The Offline Framework of the Pierre Auger Observatory

Lukas Nellen
(for the Pierre Auger collaboration)
ICN-UNAM
lukas@nucleares.unam.mx

Supported by PAPIIT IN114924

ICHEP 2024 — Prague — 2024-07-19

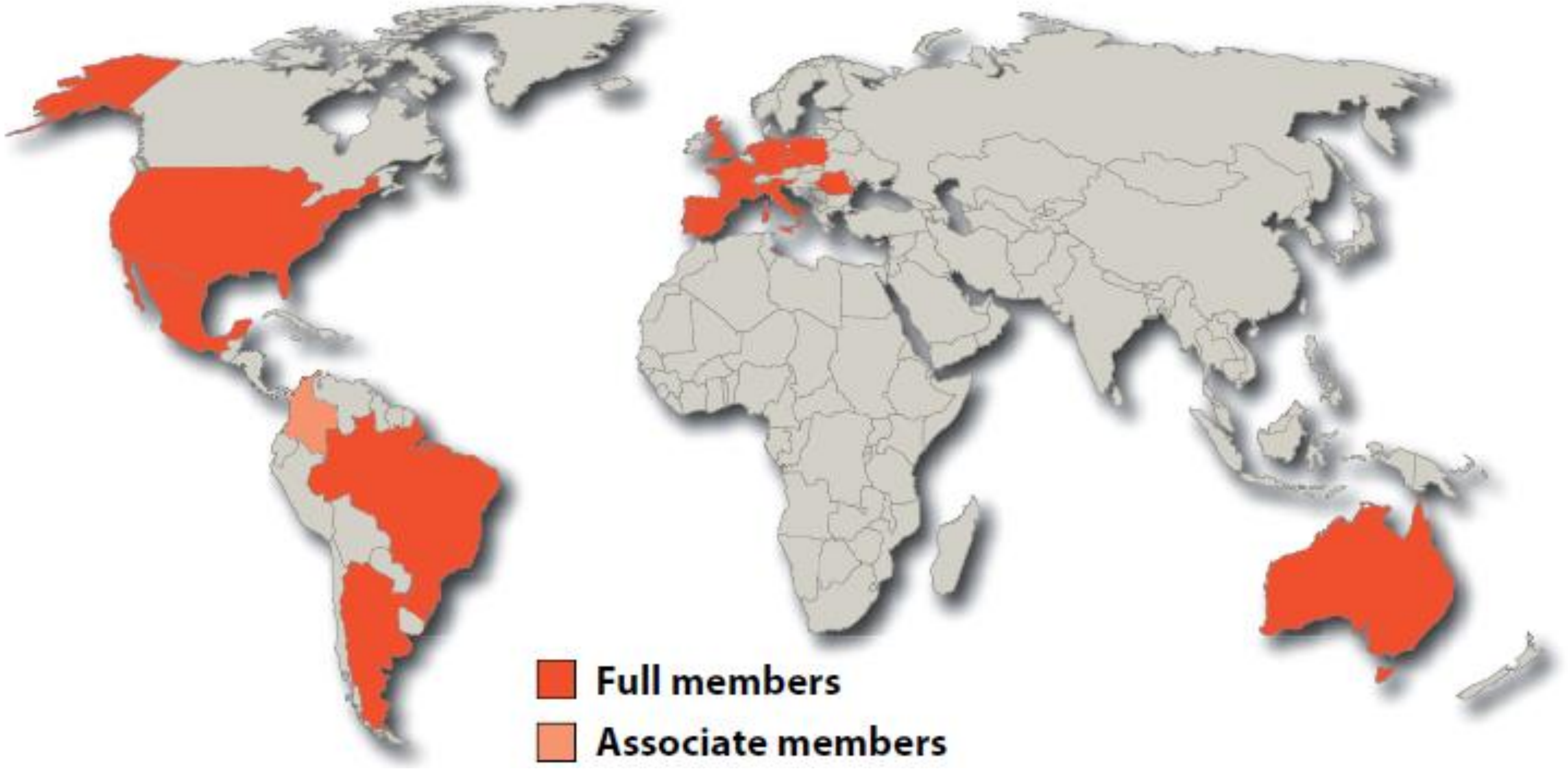
Auger Offline framework

- Started over 20 years ago: first commit from *31st of January, 2003*
  - CVS → SVN → git(lab)
- Flexible and extensible
- Standard framework for
  - Detector Simulations
  - Calibration
  - Reconstruction
  - Data preparation
  - (Some) Analysis
- Input
  - Air Shower simulation
  - Locally generated (e.g., to simulate calibration)
  - Offline format
- Output
  - Offline format
  - raw data
  - Auger Data Summary Trees (ADST): for further analysis
- Implementation in C++
  - C++98 → C++ 11/14 (→ C++ 17/20)

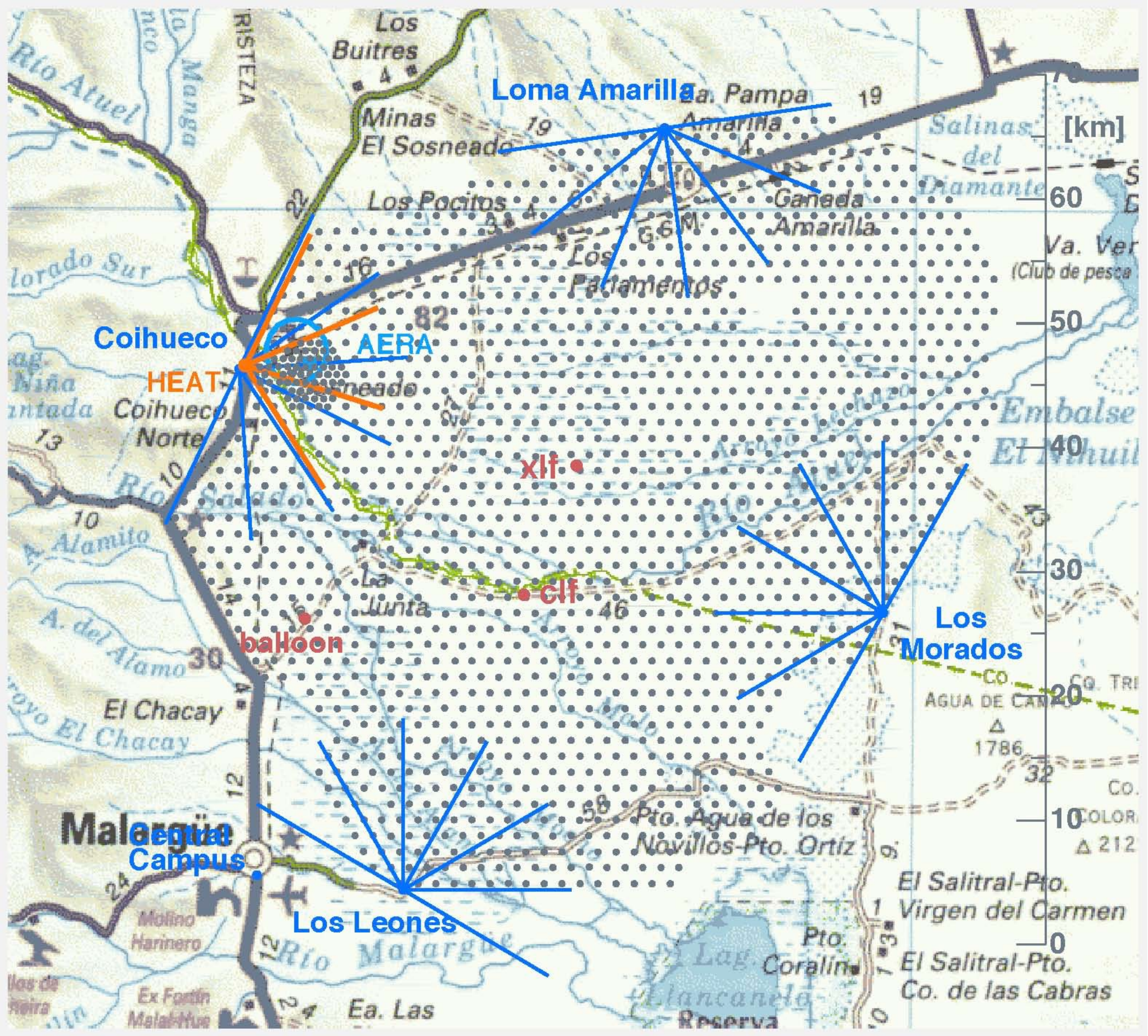
The Pierre Auger Collaboration

17 countries , ≈460 collaborators

Argentina – Australia – Belgium – Brazil – Colombia – Czech Republic – France – Germany – Italy – Mexico – Netherlands – Poland – Portugal – Romania – Slovenia – Spain – ~~United Kingdom~~ – United States



