A Simple Small Software Solution for GSI Tabularized Datasets

Jing Guo
GMAO, NASA/GSFC, August 14, 2013

1. GSI Tabularized Datasets

- GSI uses file `satinfo.txt` to manage data about radiance observations, for every sensor-instrument-satellite and channels (*sis-channel*), in table form.
- In particular, flag `iuse` is used to configure how observations of a given *sis-channel* entry should be used by GSI.
- Similar approaches are also used for other classes of observations, including `ozinfo.txt`, `convinfo.txt`, etc., with different contents and similar formats.

!sensor/instr/sat	chan	iuse	error	error_cld	ermax	var_b	var_pg
amsua_n15	1	1	3.000	9.100	4.500	10.000	0.000
amsua_n15	2	1	2.000	13.500	4.500	10.000	0.000
amsua_n15	3	1	2.000	7.100	4.500	10.000	0.000
amsua_n15	4	1	0.600	1.300	2.500	10.000	0.000
amsua_n15	5	1	0.300	0.550	2.000	10.000	0.000
amsua_n15	6	1	0.230	0.230	2.000	10.000	0.000
amsua_n15	7	1	0.250	0.195	2.000	10.000	0.000
amsua_n15	8	1	0.275	0.232	2.000	10.000	0.000
amsua_n15	9	1	0.340	0.235	2.000	10.000	0.000
amsua_n15	10	1	0.400	0.237	2.000	10.000	0.000
amsua_n15	11	-1	0.600	0.270	2.500	10.000	0.000
amsua_n15	12	1	1.000	0.385	3.500	10.000	0.000
amsua_n15	13	1	1.500	0.520	4.500	10.000	0.000
amsua_n15	14	-1	2.000	1.400	4.500	10.000	0.000
amsua_n15	15	1	3.000	10.000	4.500	10.000	0.000
hirs3_n17	1	-1	2.000	0.000	4.500	10.000	0.000
hirs3_n17	2	-1	0.600	0.000	2.500	10.000	0.000
hirs3_n17	3	-1	0.530	0.000	2.500	10.000	0.000
hirs3_n17	4	-1	0.400	0.000	2.000	10.000	0.000
hirs3_n17	5	-1	0.360	0.000	2.000	10.000	0.000
hirs3_n17	6	-1	0.460	0.000	2.000	10.000	0.000
hirs3_n17	7	-1	0.570	0.000	2.500	10.000	0.000
hirs3_n17	8	-1	1.000	0.000	3.000	10.000	0.000
hirs3_n17	9	-1	1.100	0.000	3.500	10.000	0.000
hirs3_n17	10	-1	0.600	0.000	2.500	10.000	0.000
hirs3_n17	11	-1	1.200	0.000	3.500	10.000	0.000
hirs3_n17	12	-1	1.600	0.000	4.500	10.000	0.000

