Assessment of Data Quality in Atlas

Michael G. Wilson

European Organization for Nuclear Research

for the ATLAS Data-Quality Group

International Conference on Computing
in High Energy and Nuclear Physics
(CHEP 07)

Victoria, BC Canada

5 September 2007

Topics

Methods for assessing the quality of recorded data must be fast, reliable and thorough

- Overview of data-preparation processes and their monitoring
- Tools for conducting assessments
- Experience from detector and software commissioning

Considerations and Challenges

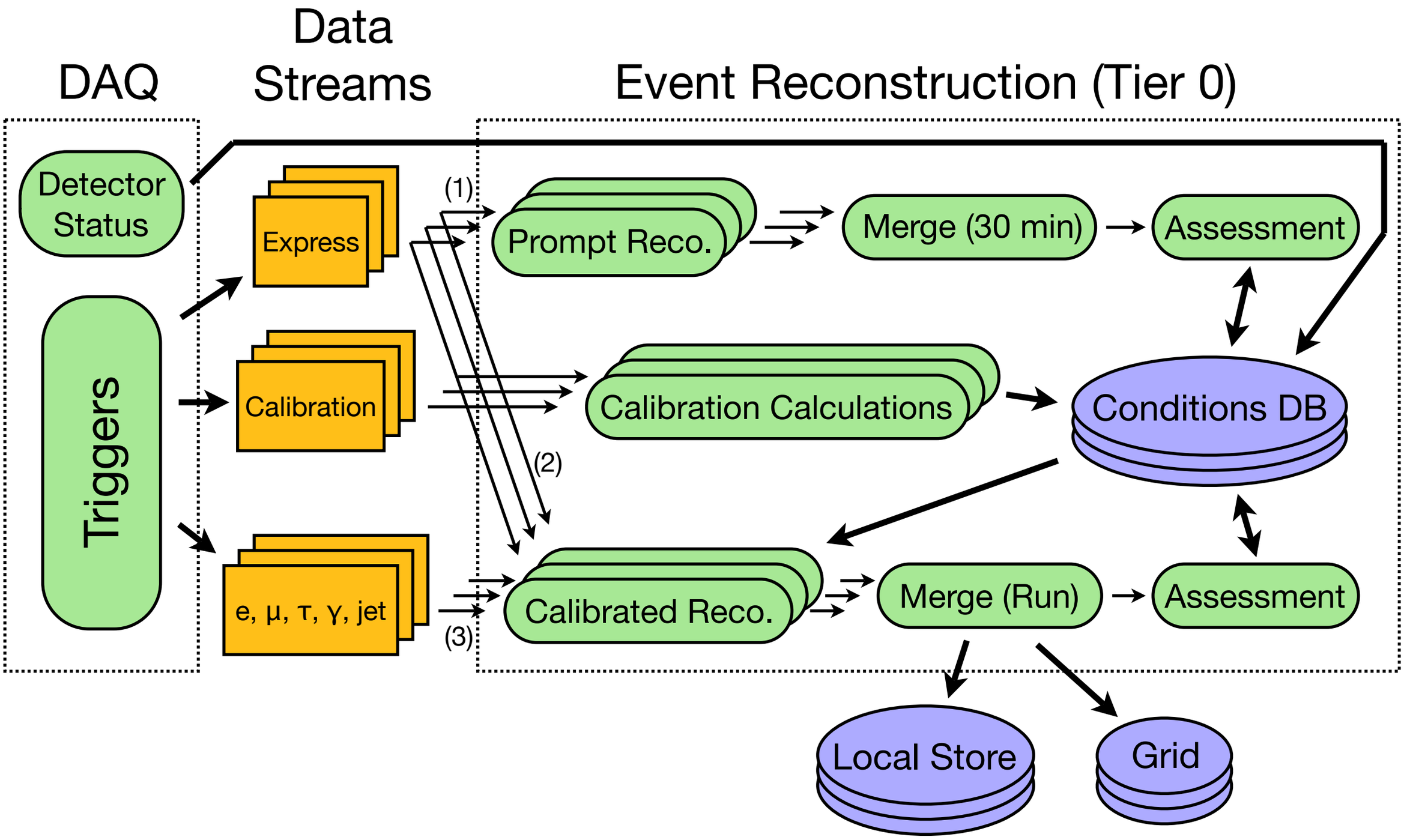
Data are processed once on a dedicated farm at CERN before being distributed on the Grid