The Auger Site in Malargüe



1660 surface detector stations, 1.5 km spacing

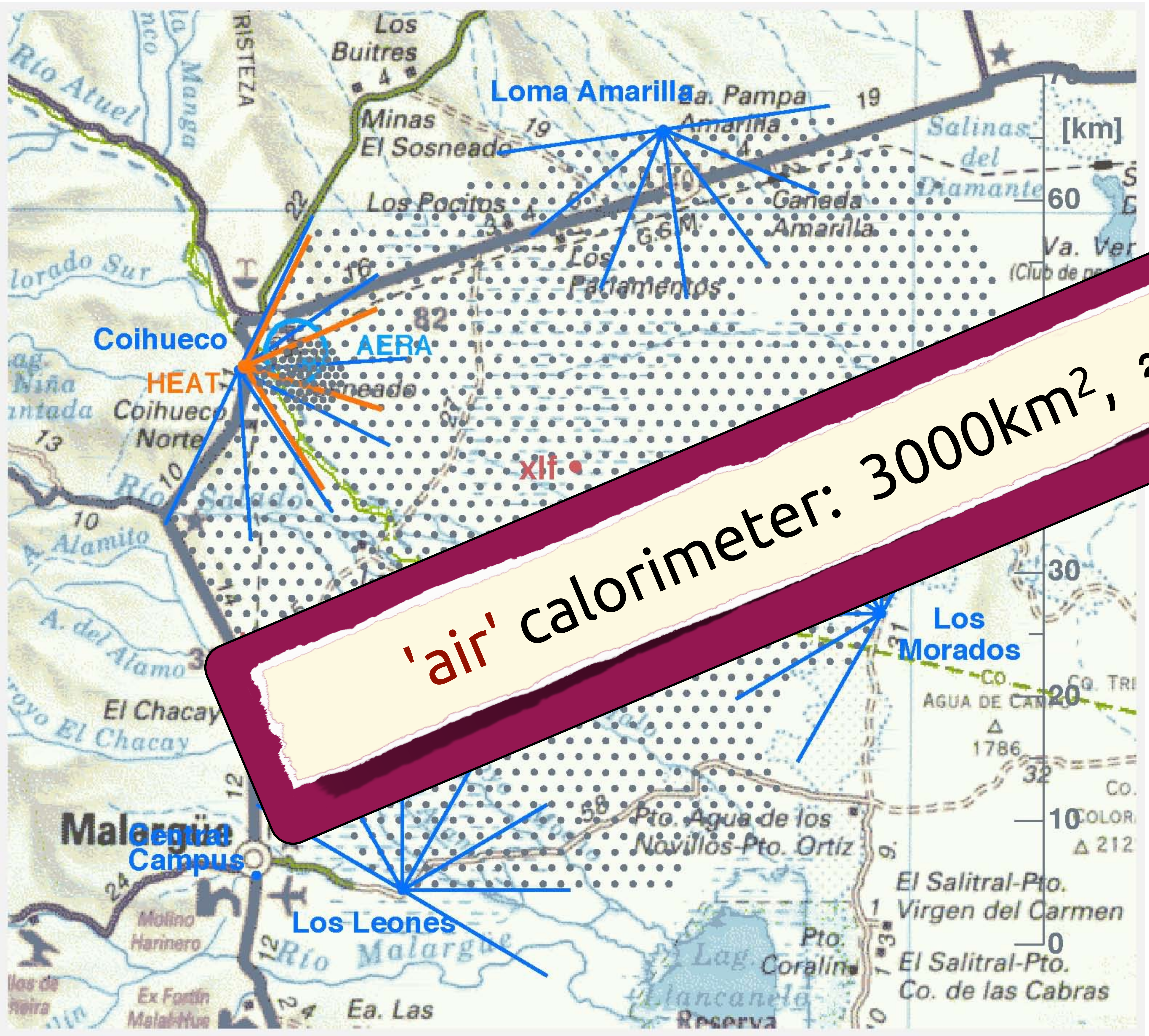
Infill: 750m spacing
+ buried μ detectors

4 Fluorescence detector sites

- 6 telescopes each
  - +3 elevated
- 27 telescopes in total
- Full coverage of the surface array
- Capability to detect stereo events
- Quadruple events seen

Low Energy Extensions
Radio Detectors

The Auger Site in Malargüe



'air' calorimeter: $3000 km^2$, $\approx 2.5 \cdot 10^{10}$ T

1660 surface detector stations, 1.5 km spacing

Infill: 75 ... ing

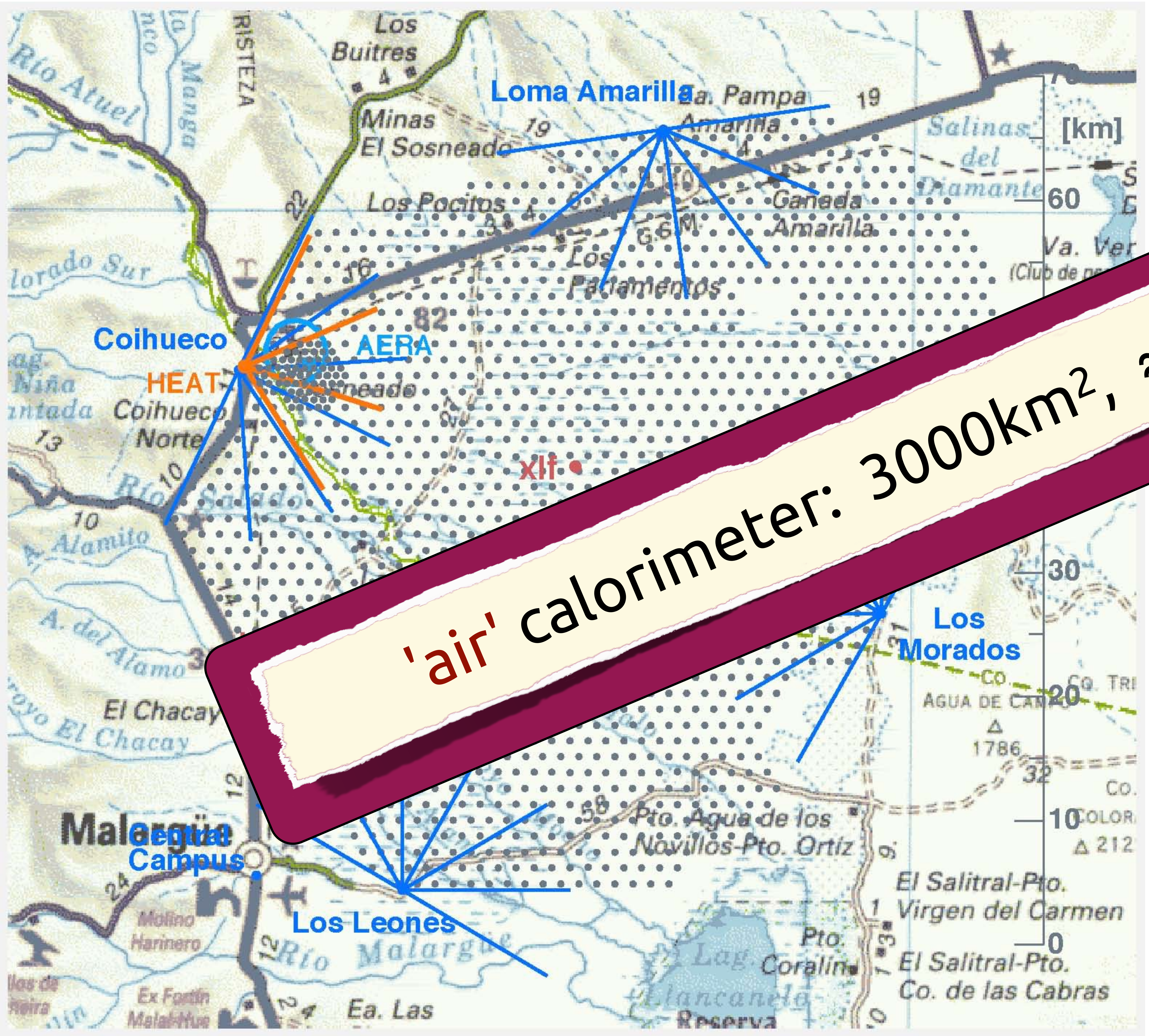
... tors

... escence detector sites

- ✱ 6 telescopes each
  - ✱ +3 elevated
- ✱ 27 telescopes in total
- ✱ Full coverage of the surface array
- ✱ Capability to detect stereo events
- ✱ Quadruple events seen