2. The original problem

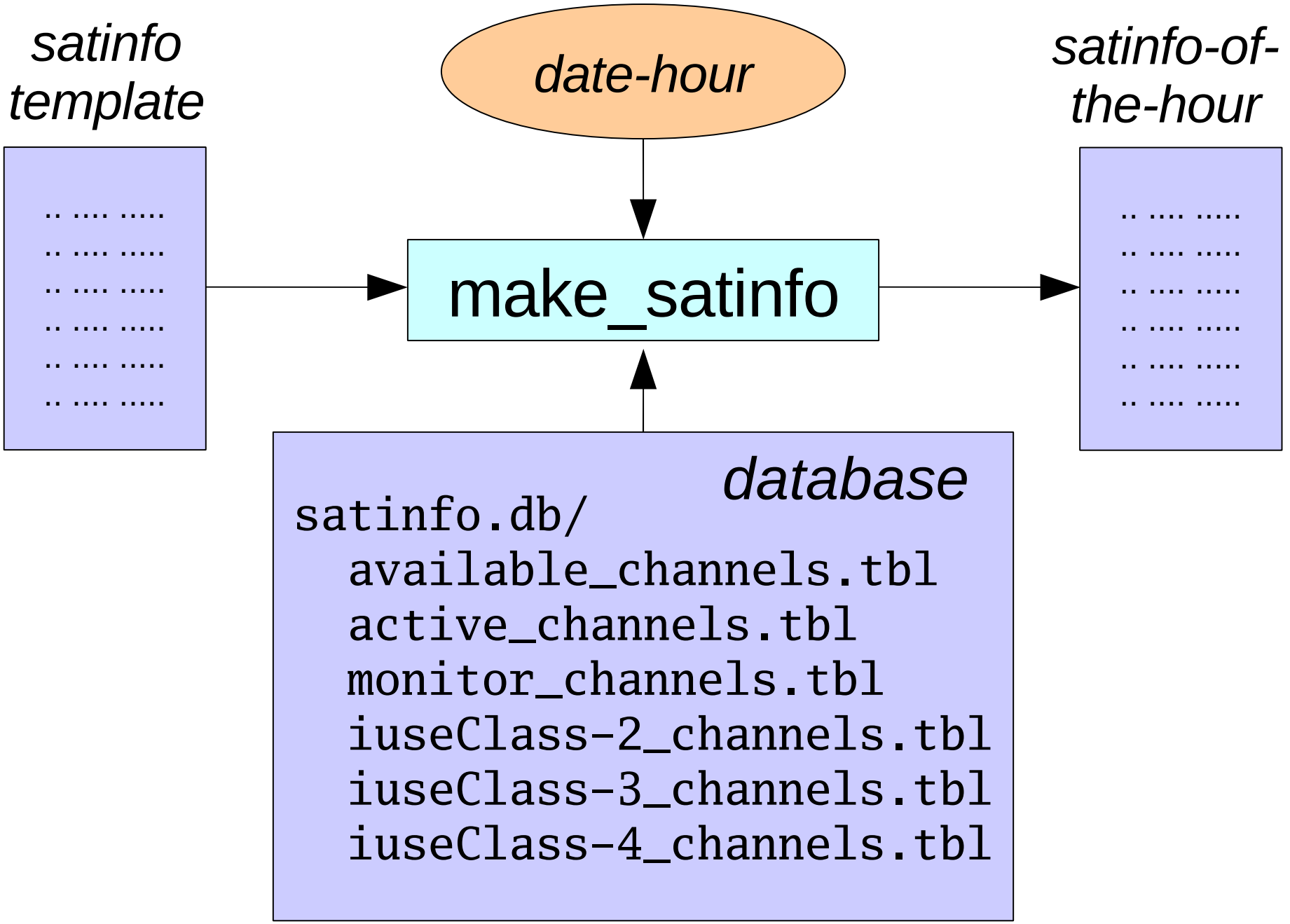
- Values of `iuse` are functions of time. A single static *info*-file sometime can not do it, in particular for reanalysis applications.
- Corporate Knowledge about `iuse` vs. *time* is scattered with too many details in different people, and fading.
- Customization vs. data sharing for different users and projects, requires a centralized, “officially” maintained, user-configurable, *satinfo* repository.

3. The GEOS DAS solution

- A separate *satinfo*-database is used in GEOS DAS, to maintain the information about `iuse` with respect to the time change.
- Emily Liu (and Joanna Joiner?) created the original database for satellite radiance observations, including the initial development of the software to generate a "*satinfo-of-the-hour*" file on the fly.
- Her solution was later simplified, once we had a better understanding of the issue, the information content, as well as the algorithm.
- The original `iuse` flag of -1/0/1, has now evolved into multiple values for *`satinfo`*. The implementation had to be extended to address that.

