

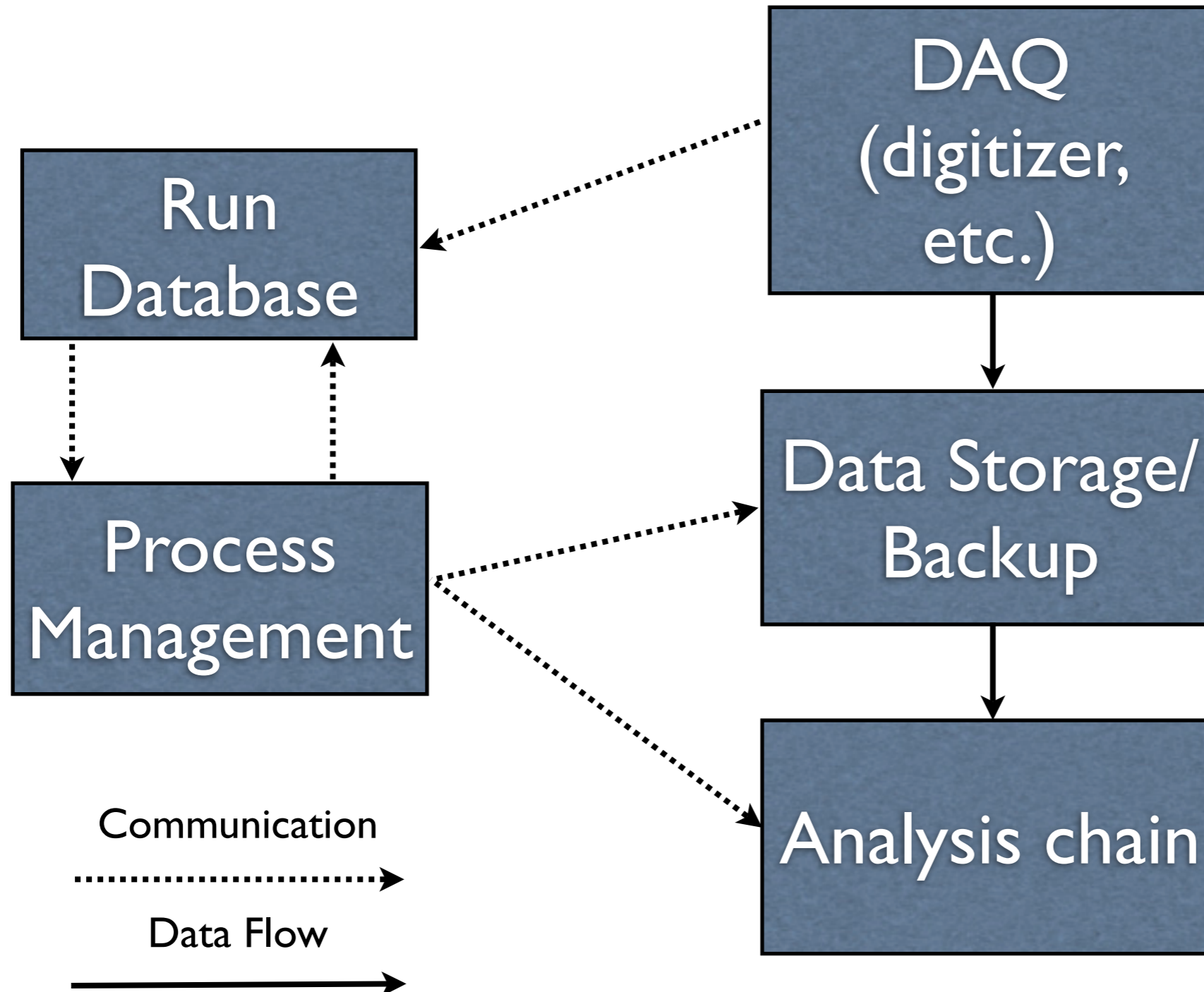
Run DB for MAJORANA R&D

Michael Marino
MaGe Meeting, München
19 Jan 2010

Introduction: the problem

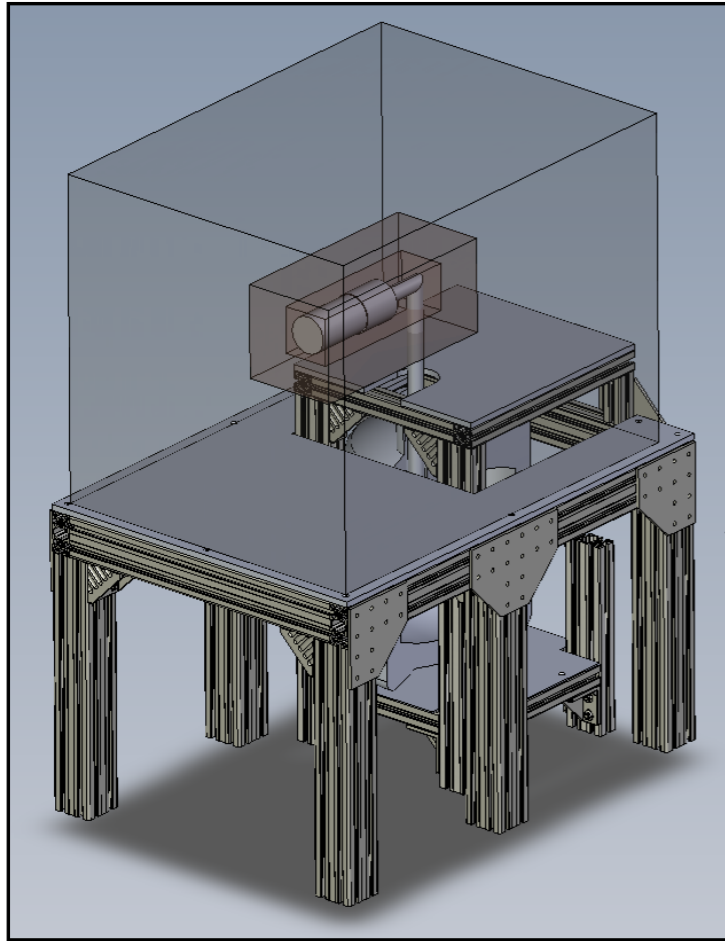
- Compilation of data from DAQ (e.g. raw data, detector health, environment, etc.) must be aggregated into a DB.
- larger # of files leads quickly to management nightmare
- DB also used for data process management (storage, analysis chain, etc.)

'Run' Database



Introduction: investigate solutions

- Determining DB needs (this session).
- Majorana R&D detectors taking physics data provide opportunity for testing actual implementation.
- e.g. 2 BeGe, P-PC detectors underground (consecutively) in Soudan, MN.

