

DB RESOURCES
DB POLICY AND POLITICS
DB SOFTWARE AND TECHNOLOGIES
EVENT DATA PRESERVATION STATUS
ADDITIONAL DATA
SOFTWARE PRESERVATION

OPERA DATABASE EXPORT AND DATA/SW PRESERVATION

DB RESOURCES

Physical machine received at CCIN2P3 with 2 TB disk

- Dedicated to OPERA for 2016/2017/2018 for Oracle DB dump
- Not part of Linux batch clusters

Account activated at CERN for data preservation

- Access to EOS (disk) and CASTOR (tape) granted
- Transfer through xrdcp

DB POLICY AND POLITICS

REMINDER: CCIN2P3 requested to free the DB by March 2018

Formal letter sent to extend DB usage to full 2018 – request accepted.
Final deadline: December 31st 2018

To-Do list:

- 1) Check status of already published data (completeness / integrity)
- 2) Check presence of more data in lab DBs (e.g. shower scan volumes)
- 3) Check presence of physics flags (charm-tau-nue)
- 4) Check all CS results are published
- 5) (maybe?) Dead Material scanback if available (otherwise we can't reproduce some plots) – only if there is enough time!

Support by me and Valeri is available to check/solve some technical problems!

DB SOFTWARE AND TECHNOLOGIES

OracleDumpManager (.NET/Mono exe)

Set of bash scripts

Full chain activated with a single command or event list

Input format:

```
BRICK EVENT ID_FEEDBACK
```

```
1123062 10240014359 1000010011679071
```

```
1050094 10310004085 1000010013001023
```

```
1134018 10209046135 1000010012731215
```

```
1046052 10270007687 1000010013000435
```

```
1138277 10233027637 1000010013003253
```

```
1120844 10317043625 1000010012735188
```

DB SOFTWARE AND TECHNOLOGIES

Each event is completely contained in a single directory

Create dump directory exp_evXXXXXXXXXX_bkYYYYYYYYYYY

Extract event-related electronic data

Extract brick-related data (ALL – will cause duplication for multi-events)

Extract feedback views for LAST feedback for that event

Convert to ASCII (in parallel with extraction of next event – generate directory with name ascii_exp_evXXXXXXXXXXXX_bkYYYYYYYYYYY)

Copy over Internet (after ASCII, in parallel with extraction of next event)

All actions are logged and extraction log is saved
Single corrupt rows or fields are documented and skipped

EVENT DATA PRESERVATION STATUS

Work to be done, in order of dataset size

Laboratory	Total DS	At CERN	CHECK1	CHECK2	To do	To do (%)	Notes
BERN	1262	5	312	108	1257	19.3	Checks only partially reliable (volume definition missing)
BOLOGNA-PADOVA	430	0	357	313	430	6.6	263 events selected to start (<i>ex officio</i>)
NAGOYA	3080	2887	2944	2887	193	3	Missing events closed in scanback and correction of physics flags (tau/charm/nue)
NAPOLI	679	495	515	161	184	2.8	
FRASCATI	64	0	42	16	64	1	Checks to be done, could start
BARI	317	279	298	287	38	0.6	Some events stuck because of missing CS data
ANKARA	35	0	22	15	35	0.5	Should start
DUBNA	28	0	26	13	28	0.4	Checks only partially reliable (volume definition missing), 14 manually certified
LEBEDEV	26	0	6	0	26	0.4	Should start
SALERNO	500	481	498	492	19	0.3	Some events stuck because of missing CS data
SINP-MSU	18	0	15	5	18	0.3	Should start
ROMA	70	60	63	60	10	0.2	Several checks to be done, including for events already at CERN
Total	6509	4207			2302	35.4	

C. B

EVENT DATA PRESERVATION STATUS

CS data missing because of failed replication from OPITA to OPFRA

Link to LNGS established, need to recheck all publications (I should take a few days with some support by Nicola and Natalia – CS data are small in size)

Global figures

Disk occupancy at CERN for 4207 events (about 63% of total)

9.6 TB (binary format)

13 TB (ASCII format)

Average export speed is 28 events/day

82 days needed if we had all data available in the DB

BUT WE DON'T HAVE ALL DATA IN THE DB AND HUMAN RESOURCES ARE CRITICAL TO ENSURE DATA QUALITY FROM LABORATORIES!