Low Energy Extensions
Radio Detectors

The Auger Site in Malargüe



'air' calorimeter: 3000km², ≈2.5·10^10 T

1660 surface detector stations, 1.5 km spacing

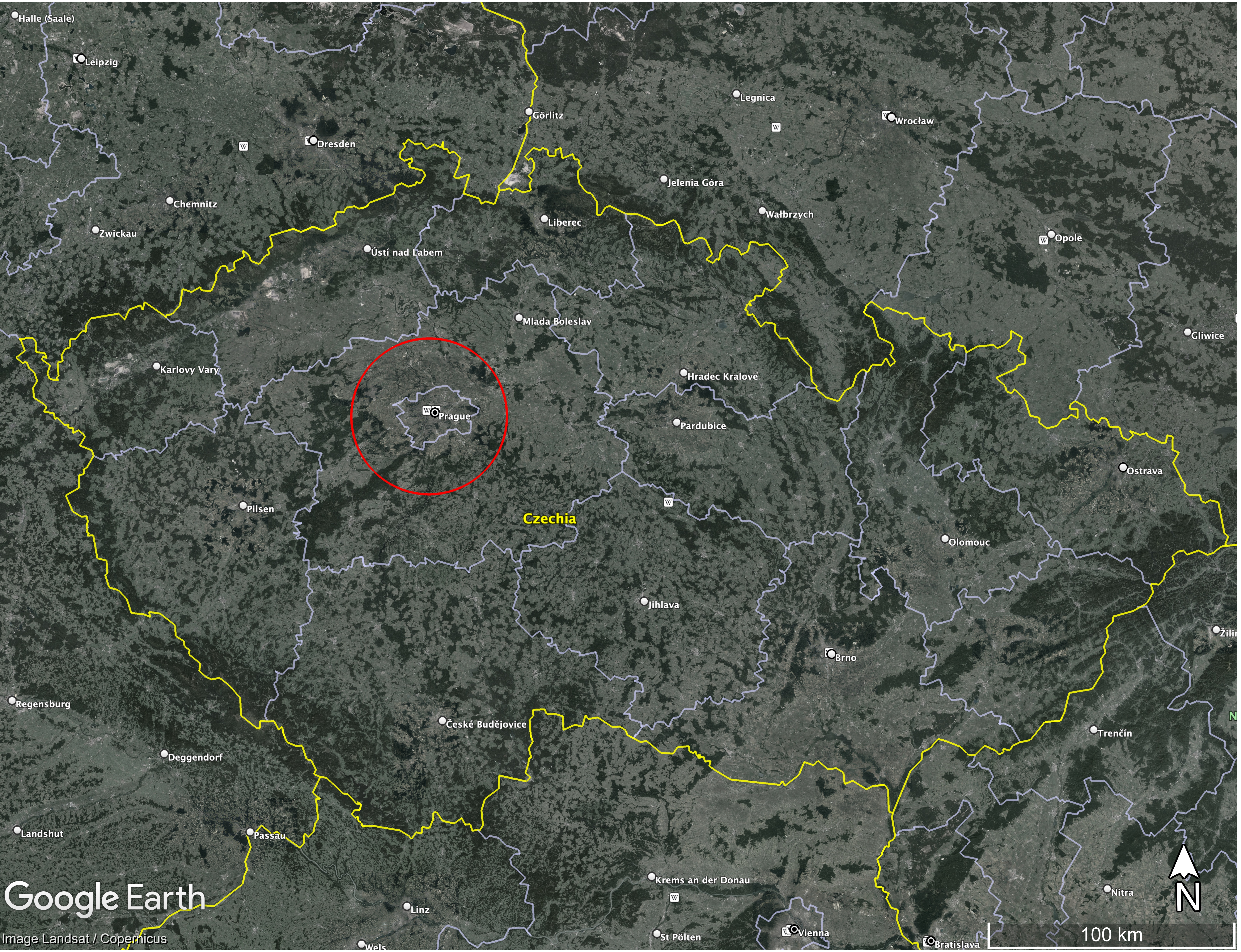
Infill: 75 ... ing ... tors

Upgrade (Phase II)

- ✱ Faster Electronics
- ✱ Scintillator on detector stations
- ✱ Buried muon detectors (in part of the array)
- ✱ Small PMT to expand dynamic range
- ✱ Antennas for radio detection

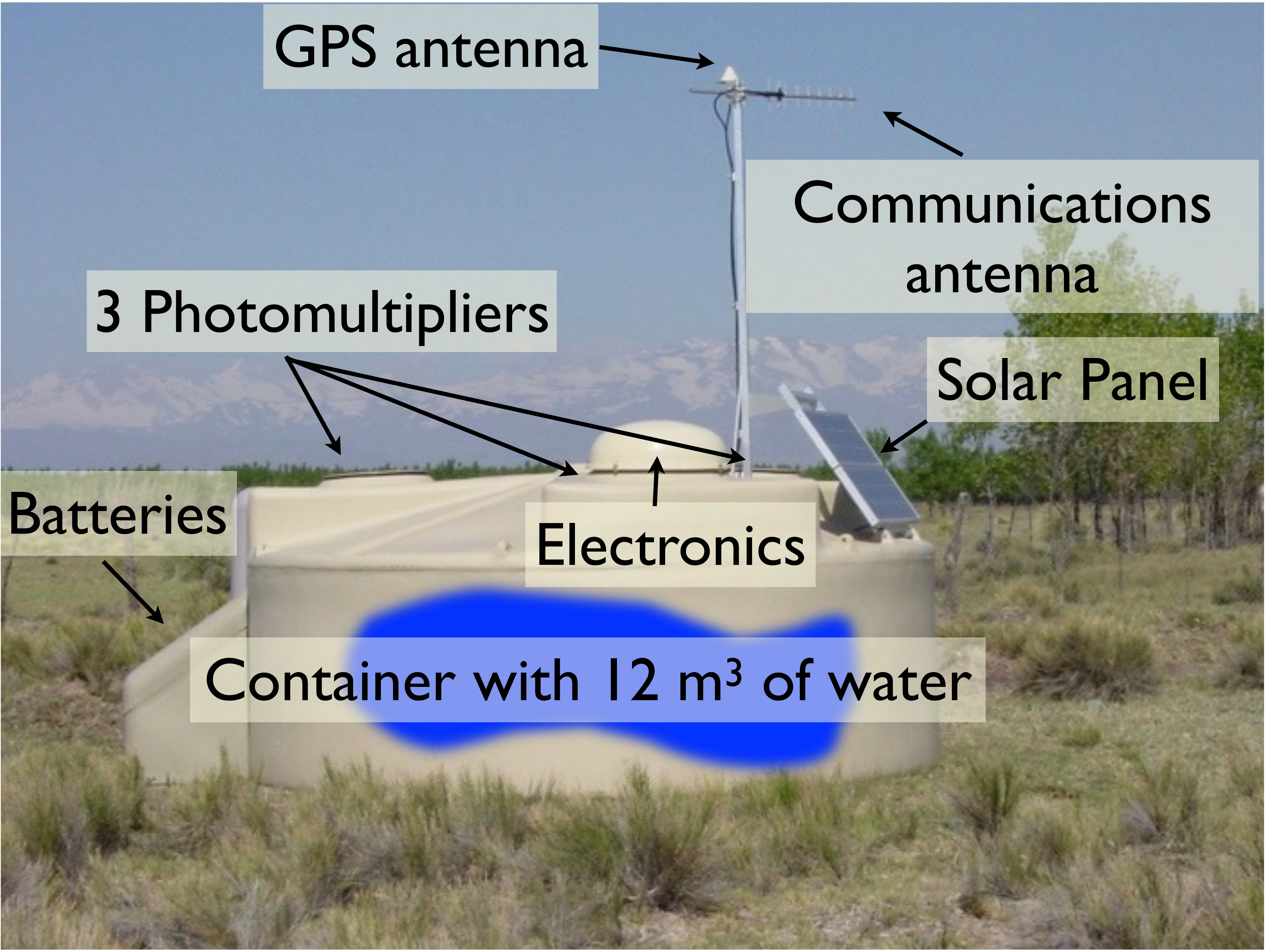
- ✱ Finalizing installation
- ✱ Commissioning