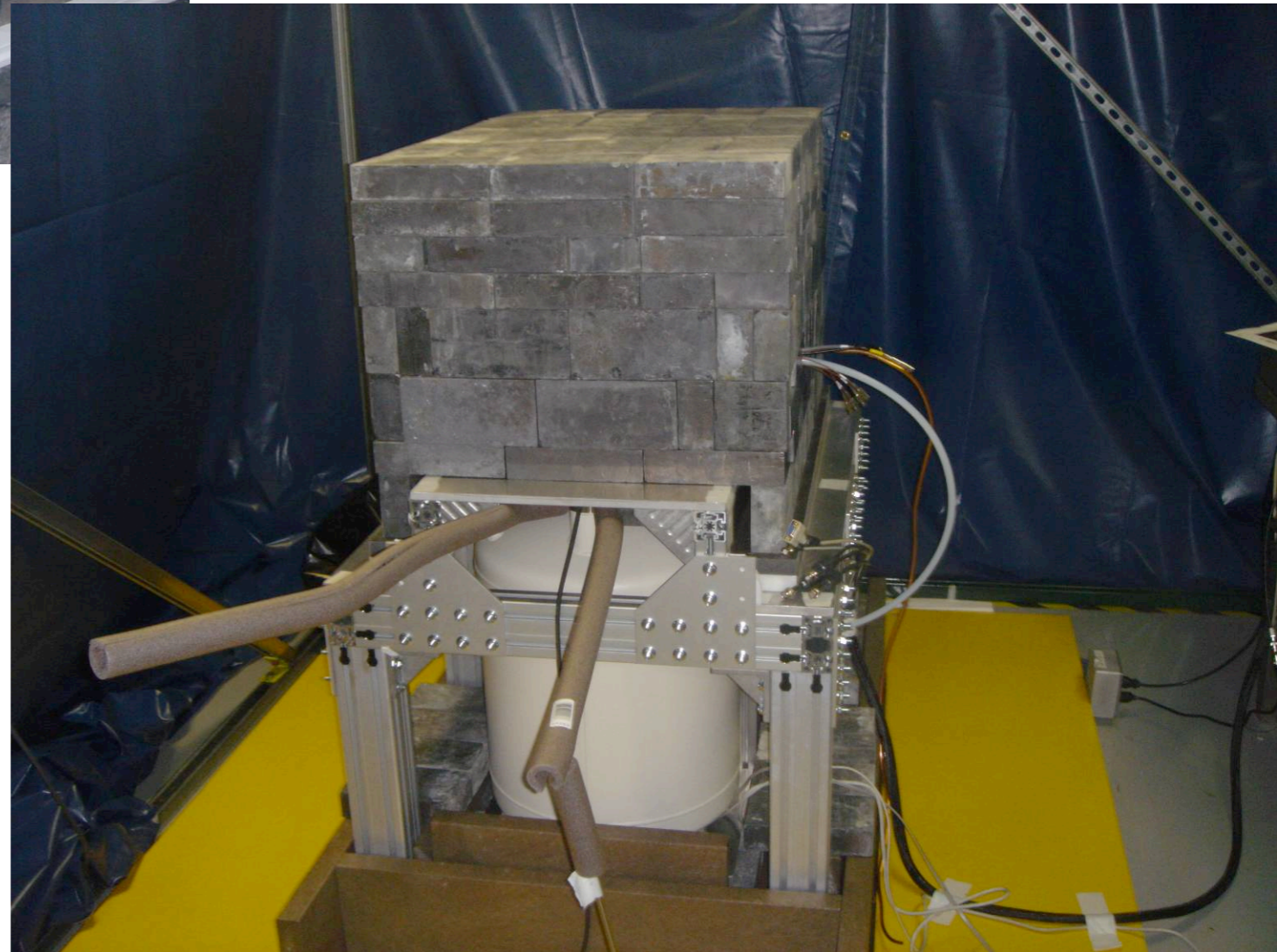