4. Basics about *info-of-the-hour*

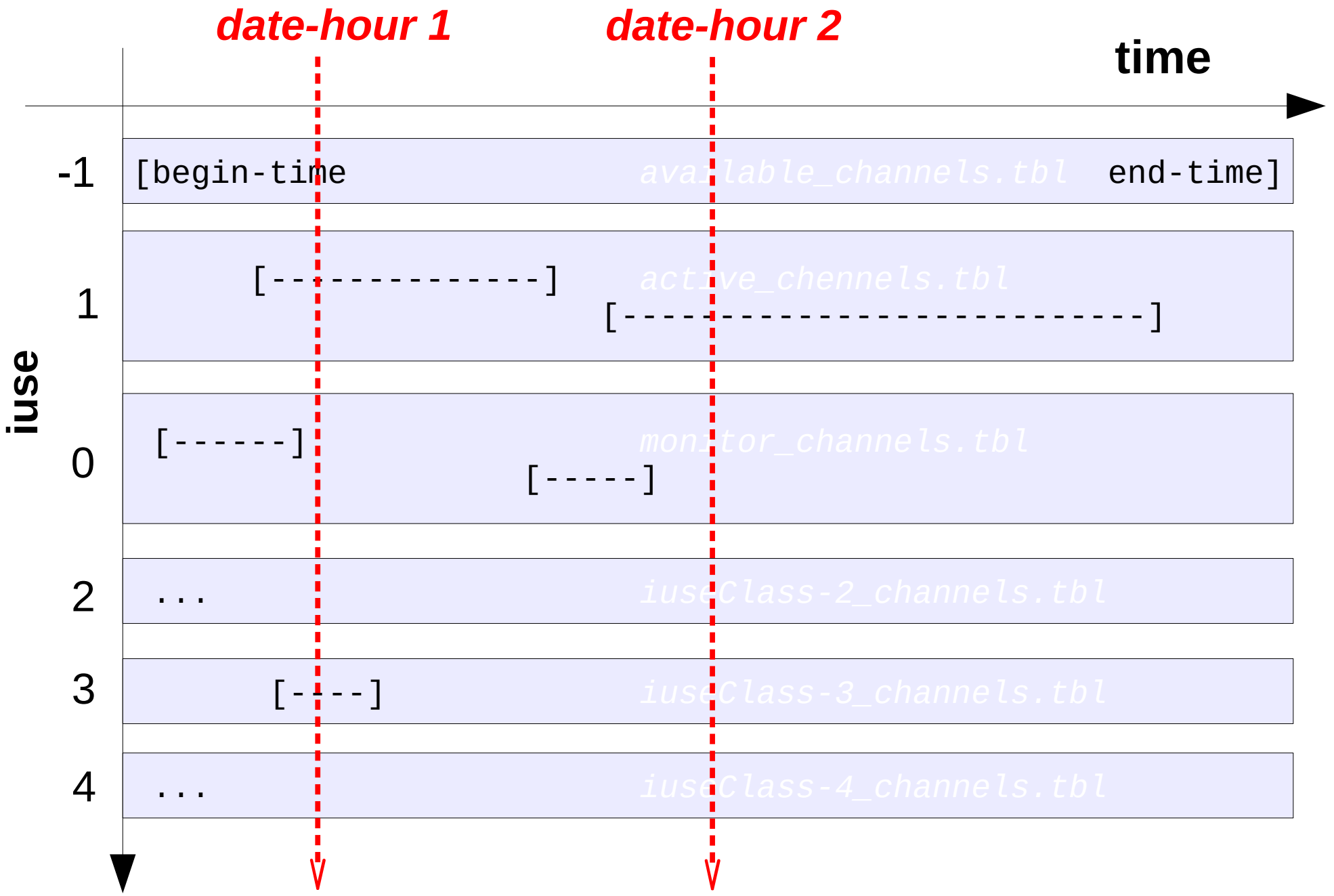
- *template*: an all-in-one *satinfo* file is used as a "template" for all *sis-channel* entries for the entire assimilation period.
- *database*: a directory with multiple *table*s for each different `iuse` values, is used to store information at different periods for every *sis-channel* entries.
- *info-of-the-hour*: given a *date-hour* (*yyyymmdd:hhmmss*), a off-line program will run "on-the-fly" to produce a *satinfo-of-the-hour* file for GSI, based on the information from the *template* and the *database*, and archived by individual experiments.


satinfo
template
.. ..
.. ..
.. ..
.. ..
.. ..
.. ..
date-hour
make_satinfo
satinfo-of-
the-hour
.. ..
.. ..
.. ..
.. ..
.. ..
.. ..
database
satinfo.db/
available_channels.tbl
active_channels.tbl
monitor_channels.tbl
iuseClass-2_channels.tbl
iuseClass-3_channels.tbl
iuseClass-4_channels.tbl