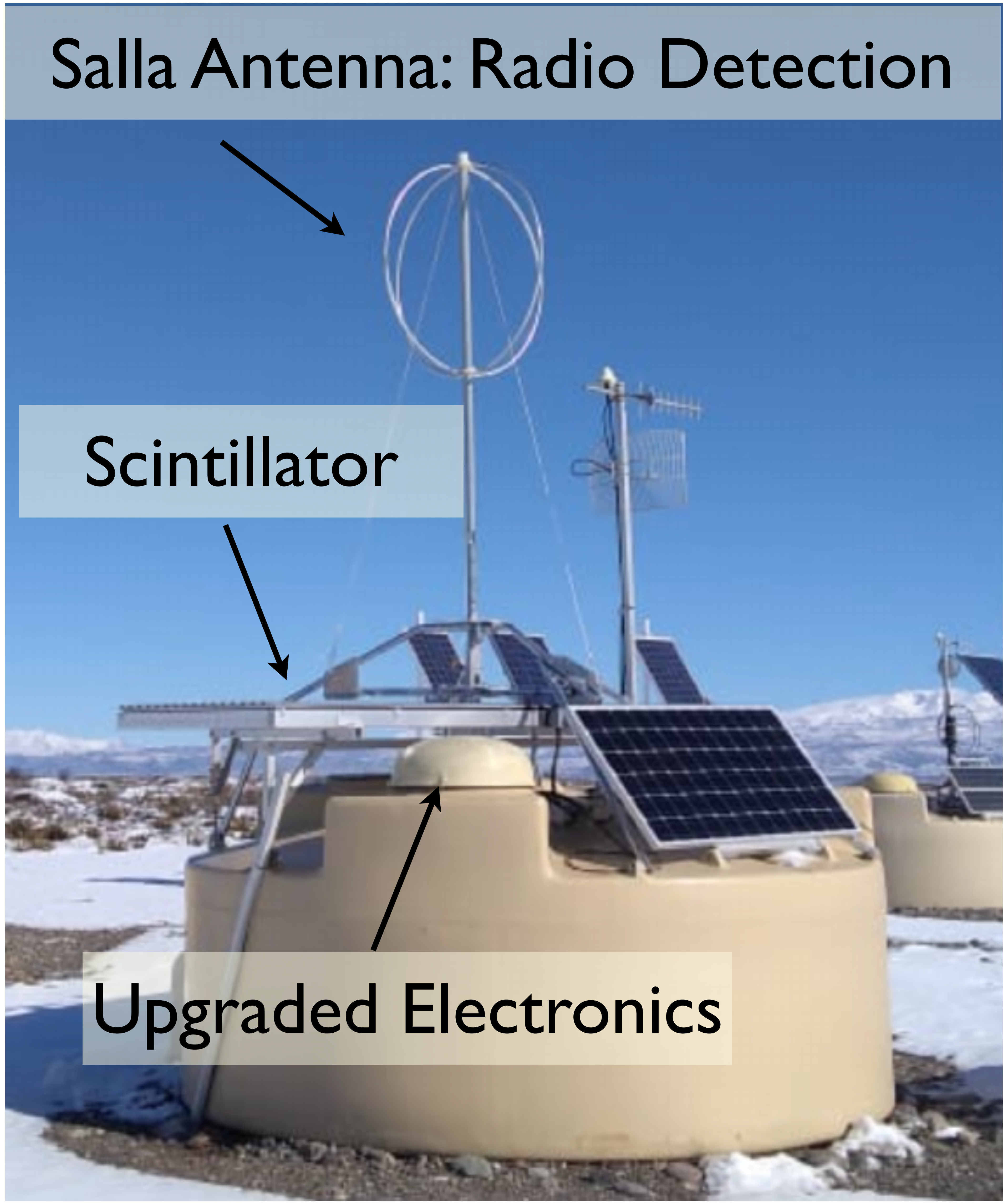
Auger size in the Czech Republic



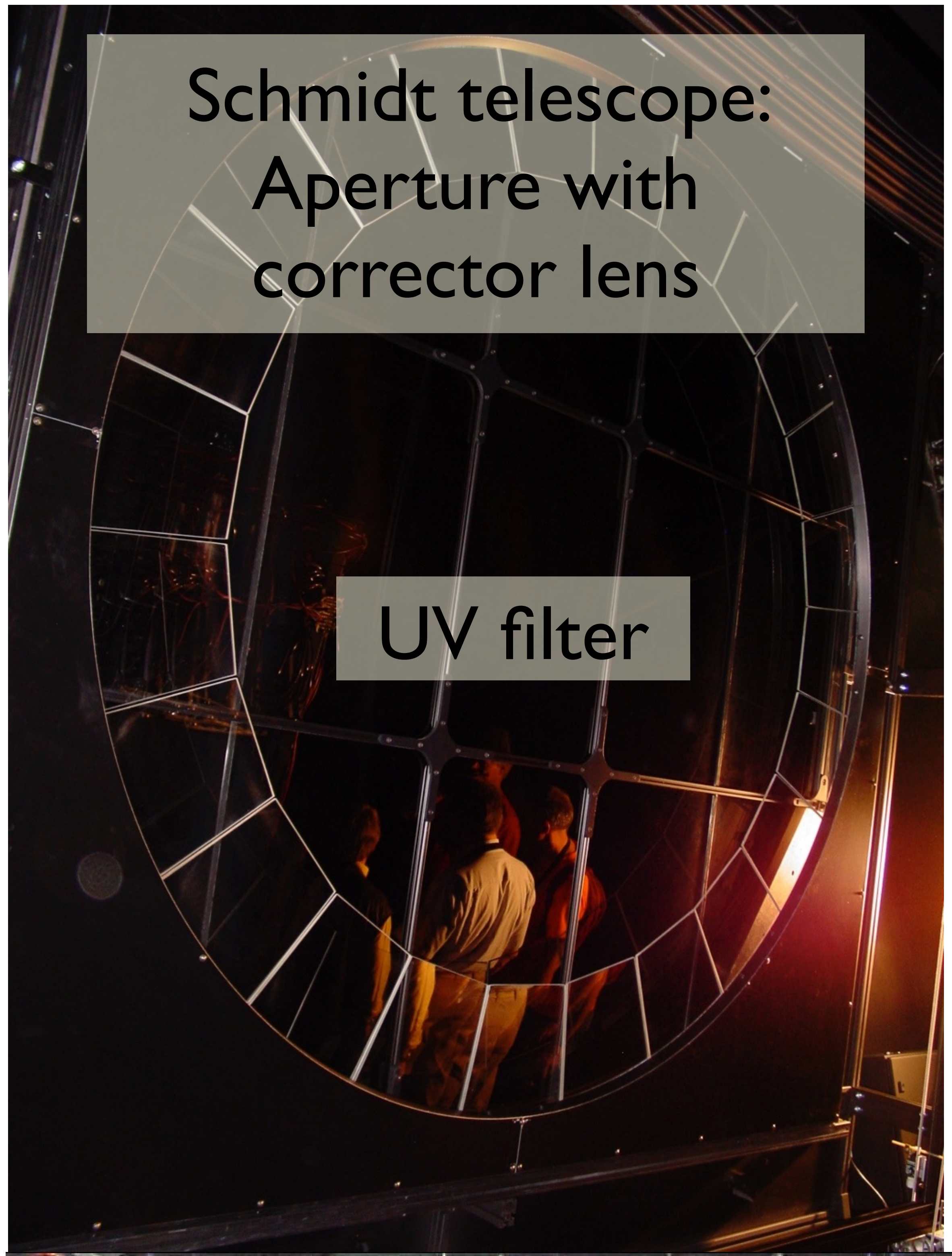
- Circle: 3,000 km² ⌀ 60km
- Centered over Prague Conference Center
- Luxembourg: 2,586.4 km²