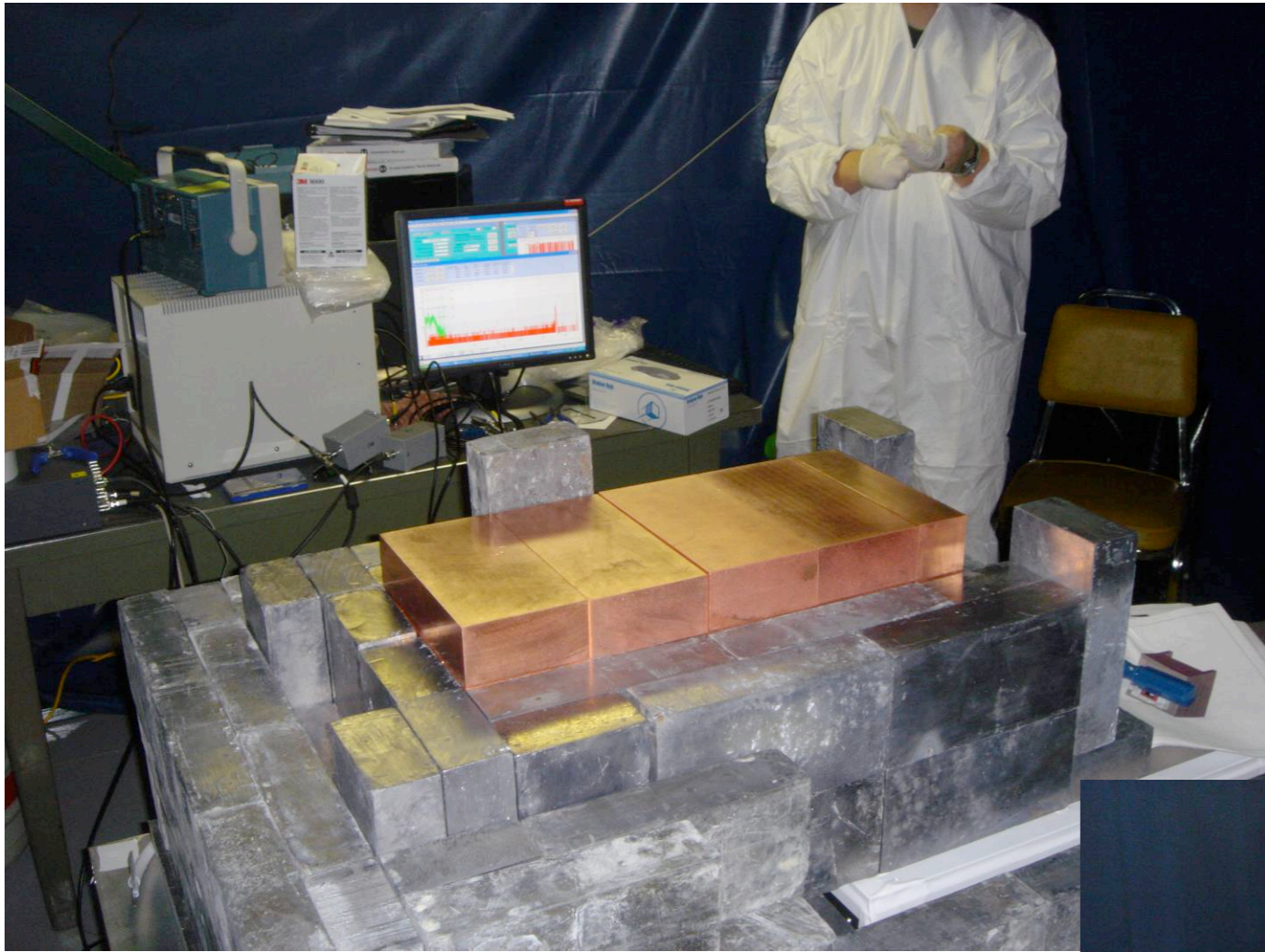