5. About the *database*

- The *database* contains multiple *tables*, is maintained through the GMAO configuration management process, as a part of our local GEOS DAS software system.
- All *table*s have the same format, with a *begin-end* time mask, in a *stem-leaf* (*sensor-channels*) form, as editable text files. A given *table* corresponds to a single `iuse` value,
- For the same given *date-hour* of the same *sis-channel*, overlapping table entries imply later entries always override earlier entries, in the same *table* file as well as in different *table* files.

```
# !REVISION HISTORY:
#
#  17Mar2009 Todling  Merge w/ earlier removal of 31 AIRS
#  18Mar2009 Liu      Remove channels 325 and 484 of AIRS
[...]

# NOAA-17 (nm)
n17  20020715 000000  20031027 240000  hirs3 14    2 3 4 5 6  ...
n17  20020715 000000  20031027 240000  amsua 10      4 5 6  ...
n17  20020715 000000  20031027 240000  amsub  5  1 2 3 4 5

n17  20031028 000000  20091222 120000  hirs3 14    2 3 4 5 6  ...
[...]

# SSM/I
f08  19870710 000000  19880207 240000  ssmi 7  1 2 3 4 5 6 7
f08  19880208 000000  19901231 240000  ssmi 6  1 2 3 4 5 7
f08  19910101 000000  19911204 240000  ssmi 5  1 2 3 4 5
f10  19901209 000000  19971113 240000  ssmi 7  1 2 3 4 5 6 7
f11  19911205 000000  19991218 240000  ssmi 7  1 2 3 4 5 6 7
f13  19950503 000000  20081231 240000  ssmi 7  1 2 3 4 5 6 7
f13  20090101 000000  21000531 240000  ssmi 0          #     ...
[...]
```


date-hour 1
date-hour 2
time
iuse
-1
[begin-time
available_channels.tbl
end-time]
1
[--------------]
active_chennels.tbl
[-------------------------]
0
[------]
monitor_channels.tbl
[-----]
2
...
iuseClass-2_channels.tbl
3
[----]
iuseClass-3_channels.tbl
4
...
iuseClass-4_channels.tbl

6. The implementation

- *Objective*: simple, small, but flexible with reasonable limitations.
- *Buffering a record before read*: to allow in-table remarks or comments and context-dependent read().
- *List-directed-input*: use "free-format" on Fortran "*internal-file*" (input buffer), for easy table editing.

```
character(len=256)::  buffer
lu=10
open(lu,file='satinfo.tmpl',...)
call getarec(lu,buffer,ier)
do while(ier==0)
  read(buffer,*) a, b, mv
  read(buffer,*) a, b, ii, v(1:mv)
  call getarec(lu,record,ier)
enddo
```

7. Current applications

- This *info-of-the-hour* process has been used in GEOS DAS for years, behind the scene quietly to most users.
- Similar processes have been applied to `ozinfo.txt` and `convinfo.txt`.
- It does not require any change from GSI, but only a step in the GEOS DAS GSI-setup process, self-explaining and requiring little support.
- Its size is growing, but still manageable.