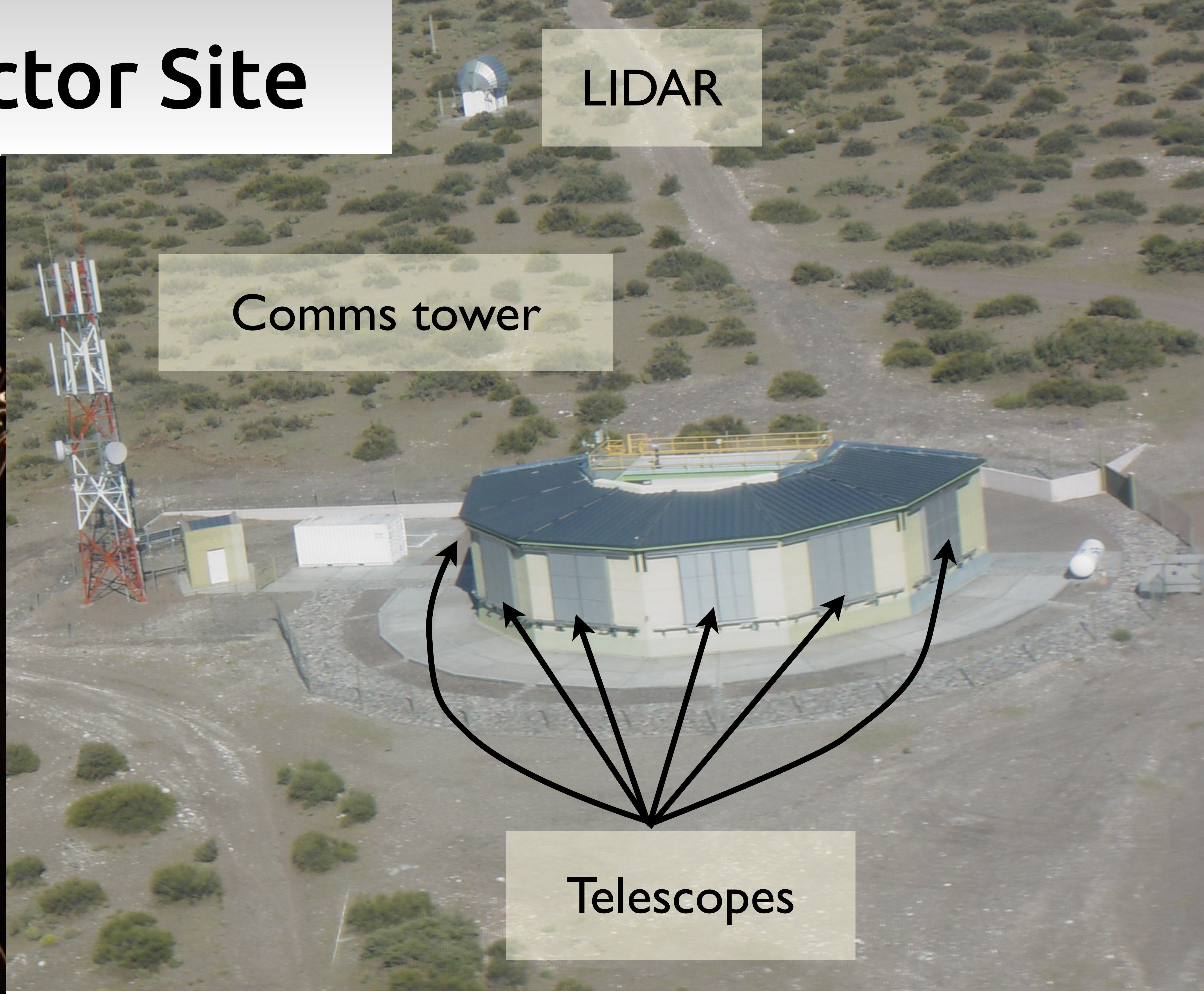
A surface detector station

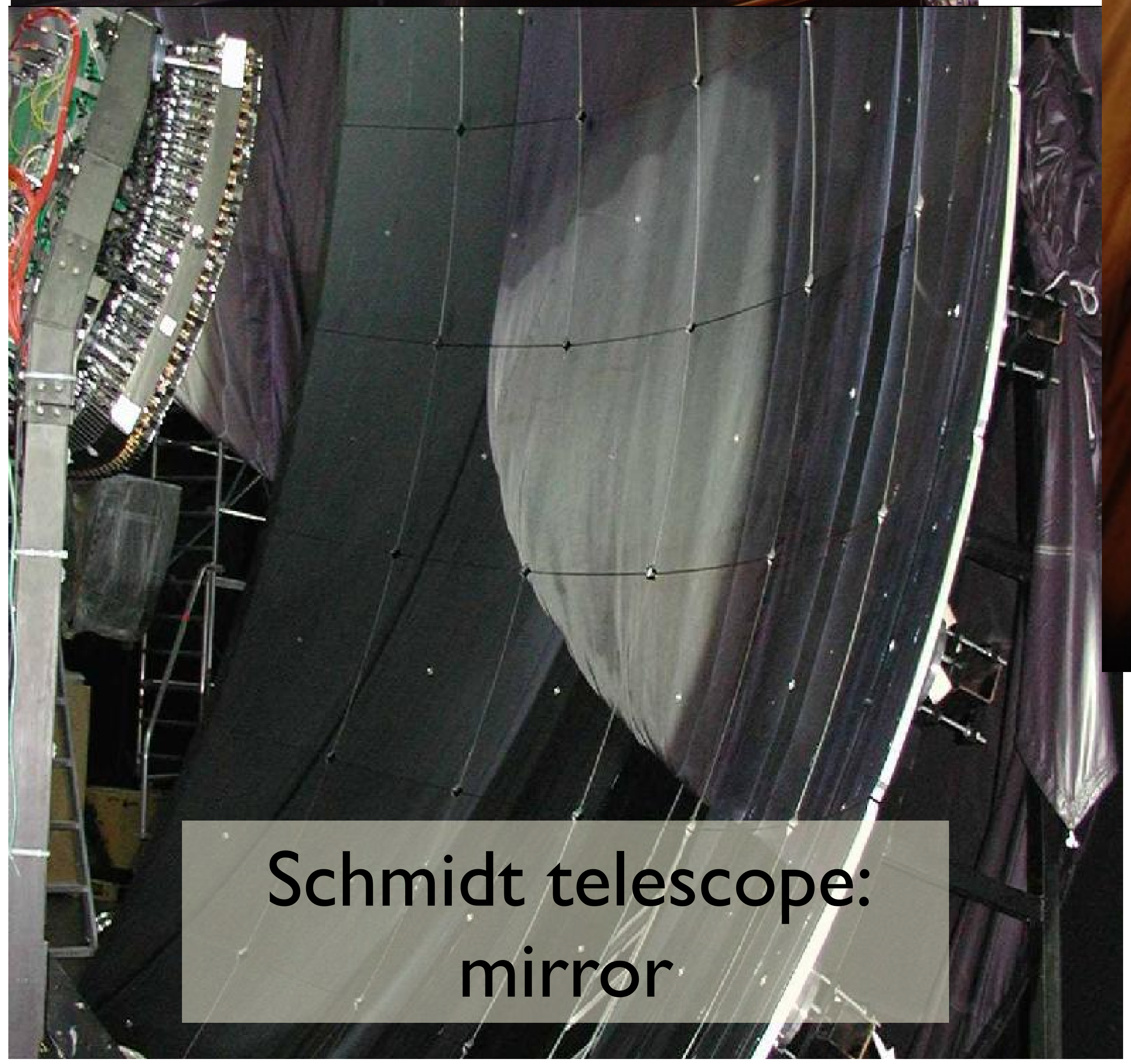


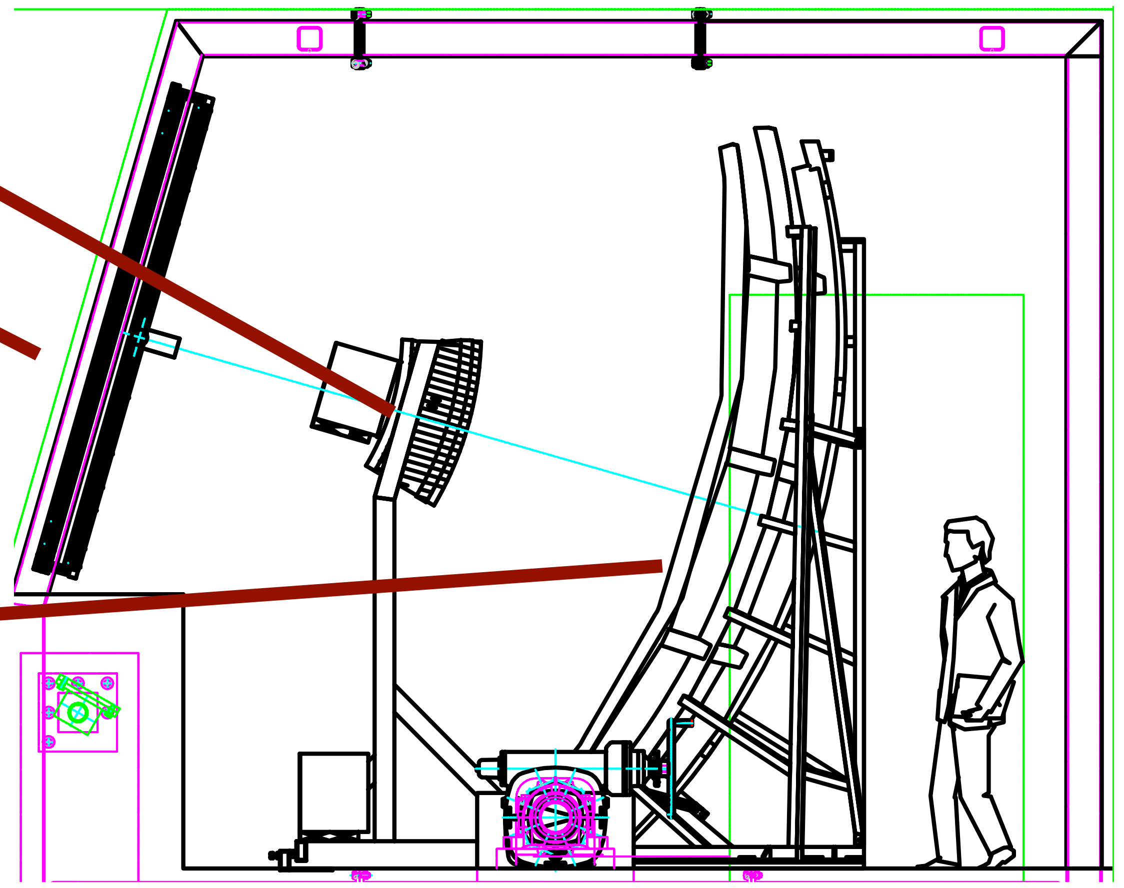


A Fluorescence Detector Site

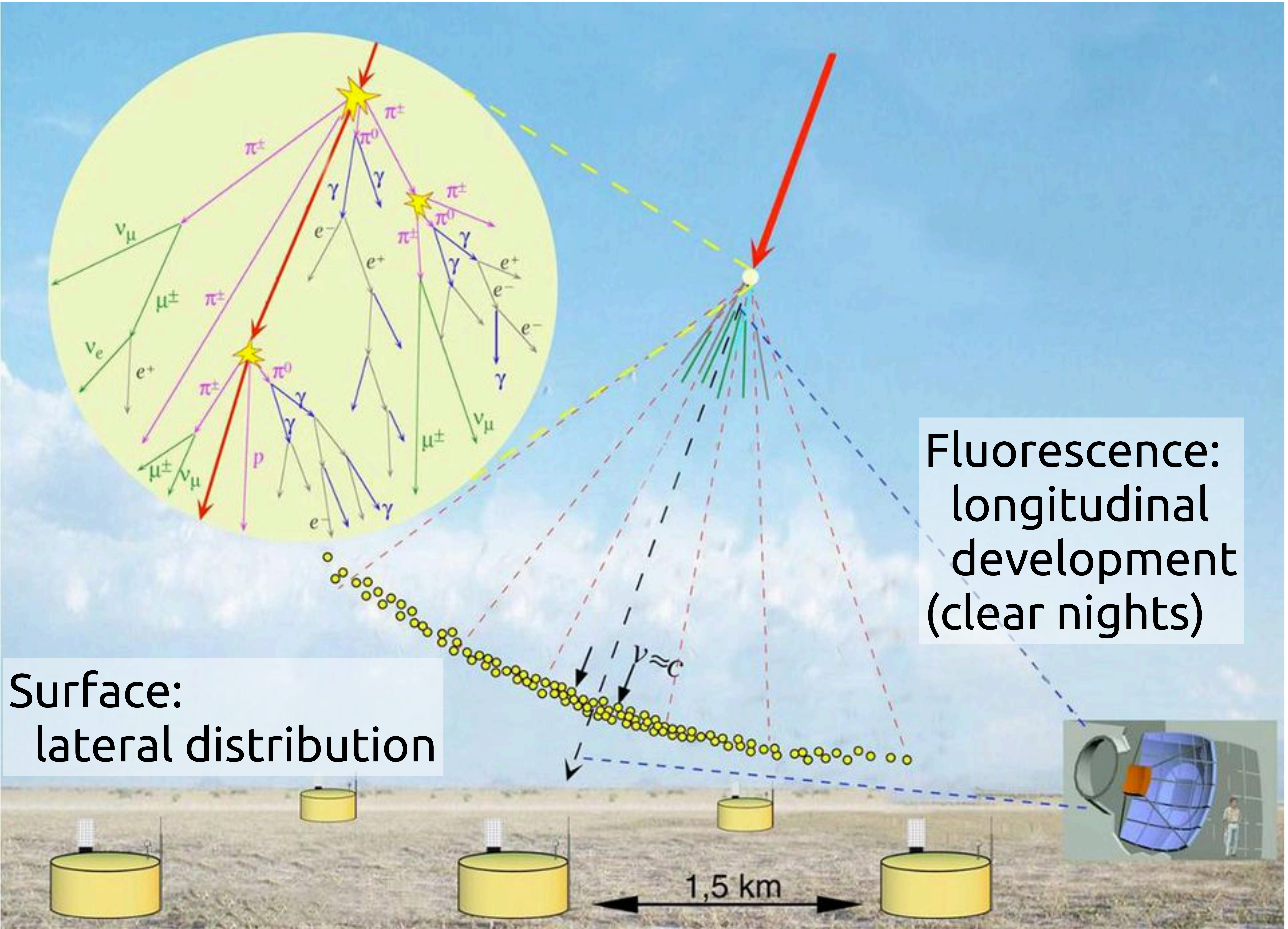








Hybrid Air shower detection



- High-quality hybrids