grey vertical wall

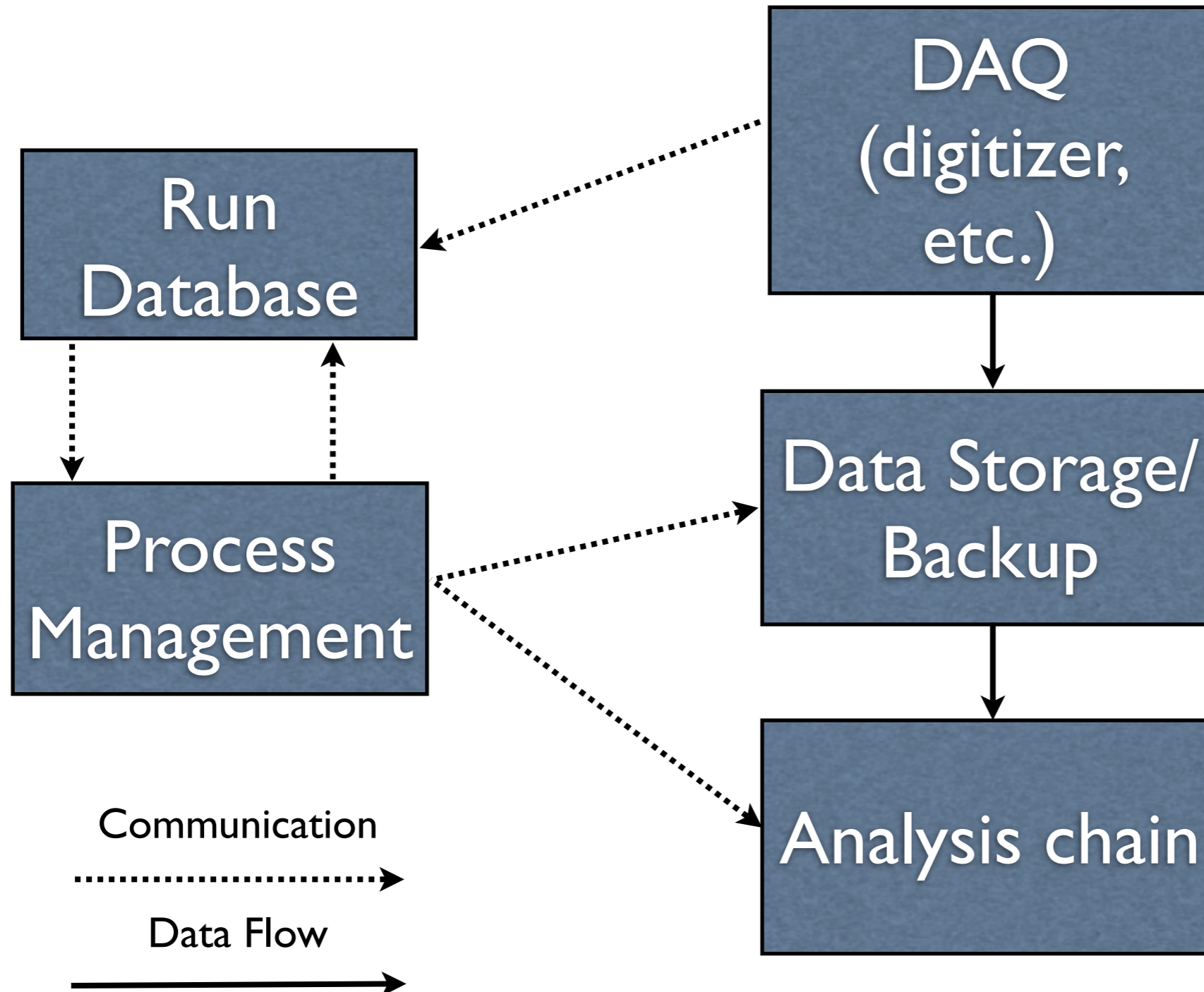


Slide: J. Orrell

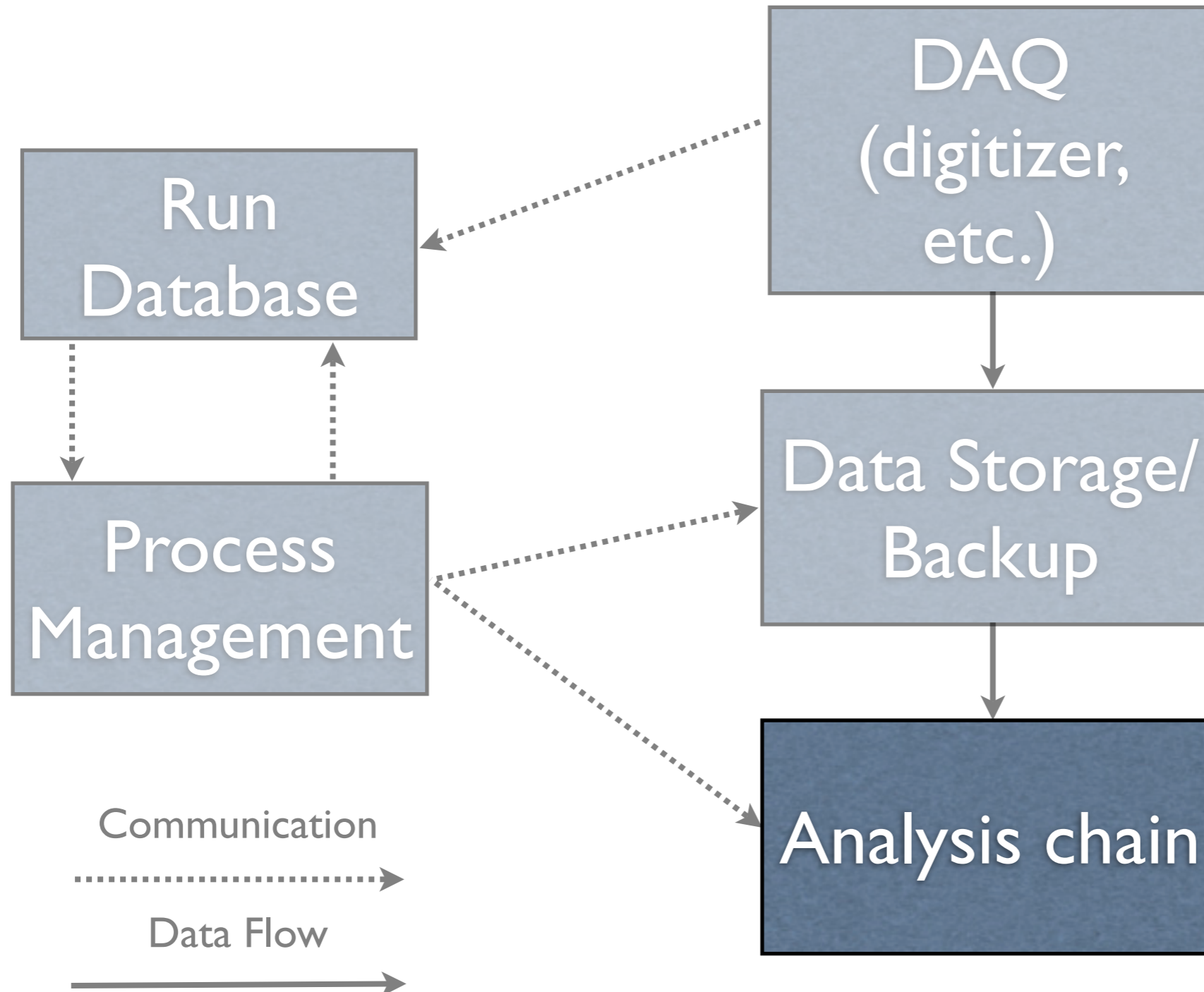
Some more pictures



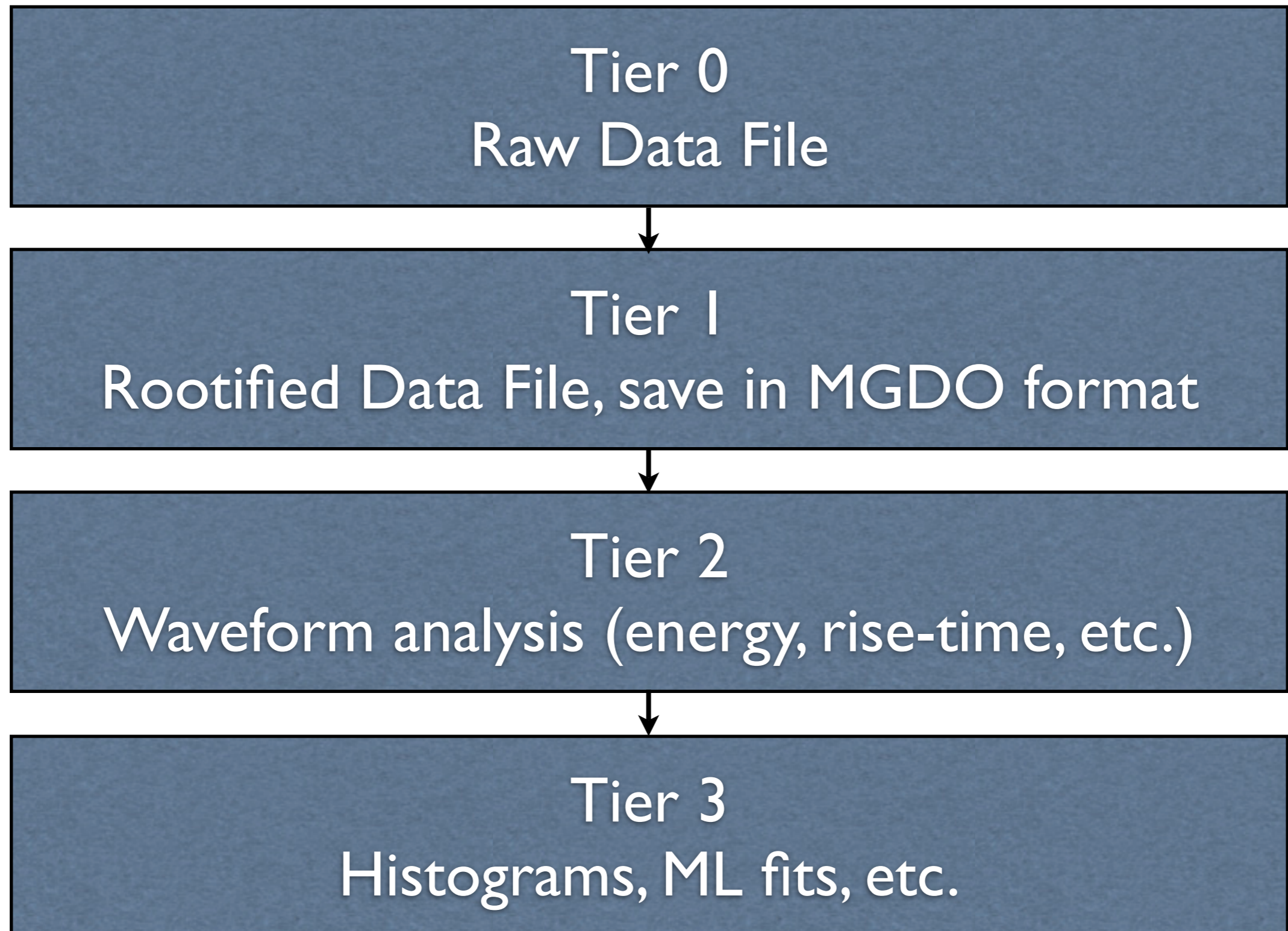
'Run' Database



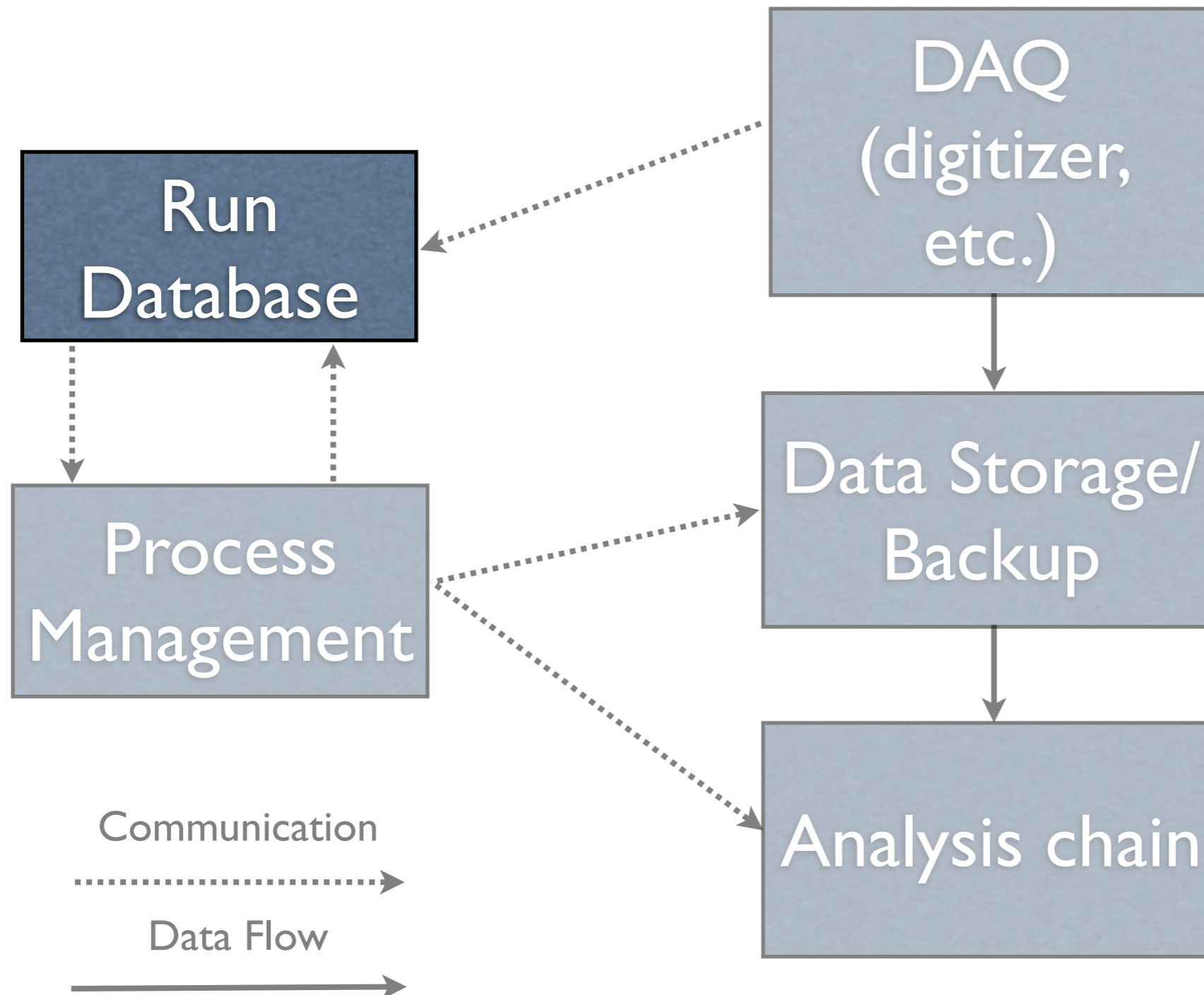
'Run' Database



Analysis chain



'Run' Database



CouchDB

- Document-based DB, schema-less, meaning **no tables, JSON**
- HTML server, talk to it in any language you like (python, perl, Ruby, C, web browser, etc.)
- Automatic document versioning, allowing data update after initial insertion.
- ‘Views’ replace queries, automatic caching **significantly** speeds up database response
- Replication allows simple DB backup, mirroring

