

Raw data storage concept for Atlas Pxl production

Dirk Hoffmann -CPPM-, 19.03.04

1. All institutes store the data taken in their production site(s) on one or several internet sites, accessible via HTTP protocol. They are responsible for backup and long term availability of the data.
2. Starting from an arbitrary HTTP address (root, URL), the raw data of production tests and characterisation measurements are organised in a tree according to the scheme

PXLPROD_ROOT/ItemType/ItemPdbNumber/TestType/TestNumber/...

The following table defines how the different path fractions will be defined in practice.

| <i>variable (path fraction)</i> | <i>valid values</i> |
|---------------------------------|---|
| ItemType | identical to the PDB registry number (14 digits) |
| ItemPdbNumber | BareModule, FEchip, FlexLoaded, FlexMCC, ... |
| TestType | ASSY, BOW, bow, BOWX, bowx, BURN, RECV, ONSTAVE, fdacs, config, ... |
| TestNumber | 2 digits |

This list will be updated by C. Gemme (Feb '04). Space characters should be avoided in path fragments.

Assuming that there is such a PPRDB (pixel production raw data base) installed at Genova with

```
PXLPROD_ROOT = "http://atlas1.ge.infn.it/pixprod",
```

a valid directory in this PPRDB would be

```
http://atlas1.ge.infn.it/pixprod/BareModule/20210220881402/BARE/01/
```

Another PPRDB at Marseille could be located at

```
PXLPROD_ROOT = "http://marpix1.in2p3.fr/Pixel/modules/html/"
```

with valid directories like

```
http://marpix1.in2p3.fr/Pixel/modules/html/BareModule/Ge_Mod24/ArrivalC  
heck/
```

3. A given file named xyz.ext in any directory may have a counterpart xyz.ext.comment containing a short description in plain text (human readable) format (e. g. noise.root, noise.root.comment).

4. These sites must be in general open for read access by other users and in particular a scanning robot, which can list directories with their files. It is allowed to site owners to hide (no read access) certain directories before officially releasing them. Once data are released, the access can be of type "world read". The scan robot establishes a central PPRDB (PPRDB catalog or master), similar to the data structure built up in <http://atlas1.ge.infn.it/pixprod/>. [A sample scan robot program, which produces a complete tree starting from an HTTP directory address (root) is available under [/afs/desy.de/user/h/hoffmann/bin.sh/wwwlsR.](#)] The PPRDB catalog establishes

- differences (vanished, unavailable files, new entries, ...)
- links to the PDB (validity of names and part numbers)
- consistency (double entries accross PPRDB sites/roots)

and informs the PPRDB site administrator by email about anomalies.

5. There is no restriction on the number of different sites or PPRDB roots in the collaboration. It is recommended to use one unique site per institute. A new root can be registered on the PPRDB catalog site.

6. The PPRDB catalog server fills the PDB appropriately according to the information that it gathers from the registered sites.

7. Local directory structures in each site can be maintained on each site server, but must not appear inside the PXLPROD_ROOT tree structure. Each site is responsible for its contents and should only accept foreign data after careful verification (source, but also transfer integrity). Internal data organisation like symbolic links to same data are within the responsibility of the local site responsible, who can avoid duplication of redundant data in this way.