- Limited by FD duty cycle
- Calibration for SD only analysis

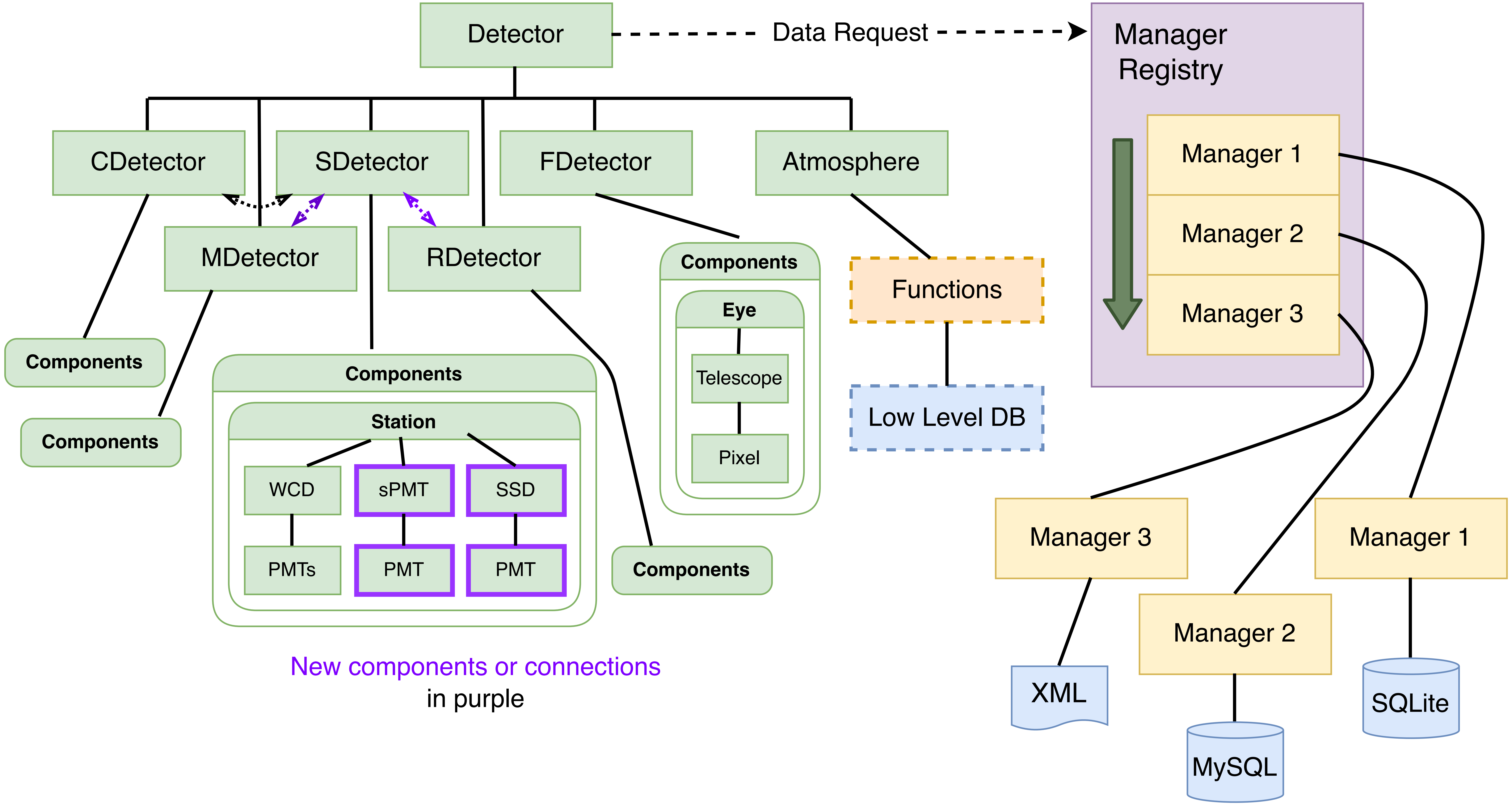
Framework Hierarchy

- Software components structured in non-cyclic hierarchy
  - Access only to components lower in hierarcy
  - Clean dependencies for parallel building