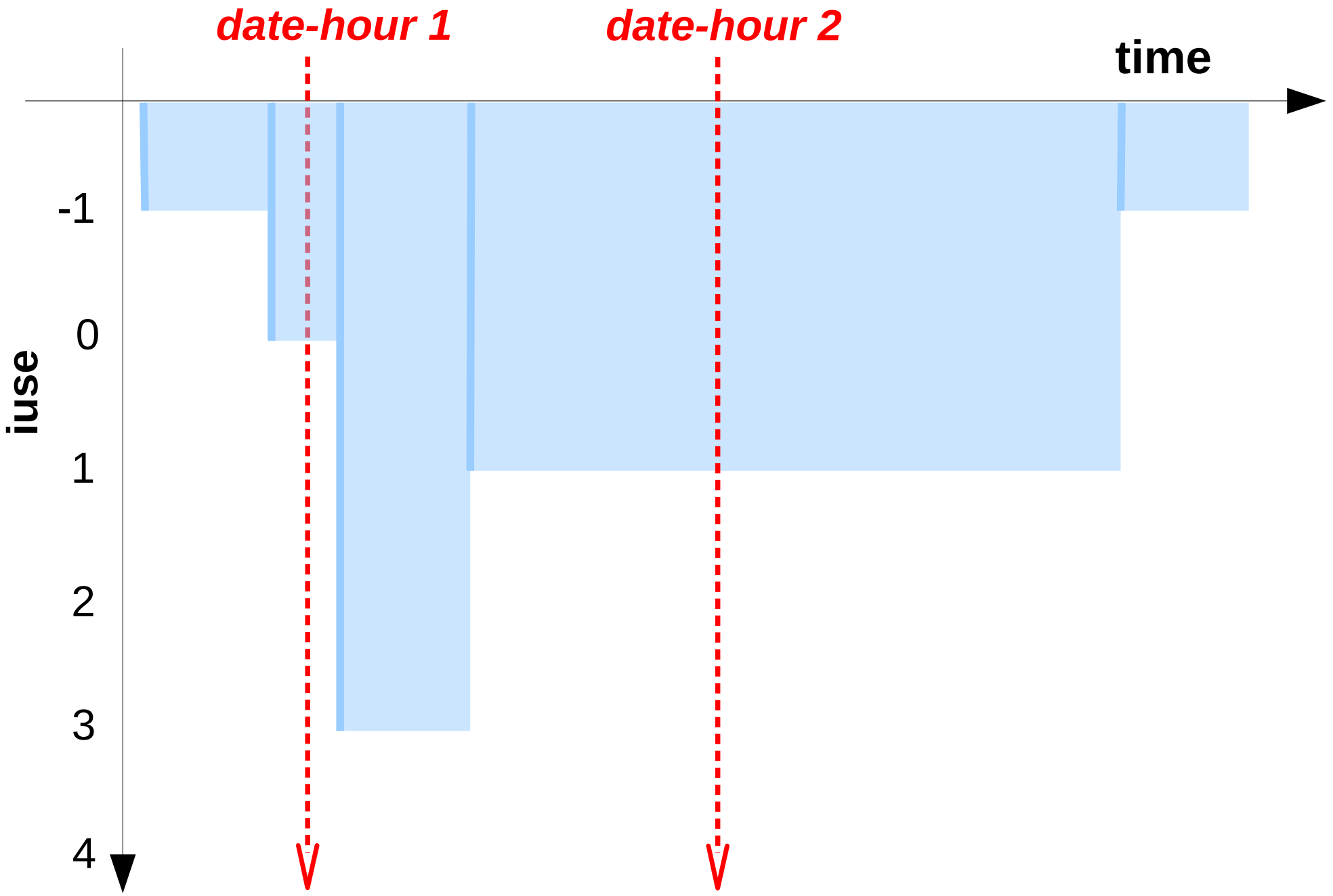
```
427 satinfo.db/available_channels.tbl
798 satinfo.db/active_channels.tbl
 23 satinfo.db/monitor_channels.tbl
 17 satinfo.db/iuseClass-2_channels.tbl
 17 satinfo.db/iuseClass-3_channels.tbl
 66 satinfo.db/iuseClass-4_channels.tbl
```

