

The astro-geodetic VLBI format consists of a description of internal representation of the data in memory, a description of two external datafile formats on disk and a user interface (GVH).

In the future, two datafile formats will be available on disk.

- A binary format (bgv) which follows closely the internal structure of the data in memory and is oriented for processing experiments in automatic modes. This format could replace, one day, the actual CALC database.
- An ASCII format (PIVEX) more oriented for reading by a human being. It will replace, as soon as possible, the NGS files in IVS Data Centers.

The GVH procedures is a set of subroutines for the data manipulation. It provides the service of reading information from the input file(s) to the user program and writing information from a user program to the output file(s). User access to the information, at least for the binary files (bgv), will be through this standard interface. Generators of data for submission to IVS, essentially only the Operation Centers, will use the routines that pass information into PIVEX. Users doing analysis will only need to use the subroutines that retrieve information. Users may, of course, write their own code to access the PIVEX files.

The geo VLBI handler is able to load more than one experiment in memory and is an efficient tool to manipulate the data. The datafiles contain only observation informations like output of fringe fitting, parameters of observing session, correlator comments, calibration information or history records and is identical either in binary and ascii form. Some of the key structural points are that X and S band is merged with common access code for data item, all data are trated as two-dimensional although second dimension can be one and the format will be self-documenting and extensible by table of contents.

A description of the PIVEX file structure resulting from intensive discussions is given hereafter.

4. PIVEX Data Organisation

4.1. Supported Data

The data supported are of three kinds: text information, description of numerical data and numerical data themselves.

The text information is organized in one or more chapters. Each chapter has one line of title and a body consisting of one or more lines of variable length.

The numerical data consist of several arrays of access codes called **lcodes**. They could be of fixed or variable length and contain the data.

The description of numerical data is an array of fixed length which describe all access codes. It consists of a name, type, class, dimensions and a short explanation for each **lcode**. The following types are actually supported:

INTEGER*2 REAL*4 CHARACTER*1
INTEGER*4 REAL*8

The format can be easily extended to other types. In order to avoid data redundancy, as much as possible, all information is classified as session, scan, station or baseline.

4.2. File Structure

Each datafile has the following structure:

- one line of identification label which tells the name of the format and the revision date.

PVX format of 2001.11.25 32 bit address

- one or several segments consisting of 5 sections (preamble, text, toc, data and heap).

Each section consists of a reserved keyword and a body. A brief description and a short example (in italic) of the different sections follows.

- Preamble: keyword **@PREA**
  The body of this section consists of records of variable length, it contains general information on data type, class definition, generator ... For example:

  @PREA section_length: 13 keywords
  DEF_TYPE: 1 CHARACTER ASCII
  *DEF_TYPE: 2 INTEGER*2 IEEE-231*
  DEF_CLASS: 81 Session
  DEF_CLASS: 84 Baseline
  GENERATOR: ex1 GVH release of 2001.11.28
  FILENAME: sample.bgv
  CREATED_AT: 2002.02.26-15:22:07

- Text: keyword **@TEXT**
  The section consists of records of variable length separated by terminator and organized in chapters and paragraphs. History of versions, correlation and post-correlation information are written in this section.

  @TEXT section_length: 2 chapters
  @@chapter 1 History of version 1
  Dbedit: na444, NEOSA — geo_export
  VLBI experiment NA444 (NEOSA)
  @@chapter 2 History of version 2. Created 2001-11-04T15:48:12 UTC
  CALC 9.12 Ver. 2001.01.12 Tue Nov 20 10:48:02 2001 leo

- Tocs: keyword **@TOCS**
  It consists of records of fixed length. It is the table of contents in the form lcode name, class, type, two dimensions and a short description.

  @TOCS

NUMB_OBS	*SES*	*I4*	1	1	*Number of observations in the session*
SOURCE	*SCA*	*C1*	8	1	*Source name*
LO_FREQ	*STA*	*I2*	22	16	*Local Osciallator frequencies per channel in MHz*
GRIONFRQ	*BAS*	*R8*	2	1	*Effective ionosphere frequency for group delay (MHz)*
PHASE_AP	*BAS*	*R4*	−16	−512	*Fringe phase per channel, per AP (rad)*

- Data: keyword **@DATA**
  The section consists of records of variable length. It contains the observations organized in four classes: session (**@@Session**), scan (**@@Scan**), station (**@@Station**), baseline (**@@Baseline**). It is value of lcodes of fixed length (dimensions are specified in the Toc section).

```
@DATA
@@Session
  NUMB_OBS 20
  CORPLACE HAYSTACK
@@Scan        1
  SEC_TAG 1.01
@@Station GILCREEK        1
  LO_FREQ 234 212 494 407 318 242 1093 463 511 349 1007 864 1129 750 850 554 887
441 731 221 822 641 564 843 605 959 1122 736 295 360 973 458
@@Station KAUAI        3
  LO_FREQ 1178 1048 368 684 417 320 1050 657 542 1167 582 550 828 887 1144 466 218
883 1132 709 1127 316 393 916 734 282 945 838 1120 692 901 598
@@Baseline GILCREEK KAUAI        1
  GROBSDEL .5040884613990784 .1870681047439575
@@Baseline GILCREEK NRAO85 3 2
  GROBSDEL .1870681047439575 4.285985231399536E-02
@@Scan        2
  SEC_TAG 2.02
```

- Heap: keyword **@HEAP**
  It is one record of variable length. It contains the value of lcode of variable length. The maximum dimension is specified in the Toc section with negative value and the Data section contains the descriptors.

```
@HEAP
  PHASE_AP 0.123 0.456 1.3456
```

A summary of the PIVEX file structure is shown in figure 1.