- Separate data holders from algorithms

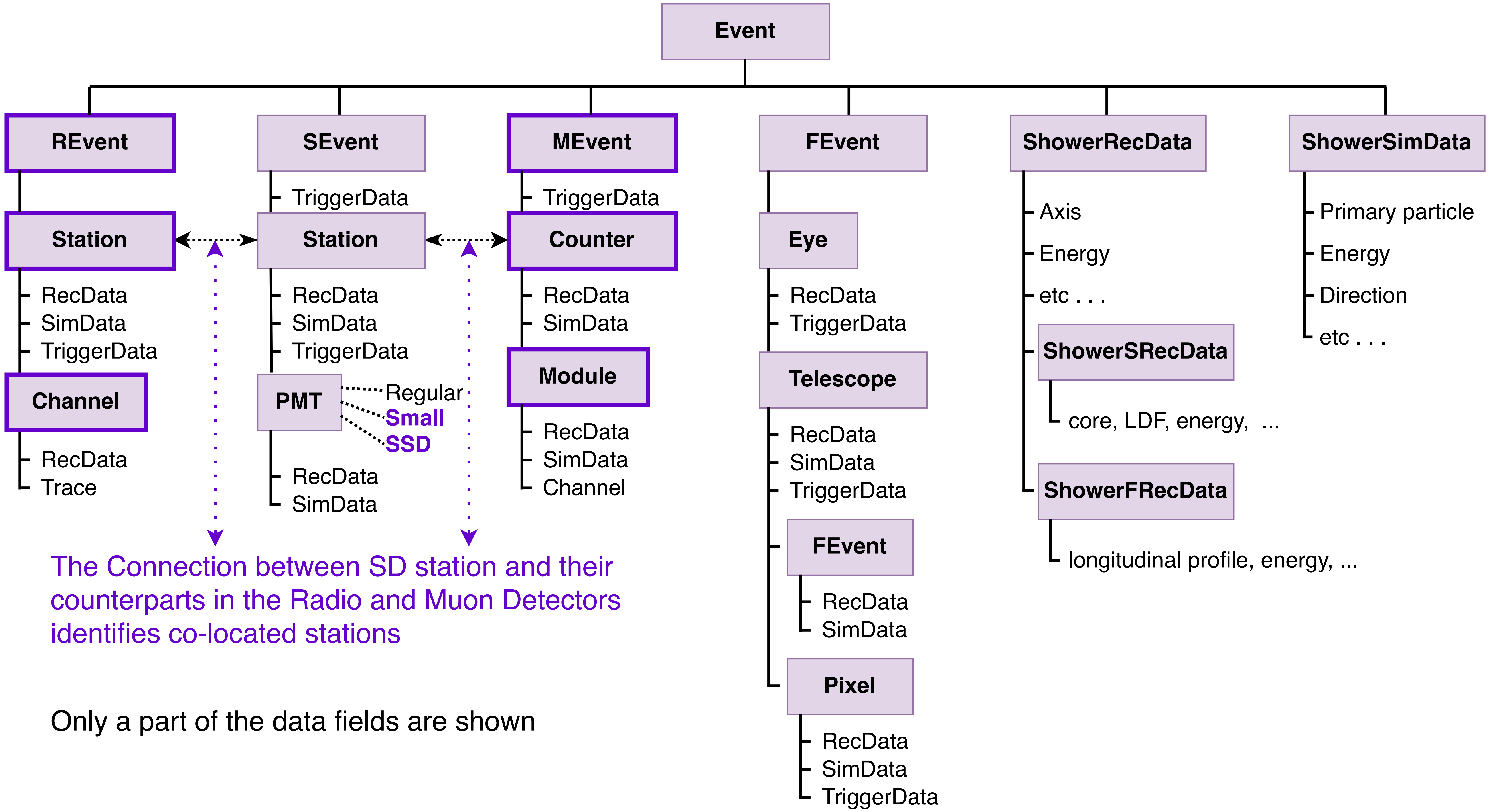
- ADST – can be used stand-alone
- Utilities
- Framework
  - Detector, Event, RunControl
- Tools – used in multiple modules, but need Framework
- Modules – Simulation and reconstruction happens here

The Detector – slowly changing



- Structure follows detector hierarchy
- Atmosphere is part of the detector
- Managers as abstraction for data access
  - Configurable

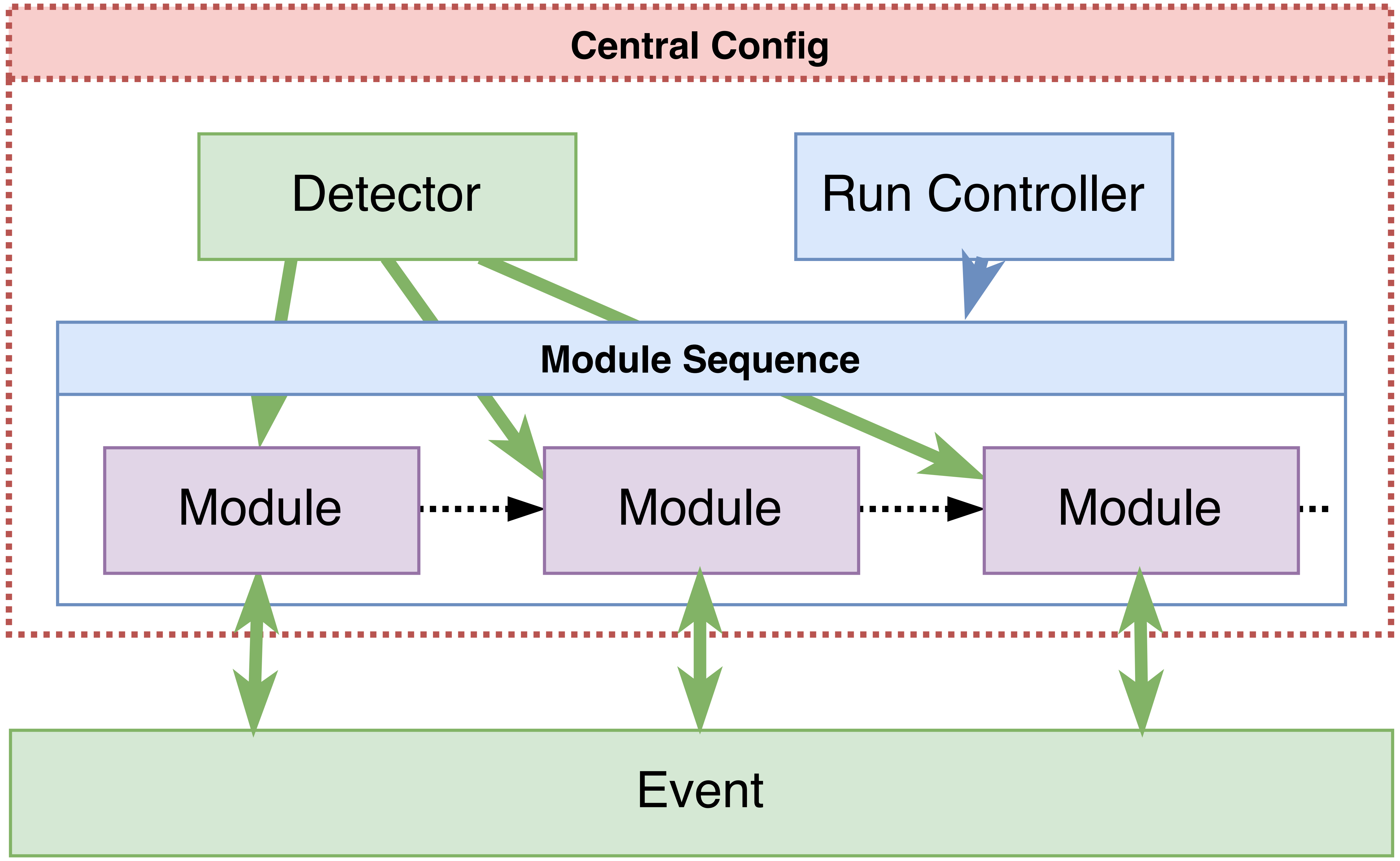
Event



Only a part of the data fields are shown

- Structure parallel to detector
- Mostly write-only
  - Delete only when unavoidable
- Not all fields shown

Control Flow



- Application: sequence of steps
  - Encapsulated as Modules
- RunController
  - Configures sequence
  - Schedules execution
- CentralConfig configures
  - Detector
  - RunController
  - Modules
- Detector is read-only
- Event transports information between modules

Event I/O

- Libraries with support from Modules

- Initial filling of the event at the start of the loop
  - From external simulations
  - Generated or parametrized
  - Generally a thin layer around input libraries

- Different output formats
  - Offline format for loss-less event storage
    - Can resume module chain
  - Other formats only store select data
    - Raw data
    - ADST files
    - JSON for Open Data