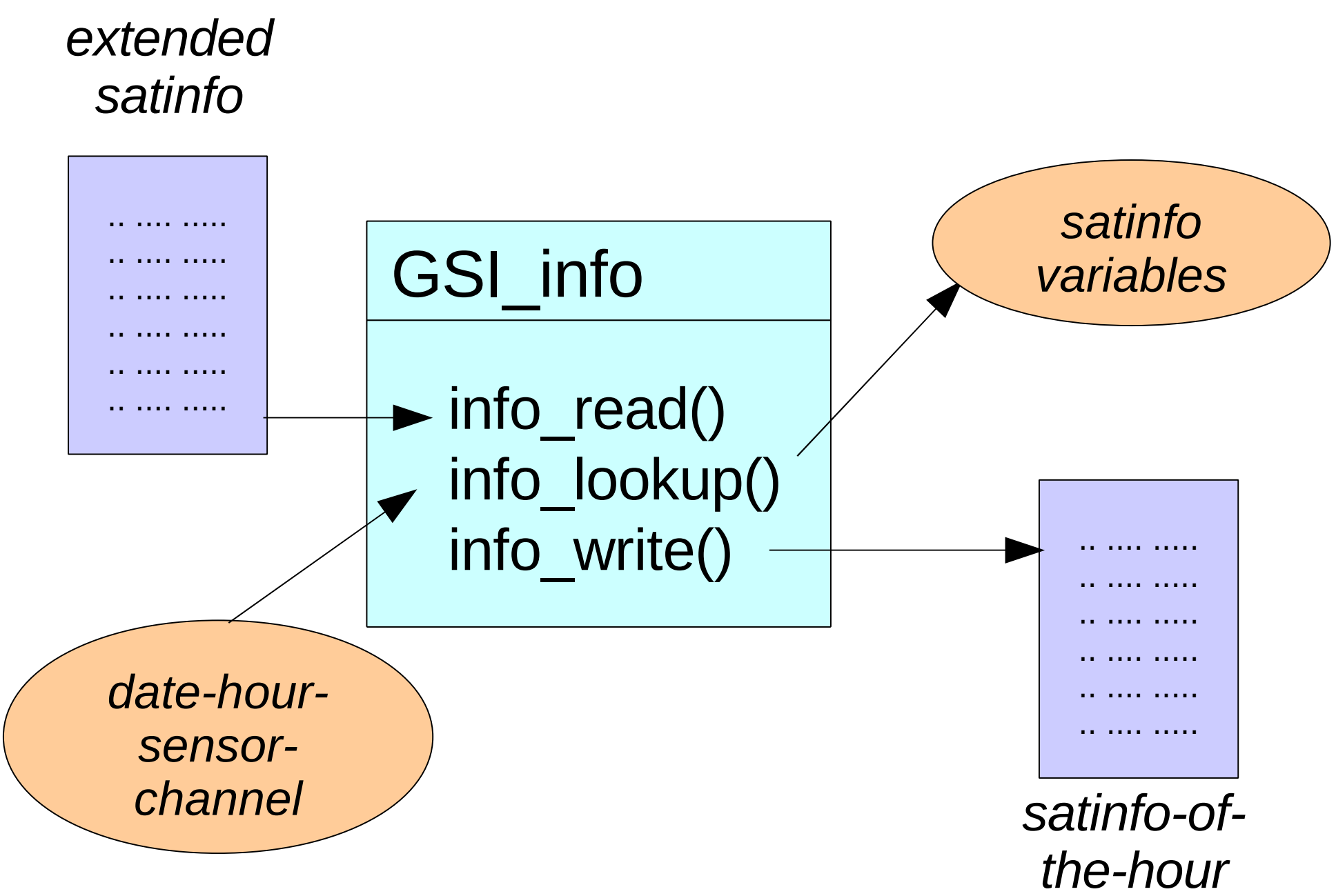
8. Future improvements?

- Time-manage information more than just `iuse` in the same *satinfo* file?
- Merge the database into an integrated (or extended) *satinfo* to minimize maintenance overhead?
  - Highly related information are managed at two different places.
  - What you see (in *template*) is NOT what you get (in an *info-of-the-hour* file).
- Then it means the following.
  - Should be backward-compatible.
  - Can be integrated into GSI for real-time processing.

Extended *satinfo* design

```
![yyyymmdd:hhmmss]   sis   chan iuse error  error_cld  ermax ...

                   amsua_n15  7   1   0.250    0.195    2.000 ...
                   amsua_n15  8   1   0.275    0.232    2.000 ...
                   amsua_n15  9   1   0.340    0.235    2.000 ...
                   amsua_n15 10   1   0.400    0.237    2.000 ...

# If something had happened to channel 11 on 06Z, Apr. 1, 2002
                   amsua_n15 11   1   0.600    0.270    2.500 ...
[20020401:060000] amsua_n15 11  -1   0.600    0.270    2.500 ...
[20020401:120000] amsua_n15 11   1   0.600    0.270    2.500 ...

                   amsua_n15 12   1   1.000    0.385    3.500 ...
                   amsua_n15 13   1   1.500    0.520    4.500 ...

# If you don't like to use channel 14 at some point
                   amsua_n15 14   1   2.000    1.400    4.500 ...
[20001030:000000] amsua_n15 14  -1   2.000    1.400    4.500 ...

                   amsua_n15 15   1   3.000   10.000    4.500 ...
```


date-hour 1
date-hour 2
time
iuse
-1
0
1
2
3
4


extended satinfo
.. ..
.. ..
.. ..
.. ..
.. ..
.. ..
GSI_info
info_read()
info_lookup()
info_write()
satinfo variables
date-hour-sensor-channel
.. ..
.. ..
.. ..
.. ..
.. ..
.. ..
satinfo-of-the-hour

End