CouchDB

- a glimpse inside.
- JSON - raw data examples
- View examples

```
{ "_id": "20090828173627",

"_rev": "37-0a77a90847d1d218c296ae374cbb1627",
"settings_data_file": {
  "last_mod_time": "2009-08-28T15:36:27Z",
  "lfn": "tier0/090828173627_settings",
  "md5hash": "43278e4704a1e4c17784b63d22c2df2d",
  "pfn": "/mnt/raid/data/Soudan/Data/BeGe/tier0/090828173627_settings"
},
"raw_data_file_tier_0": {
  "last_mod_time": "2009-08-28T15:37:49Z",
  "lfn": "tier0/090828173627",
  "md5hash": "79d4c0fb2f40c1d9736a6292f53ff3b8",
  "pfn": "/mnt/raid/data/Soudan/Data/BeGe/tier0/090828173627"
},
"number_of_entries_in_tier1_root_tree": 2400,
  "lifetime": {
    "run_milliseconds": 80234,
    "run_milliseconds_error": 0
  },
"output_data_file_tier_2": {
  "last_mod_time": "2009-09-14T11:50:26Z",
  "lfn": "tier2/090828173627_output.root",
  "md5hash": "43f17f1ec935235a2ae163a696e2228c",
  "pfn": "/mnt/raid/data/Soudan/Data/BeGe/tier2/090828173627_output.root"
},
"local_time_of_start_of_run": "2009-08-28T17:36:27Z",
"root_data_file_tier_1": {
  "last_mod_time": "2009-09-14T11:50:21Z",
  "lfn": "tier1/090828173627_rootified.root",
  "md5hash": "b4c22beac149f5bfba2c3cc6493d846c",
  "pfn": "/mnt/raid/data/Soudan/Data/BeGe/tier1/090828173627_rootified.root"
},
"run_settings": [
  "1000",
  "2500000",
  "20"
]
}
```

Views (queries)

Simple Example:

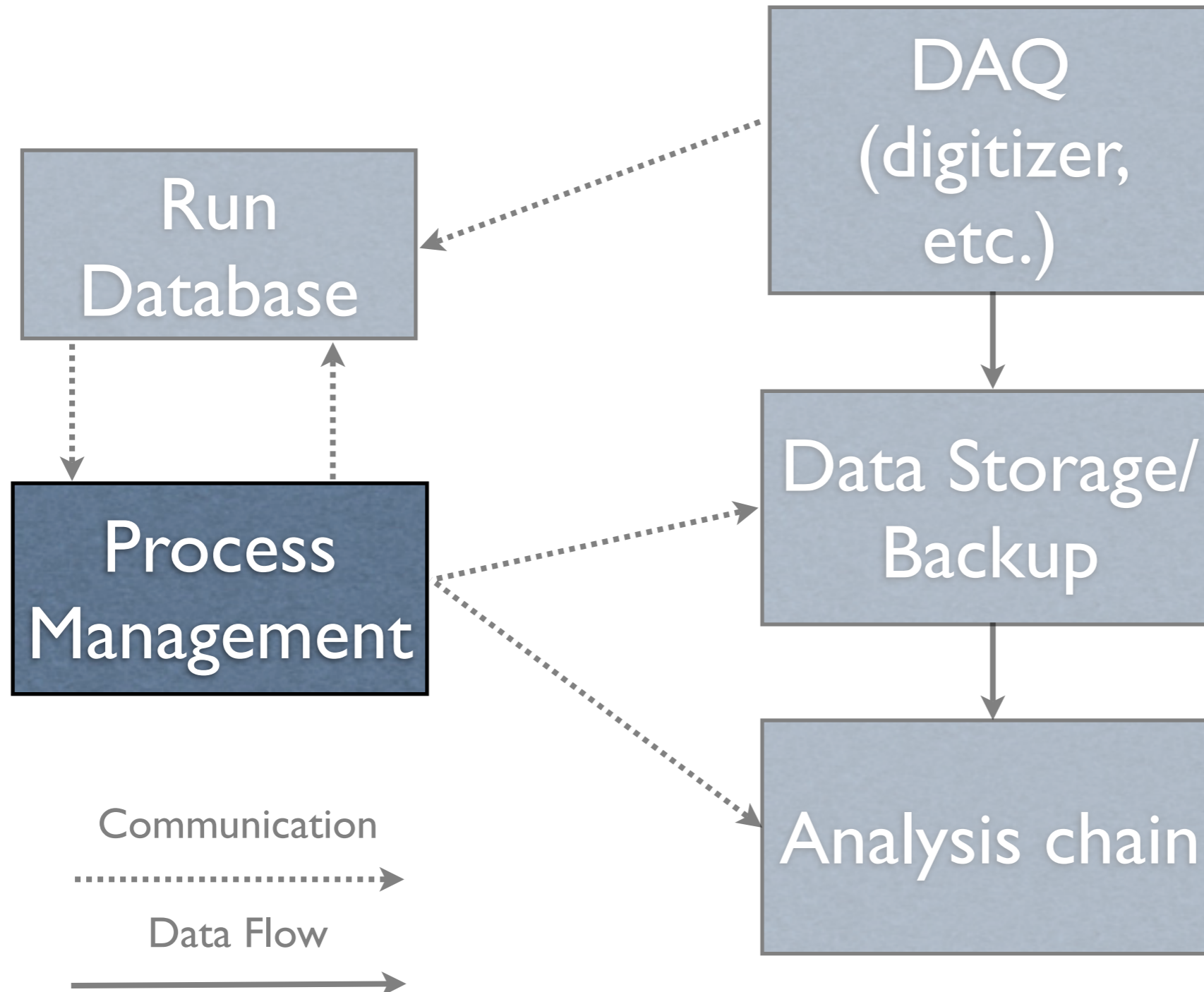
```
function(doc) {  
  
  var my_string = doc._id  
  if (my_string.search("LN2") != -1) return;  
  if (isNaN(parseInt(my_string))) return;  
  emit(doc._id, null);  
}
```