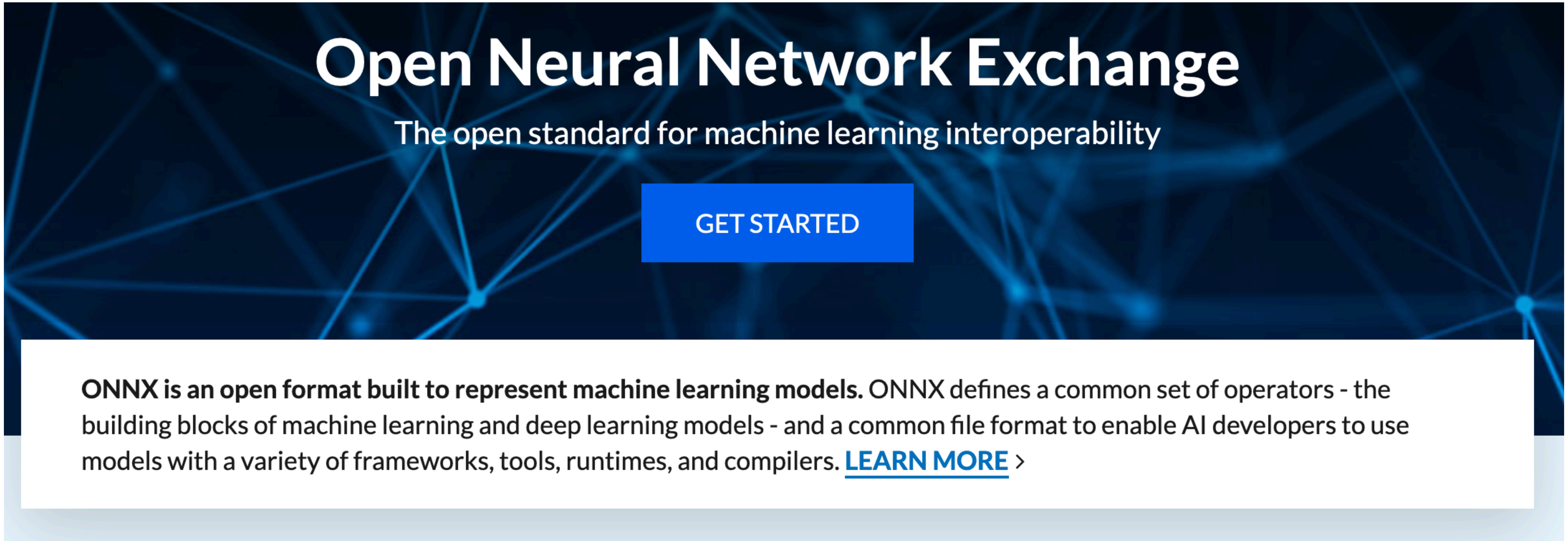
- General introspection would be nice

Some changes we had to do

- Upgraded electronics: different digitizer frequency
  - Was hard-coded
- Additional channels, PMTs
  - Was hard-coded
- Different time delays
  - PMT characteristics
  - Change of FPGA
  - Have to match for comparison of Phase I and II
- Fix non-compliant code, e.g., incomplete headers

In progress: Machine Learning support

- Currently: everybody uses their favourite framework
- Need to be able to use inference in regular modules
- Training can be done in specialized application
- Evaluating Open Neural Network Exchange (ONNX)
  - Can provide C++ representation
  - Training typically done in python-based framework



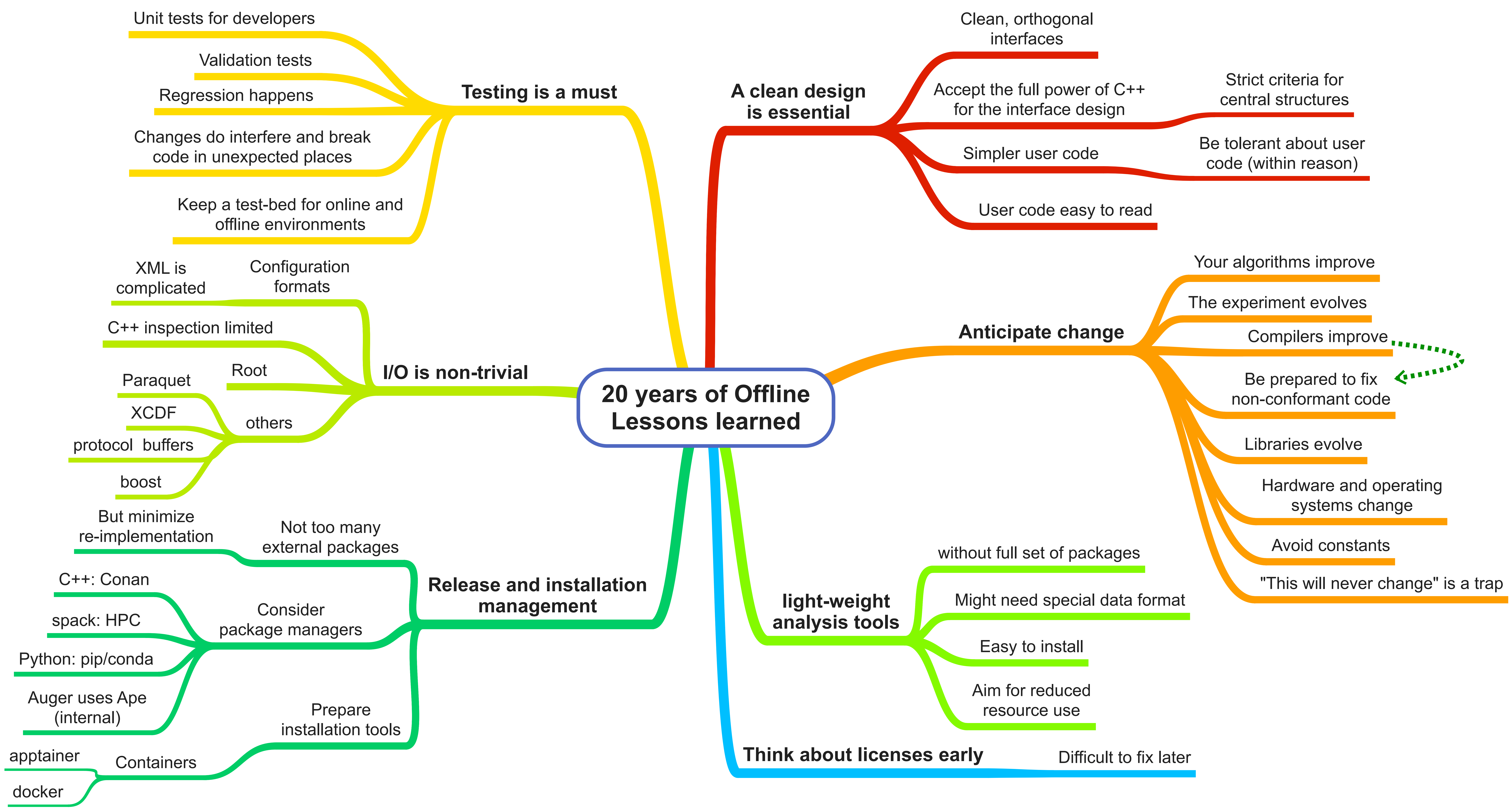
Improving distributions

- Build your own
  - Developed APE to install dependencies
    - Multi distribution, multi platform
  - Slow, but (reasonably) reliable

- Binaries is CVMFS
  - Needed for GRID production
  - Can be used by end-users with CVMFS installed

- Containeres: apptainer
  - Compiled code
  - Dependencies and compilers for development
  - Ship in native format and CVMFS unpacked
  - Positive user experience (but so far limited use)

Lessons Learned



Key lessons

- Avoid hard-coding anything
  - "This will never happen" is almost always wrong
- Licensing matters
  - It is hard to track contributors after some time
- Use a dependency manager
  - Hand installation is a headache
    - For testing
    - For users (Release!)
  - We wrote our own (APE)
    - We have community projects now
- Maintaining good documentation is difficult
- Regular user training desirable
- APIs have to be simple for users
  - Internals and implementation details can be complex
  - Assuming you have the experts in your team
- "Testing first" works
  - Some non-trivial code survived w/o major bugs for 20 years
- I/O is still a headache
- Add light-weight tools
  - Python
  - ROOT (Here: ADST)

Summary

- The Offline Framework stood the test of time
  - Could be extended to handle various extensions and a major upgrade
- Parts have been adapted by other projects
  - NA61/Shine, Jem/EUSO, HAWC, CORSIKA 8, IceCube, SWGO (?)

- Mass production of simulation not covered here
  - Auger is largest user of EGI infrastructure after LHC

- Currently focussing on commissioning of the Auger Upgrade
  - See other talks:
    - *Astro-particle* session on Saturday
    - Martin Schimassek in *Operation, Performance and Upgrade* session on Saturday
    - Viviana Scherini in *Outreach* session

Extra: Auger over Prague