EVENT DATA PRESERVATION STATUS

The following slides contain queries that have been generated on purpose to run on the **opera** account of the Central DB

Local Database administrators are **encouraged** to use these queries to recover details on a brick-by-brick and event-by-event basis

The materialized views named

DBEX_.....

are recomputed **daily** and available for you to check the status of your laboratory, accessible from your local **operapub** account, e.g.:

```
select * from opera.dbex_invalid_datasets@opfra where laboratory = 'SALERNO'
```


EVENT DATA PRESERVATION STATUS

CS publication statistics

DBEX_CS_PUBLICATION_STATUS

RUNYEAR	LAB	ASSIGNED	FLAGSRECEIVED	DATARECEIVED
2007	LNGS	1	1	0
2007	Nagoya	4	3	0
2008	LNGS	1406	1359	805
2008	Nagoya	1330	1301	654
2009	LNGS	2794	2735	1563
2009	Nagoya	2962	2869	1436
2010	LNGS	3069	2894	1553
2010	Nagoya	2987	2887	1544
2011	LNGS	3561	3336	1297
2011	Nagoya	3303	3164	1603
2012	LNGS	2513	2373	986
2012	Nagoya	2447	2345	954

EVENT DATA PRESERVATION STATUS

CS results published, missing candidates

DBEX_CS_RESULTS_MISSCANDS

```
select id_cs_eventbrick, result_status, sum(nvl2(id, 1, 0)) as ncands,
tb_brick_assign.cs_assign from
(select id_cs_eventbrick, result_status, id from tb_cs_results
left join tb_cs_candidates on id_eventbrick = id_cs_eventbrick)
inner join tb_brick_assign on tb_brick_assign.id_eventbrick = id_cs_eventbrick
group by id_cs_eventbrick, result_status, cs_assign)
```

RESULT_STATUS	CS_ASSIGN	CSDOUBLETS	CSWITHCANDS
B2B_FASTUNPACK	JP	8	0
B2B_FASTUNPACK	UE	8	0
B2B_NOSCAN	UE	2	0
BACK_TO_DETECTOR	JP	2558	28
BACK_TO_DETECTOR	UE	3721	112
BACK_TO_DETECTOR_NO_CS	JP	2525	1
BACK_TO_DETECTOR_NO_CS	UE	2445	25
BLACK_CS_DEVELOP	JP	159	10
BLACK_CS_DEVELOP	UE	351	0
CS_CAND_OK_DEVELOP	JP	6632	5951
CS_CAND_OK_DEVELOP	UE	5745	5699
CS_CAND_OK_FAST_UNPACK	JP	315	67
CS_CAND_OK_FAST_UNPACK	UE	43	40
NO_COSMIC_RAYS_DEVELOP	JP	414	141
NO_COSMIC_RAYS_DEVELOP	UE	454	414
WRONG_CS_HANDLING_DEVELOP	UE	33	11

JP: many CS without candidates

UE: in some cases another CS was flagged with results to trigger re-extraction

EVENT DATA PRESERVATION STATUS

Brick data consistency checks

Question: do we have scanning/feedback data for events that are at least connected?

Main query

```
select laboratory, event, brick, result_status, idcs, located, deadmaterial, passing, ds_any_done,
decode(sum(nvl2(sb1.id, 1, 0)),0,0,1) as issb, decode(sum(nvl2(vo2.id, 1, 0)),0,0,1) as isvol,
decode(sum(nvl2(re3.id, 1, 0)),0,0,1) as isrec from
(select laboratory, event, brick, result_status, idcs, located, deadmaterial, passing, ds_any_done, id as idop from
(select laboratory, event, brick, result_status, idcs, located, deadmaterial, passing, ds_any_done from
xv_event_location_detail where started > 0 and connected > 0)
 left join tb_proc_operations on id_parent_operation is null and id_eventbrick = brick
)
left join tb_scanback_paths sb1 on sb1.id_processoperation = idop
left join tb_volumes vo2 on vo2.id_processoperation = idop
left join tb_reconstructions re3 on re3.id_processoperation = idop
group by laboratory, event, brick, result_status, idcs, located, deadmaterial, passing, ds_any_done
```

EVENT DATA PRESERVATION STATUS

Decaysearched events, no volume data available

DBEX_DECAYSEARCH_NOTSDATA (located > 0 and ds_any_done > 0 and isvol = 0)

LABORATORY	BRICKS
ANKARA	13
BARI	19
BERN	889
BOLOGNA-PADOVA	68
DUBNA	0
FRASCATI	20
LEBEDEV	20
NAGOYA	127
NAPOLI	163
ROMA	7
SALERNO	0
SINP-MSU	3

WARNING: FEDRA and Japan system don't always record a volume

Publication not done or incomplete transfer

EVENT DATA PRESERVATION STATUS

Decaysearched events, no feedback

DBEX_DECAYSEARCH_NOFEEDBACK (located > 0 and ds_any_done > 0 and isrec = 0)

LABORATORY	BRICKS
ANKARA	5
BARI	4
BERN	428
BOLOGNA-PADOVA	7
DUBNA	2
FRASCATI	9
LEBEDEV	4
NAGOYA	121
NAPOLI	4
SALERNO	2

Publication not done or incomplete transfer

EVENT DATA PRESERVATION STATUS

Decaysearched events, no data

DBEX_DECAYSEARCH_NODATA (`ds_any_done > 0 and issb = 0 and isvol = 0 and isrec = 0`)