5. PIVEX Prototype

5.1. Tested Experiments

In order to test the reading/writing code of the geo VLBI handler, we have written a small session prototype. The code to write informations from memory into either binary or ascii files is fully tested on a Unix HP computer. Moreover, the code to load informations in memory from a binary file is available and the same operation from the PIVEX file is in progress.

We also have tested the possibility to transform the actual CALC database to PIVEX format. Three different sessions were used for that purpose:

- one intensive experiment (01DEC31_U, INT01-365), 2 stations, 18 observations;
- one NEOS experiment (01DEC11_E, NEOS-A450), 6 stations, 1954 observations;
- one RDV experiment (01MAY09_A, VLBA28), 18 stations, 24634 observations.

All the corresponding PIVEX files were created and even for the larger one, the ascii file is still manageable. The remaining programming work is to recode the GVH routines for multiple platforms.

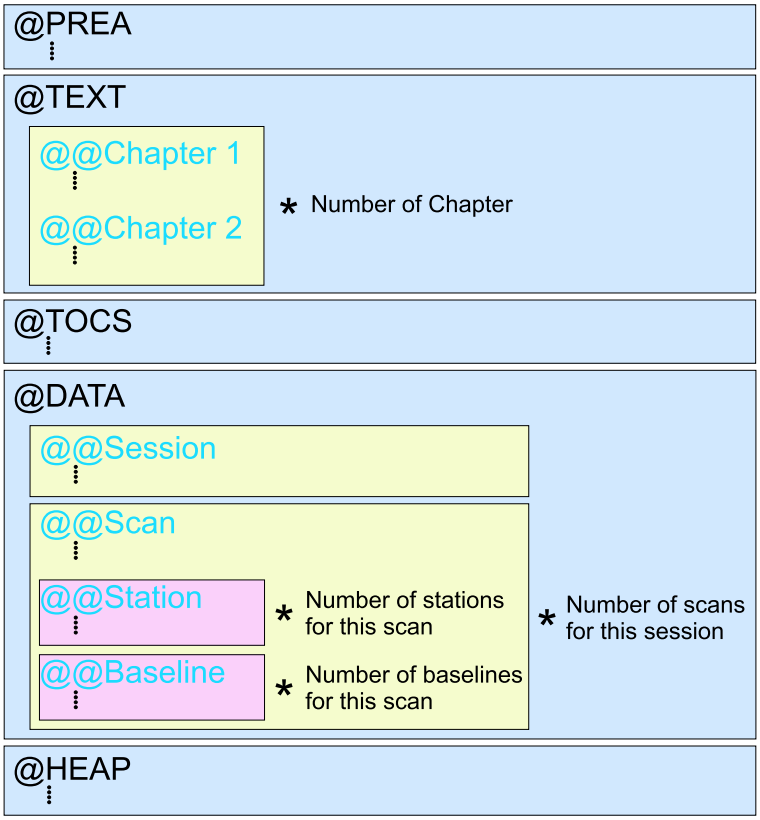

Figure 1. Schematic PIVEX structure

5.2. File Naming Convention

In order to avoid duplicate names and to recognize easily an experiment we proposed the following naming convention:

YYY_MM_DD_CODIVS_i.pvx

where YYY_MM_DD is the date, _CODIVS is the IVS code for the experiment (available in master schedule file on IVS web site) and _i is the file version (in case of re-fringing for example).
The resulting file names for the three tested experiments are:

2001_12_31_I01365_1.pvx
2001_12_11_NA450_1.pvx
2001_05_09_RDV28_1.pvx

They are available on the current PIVEX web site (http://lareg.ensg.ign.fr/feissel/pivex.html) and on IVS web site in the near future. For the binary file the extension will be .bgv with the same naming convention.

6. Future

The PIVEX proposal was accepted at the 3rd IVS analysis workshop (February 2002) in Tsukuba. A part of the remaining work is to complete the code to read the PIVEX format and to load it in memory. The next step, which consists of recoding and testing the geo VLBI handler on multiple platforms, will be realizd by S. Bolotin, if he agrees.

In the same time, a full user documentation (interface and information contents) will be written and circulated for comments in IVS.

As soon as the PIVEX content is completely specified and the code to transform CALC database into PIVEX is fully tested, hopefully before the end of the year, next experiments will be available in this two formats (CALC database and PIVEX) on the IVS data centers.

After that step, the old experiments will also be transformed to PIVEX format on IVS Data Centers going back to the past.

In the future, any revision of the PIVEX format, extensible by table of contents, should be agreed upon by e.g. a PIVEX monitoring group under the responsibility of the IVS Analysis Coordinator.

References

[1] Petrov, L., In: Specifications of a proposed geo-VLBI data format, http://gemini.gsfc.nasa.gov/development/gvf/gvf.html, 1999.

[2] Abdelkader, K., A.-M. Gontier, In: the GLORIA VLBI tables, http://lareg.ensg.ign.fr/feissel/pivex/gloria_base.html, 1996.