- Computer farm comprises *O*(1000) nodes processing *O*(10,000) files per run (*O*(1000) files per stream)
- Validation timescale is 1–2 days after data are recorded

Monitoring frameworks and tools must accommodate needs of trigger systems, detectors, and physics-calibration groups

- Compatible with existing tools for commissioning
- Scalable and reliable for LHC running

Overview of Environment



Types of Quantities to Monitor

Physics-calibration quantities depend on stream and trigger information

- Ex., cross-check e^+e^- and $\mu^+\mu^-$ invariant-mass distributions—run similar apps on *e* and μ streams
- Within a stream, check relative stability of triggers

Quantities have different relevant timescales

- Trigger rates, per minute
- Dynamic characteristics (efficiencies, noise), per run
- Stable characteristics (relative alignments, software), per week

Histogram Production

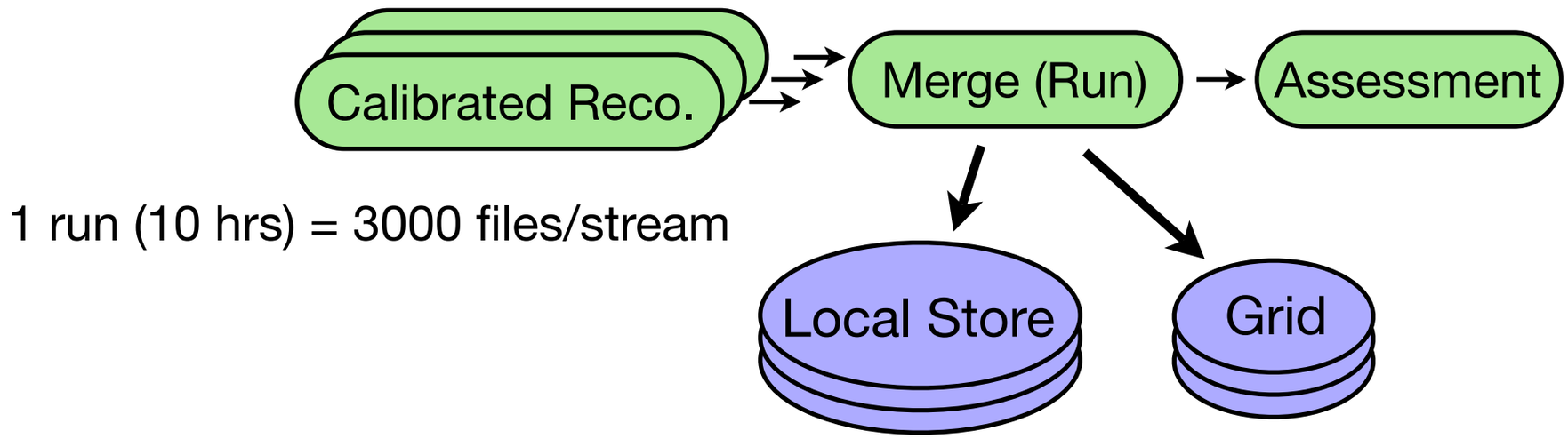
Interface provided to register histograms:

```
regHist( h, logicalPath, levelOfDetail, interval )
```

- Root ‘TH1’
- logical pathname within file
- ex., ‘shift’, ‘expert’, ‘debug’
- ex., ‘lumiBlock’, ‘run’, ‘fill’

- All details of creating and filling histograms left to systems
- Physical pathname determined at runtime; extra information (saved in `TTree`) used in merging and display
- All systems’ histograms written to one file