Complex Example:

```
function(doc) {  
  
  var my_string = doc._id  
  if (my_string.search("LN2") != -1) return;  
  if (isNaN(parseInt(my_string))) return;  
  emit(doc._id, [[doc.raw_data_file_tier_0.pfn, doc.raw_data_file_tier_0.last_mod_time],  
                [doc.root_data_file_tier_1.pfn, doc.root_data_file_tier_1.last_mod_time],  
                [doc.output_data_file_tier_2.pfn, doc.output_data_file_tier_2.last_mod_time]]);  
}
```

Field	Value
_id	"20091204180005"
_rev	"17-d543d7bc74e480f94286666e1f9153c9"
<input checked="" type="checkbox"/> all_cuts	
<input checked="" type="checkbox"/> livetime	run_milliseconds 6742687 run_milliseconds_error 0
<input checked="" type="checkbox"/> local_time_of_start_of_run	"2009-12-04T18:00:05Z"
<input checked="" type="checkbox"/> modification_time	null
<input checked="" type="checkbox"/> number_of_entries_in_tier1_root_tree	20
<input checked="" type="checkbox"/> output_data_file_tier_2	last_mod_time "2009-12-22T18:47:26Z" lfn "tier2/091204180005_output.root" md5hash "d42ac19830b530da5e5a308a1fef794e" pfn "/mnt/raid/data/Soudan/Data/BeGe/tier2/091204180005_output.root"
<input checked="" type="checkbox"/> pulser_off_set	0
<input checked="" type="checkbox"/> pulser_on_set	
<input checked="" type="checkbox"/> pulser_run_characteristics	chan_2 chan_1 chan_0
<input checked="" type="checkbox"/> quality_assurance	qa_check_process_has_been_run null qa_accept_run null
<input checked="" type="checkbox"/> raw_data_file_tier_0	last_mod_time "2009-12-04T17:53:39Z" lfn "tier0/091204180005" md5hash "bb37a06190a19b6fecf9cb07bae571e5" pfn "/mnt/raid/data/Soudan/Data/BeGe/tier0/091204180005"
<input checked="" type="checkbox"/> root_data_file_tier_1	last_mod_time "2009-12-22T18:46:08Z" lfn "tier1/091204180005_rootified.root" md5hash "c68a065207b64e8098778f1d3f051819" pfn "/mnt/raid/data/Soudan/Data/BeGe/tier1/091204180005_rootified.root"
<input checked="" type="checkbox"/> run_settings	0 "8000" 1 "20000000" 2 "20"
<input checked="" type="checkbox"/> settings_data_file	last_mod_time "2009-12-04T16:00:05Z" lfn "tier0/091204180005_settings" md5hash "a14501b45c72084fbec99c2638d9a641" pfn "/mnt/raid/data/Soudan/Data/BeGe/tier0/091204180005_settings"

'Run' Database



Process Management

- Python daemons
 - Provides glue between CouchDB, ROOT, and MGDO.
 - Uses 'views' to determine what needs processing.
 - Data management daemons have stably handled ~100s of GBs of data over > a half year.

Summary

- DB used for process management, data flow management
- Analysis chain using CouchDB, python, MGDO, ROOT (+ some others) has been implemented, running stably