LABORATORY	BRICKS
ANKARA	5
BARI	8
BERN	361
BOLOGNA-PADOVA	2
DUBNA	0
FRASCATI	6
LEBEDEV	4
NAGOYA	112
NAPOLI	3
SALERNO	0

Publication not done or incomplete transfer

EVENT DATA PRESERVATION STATUS

Passing-through/edgeout bricks, no scanning data

`DBEX_PASSING_NOSCANDATA` (`passing > 0` and `issb = 0` and `isvol = 0`)

LABORATORY	BRICKS
ANKARA	11
BARI	0
BOLOGNA-PADOVA	8
DUBNA	6
FRASCATI	3
LEBEDEV	7
NAGOYA	174
NAPOLI	10
ROMA	5
SINP-MSU	6

Publication not done or incomplete transfer

EVENT DATA PRESERVATION STATUS

Passing-through/edgeout bricks, no feedback data

DBEX_PASSING_NOFEEDBACK (passing > 0 and isrec = 0)

LABORATORY	BRICKS
BOLOGNA-PADOVA	4
LEBEDEV	2
NAGOYA	174
NAPOLI	5
ROMA	4
SINP-MSU	9

Publication not done or incomplete transfer

EVENT DATA PRESERVATION STATUS

Apparently export-ready events

DBEX_EXPORT_READY_CHECK1 (`ds_any_done > 0 and isvol > 0 and isrec > 0`)

LABORATORY	BRICKS
ANKARA	22
BARI	298
BERN	312
BOLOGNA-PADOVA	357
DUBNA	26
FRASCATI	42
LEBEDEV	6
NAGOYA	2944
NAPOLI	515
ROMA	63
SALERNO	498
SINP-MSU	15

WARNING: FEDRA and Japan system don't always record a volume

This check alone is not sufficient, inspection is needed

EVENT DATA PRESERVATION STATUS

Probably export-ready events

DBEX_EXPORT_READY_CHECK2

(DS primary vertex location is contained in a TS volume at least 5×5 mm²)

LABORATORY	DECAYSEARCHED	TSAVAILABLE
ANKARA	35	15
BARI	317	287
BERN	1262	108
BOLOGNA-PADOVA	428	313
DUBNA	28	13
FRASCATI	64	16
LEBEDEV	26	0
NAGOYA	3080	2887
NAPOLI	677	161
ROMA	69	60
SALERNO	500	492
SINP-MSU	18	5

```
select laboratory, ev, brick, result_status, idcs, 1 as ds_any_done, decode(sum(case when maxz is null then 0 when maxz > posz and minz < posz then 1 else 0 end),0,0,1) as hasvol from
(select laboratory, ev, brick, result_status, idcs, lastidr, id_vertex, posx, posy, posz, id_volume, max(Z) as maxz, min(Z) as minz from
(select laboratory, ev, brick, result_status, idcs, idr as lastidr from
(select laboratory, event+0 ev, brick, result_status, idcs, located, deadmaterial, passing, ds_any_done from xv_event_location_detail where ds_any_done > 0)
left join (select id_eventbrick, event+0 as event, max(id_reconstruction) as idr from vw_feedback_reconstructions group by id_eventbrick, event) on id_eventbrick = brick and event+0 = ev)
left join vw_feedback_vertices vx on vx.id_eventbrick = brick and id_reconstruction = lastidr and isprimary = 'Y'
left join tb_volume_slices sl on sl.id_eventbrick = brick and minx < posx and maxx > posx and miny < posy and maxy > posy and (maxx - minx) * (maxy - miny) > 25e6
left join tb_plates pl on pl.id_eventbrick = brick and pl.id = id_plate
group by laboratory, ev, brick, result_status, idcs, lastidr, id_vertex, posx, posy, posz, id_volume
) group by laboratory, ev, brick, result_status, idcs
```

Also this check is not sufficient, inspection is needed

EVENT DATA PRESERVATION STATUS

Incomplete data transfers

DBEX_INVALID_DATASETS

LABORATORY	DATASETS
ANKARA	27
BARI	6
BERN	11
BOLOGNA-PADOVA	65
LYON	2
LNGS	13
ROMA	18
SINP-MSU	13

Interrupted transfer - data need to be deleted and published again

ADDITIONAL DATA

Additional data to be exported:

- 1) Last status of reporting views
- 2) Electronic data
- 3) BAM data
- 4) BMS data
- 5) Development data
- 6) MC data (and analysis SW if available)

All such data should be quick to export

SOFTWARE PRESERVATION

Software:

- 1) DB Schema
- 2) OpRelease
- 3) OpRec
- 4) OpEmuRec
- 5) BrickFinder
- 6) All analysis software used to create papers we published
- 7) Analysis scripts and programs (in case we want to reproduce our own result)

All this should be available as source

Also installed VM in OVA format, when applicable