Histogram Merging

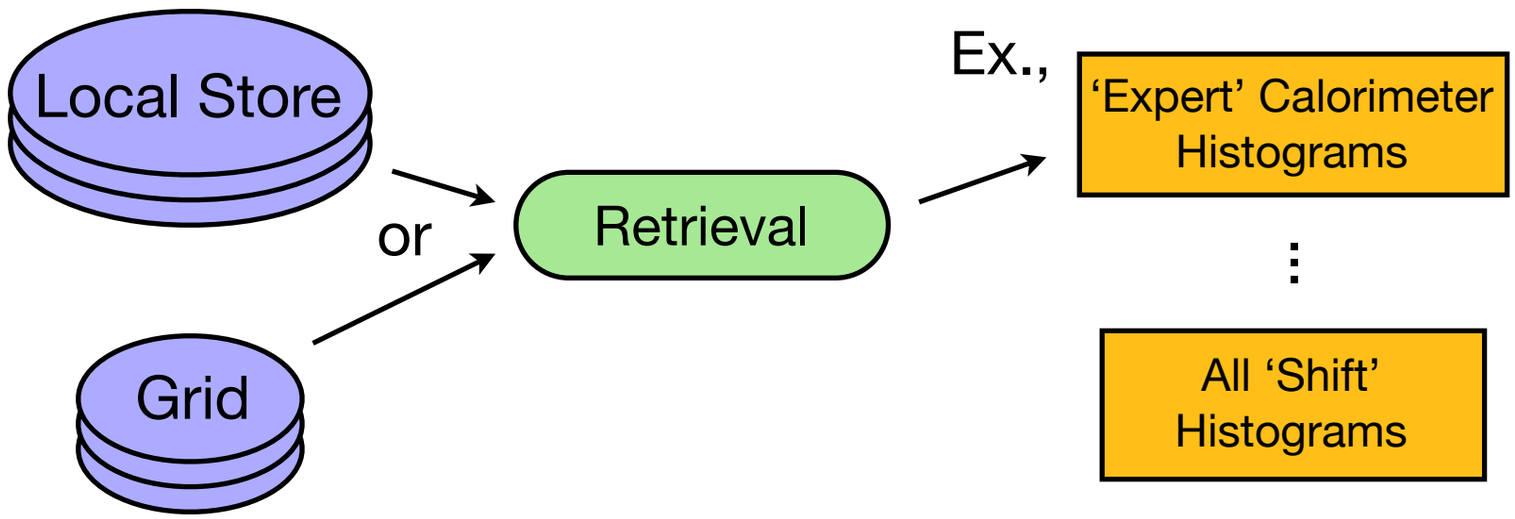


1 run (10 hrs) = 3000 files/stream

Each output file contains events from only a small fraction of a run

- Fast validation—merge express stream into ~30 minute blocks, evaluated promptly
- Full validation—merge into one file, per run, per stream
- Use ‘interval’ designations: some histograms span entire run, some span fractions of a run

Histogram Archiving and Retrieval



Archiving one file/run/stream ⇒ large file size

- A utility allows extracting histogram subsets by (run, stream)
  - Central data-quality display: all ‘shift’ histograms, from local store
  - Calorimeter expert in North America: Calo ‘expert’ histograms, from the Grid

Prompt Web Display

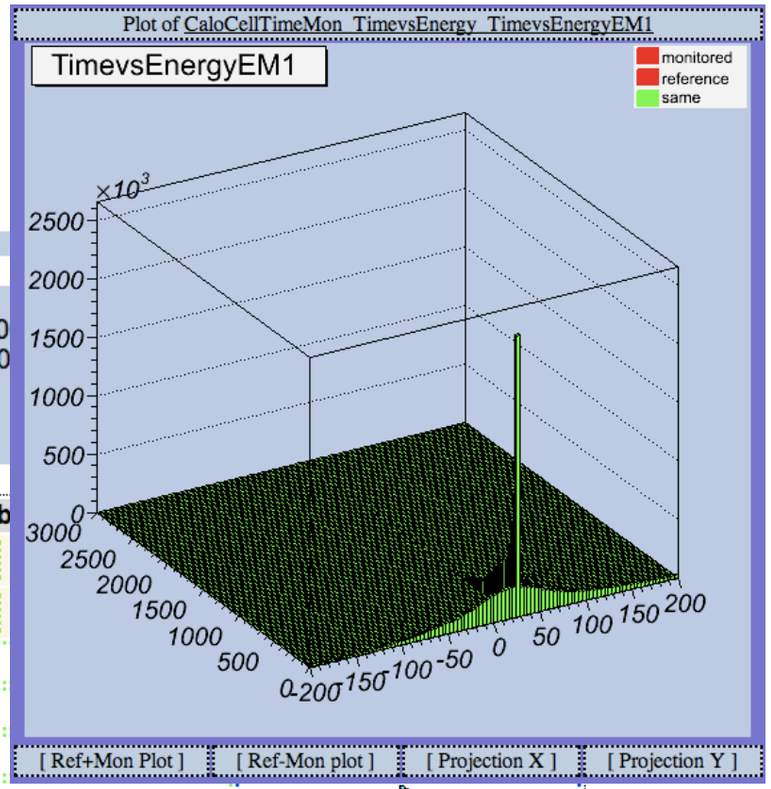
Histograms are displayed centrally on the web for fast and easy feedback

- A representative selection from each system
- Detailed histograms retrieved by experts on demand

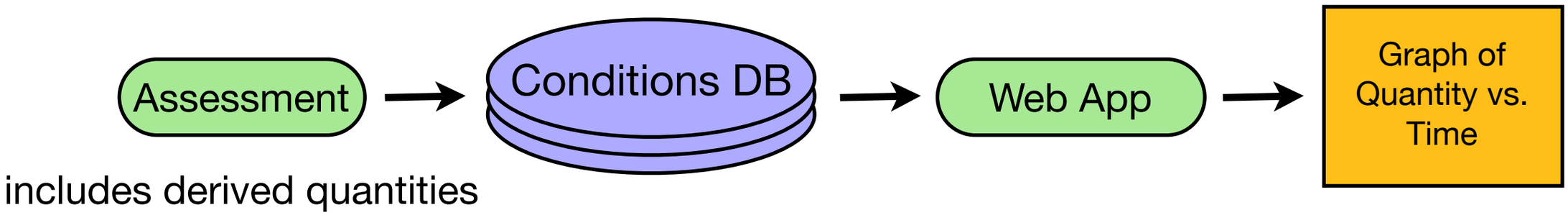
Prompt Web Display

Description: M3 Run 10915
Monitored file: /afs/cern.ch/atlas/maxidisk/d14/DCube/M3/reference/M3.combined.COSMHIST.run0010915.part0001._00
Reference file: /afs/cern.ch/atlas/maxidisk/d14/DCube/M3/reference/M3.combined.COSMHIST.run0010915.part0001._00
Date: 2007-06-15
File Descr.: M3 Cosmic-ray data
Log file: ./dq.log

name	K-S test	χ^2 test		"b
/		368	383	751
/CaloCellTimeMon/TimevsEnergy		4		4
TimevsEnergyEM1	---	1.000		
TimevsEnergyEM2	---	1.000		
TimevsEnergyEM3	---	1.000		
TimevsEnergyPSR	---	1.000		1.000
/CaloCellVecMon/DistanceInfo		47	4	51
/CaloCellVecMon/EnergyInfo		61	35	96
/CaloCellVecMon/EtaPhiInfo		105	39	144
/CaloCellVecMon/SumEvsEtaPhiInfo		36	12	48
/CaloCellVecMon/Summary			1	1
NCell	---	0.000		1.000
/CaloClusterShwrMon		3	1	4
/CaloClusterTimeMon		1	3	4
/CaloClusterVecMon/Energy		6	6	12



Monitoring Time Evolution



Monitor the evolution of detector and its performance

- Calculate quantities for each run during assessment
- Quantities are stored in Conditions DB
- A web-based application creates plots on demand
- Flexible schema for easily adding new quantities

Data-Quality Status

Based on all available information, make assessments and store in conditions DB

- A configurable calculator makes automatic assessments
- Shifters review assessments, make adjustments

Data Quality Status Browser - Query Results

UPDATE DB | RESET | Manual update

Good | Flawed | Bad | Unknown | Empty

Run 200	Calorimeters											
	LAr								Tile			
LB interval	EMBA	EMBC	EMECA	EMECC	HECA	HECC	FCALA	FCALC	TILBA	TILBC	TIEBA	TIEBC
0 - 0												
1 - 1												
2 - 3												
4 - 5												
6 - 4294967295												

Detector-Integration Tests

Detector commissioning with cosmic-ray data is ongoing

- Use common monitoring tools now to validate them (even if some systems not returning useful data)
- As detectors come online, use common tools for commissioning (instead of custom solutions)

Feedback now is very valuable

- Address integration issues, find unmonitored quantities, uncover scalability problems in software
- Learn how to monitor systems simultaneously (common reconstruction, detector synchronization, etc.)

Software-Integration Tests

Many features of the anticipated event reconstruction tested with simulated data

- Streaming model, luminosity calculations, trigger menus
- These features are an essential part of LHC data taking
- At the end of this month, will test:
  - Monitoring of different streams
  - Monitoring quantities as a function of trigger path

Conclusions

ATLAS will be ready to assess data quality as soon as the data are available

- Infrastructure has been considered and planned with participation from all systems
- First implementations of tools are ready—tuning, fixing, and extending is underway