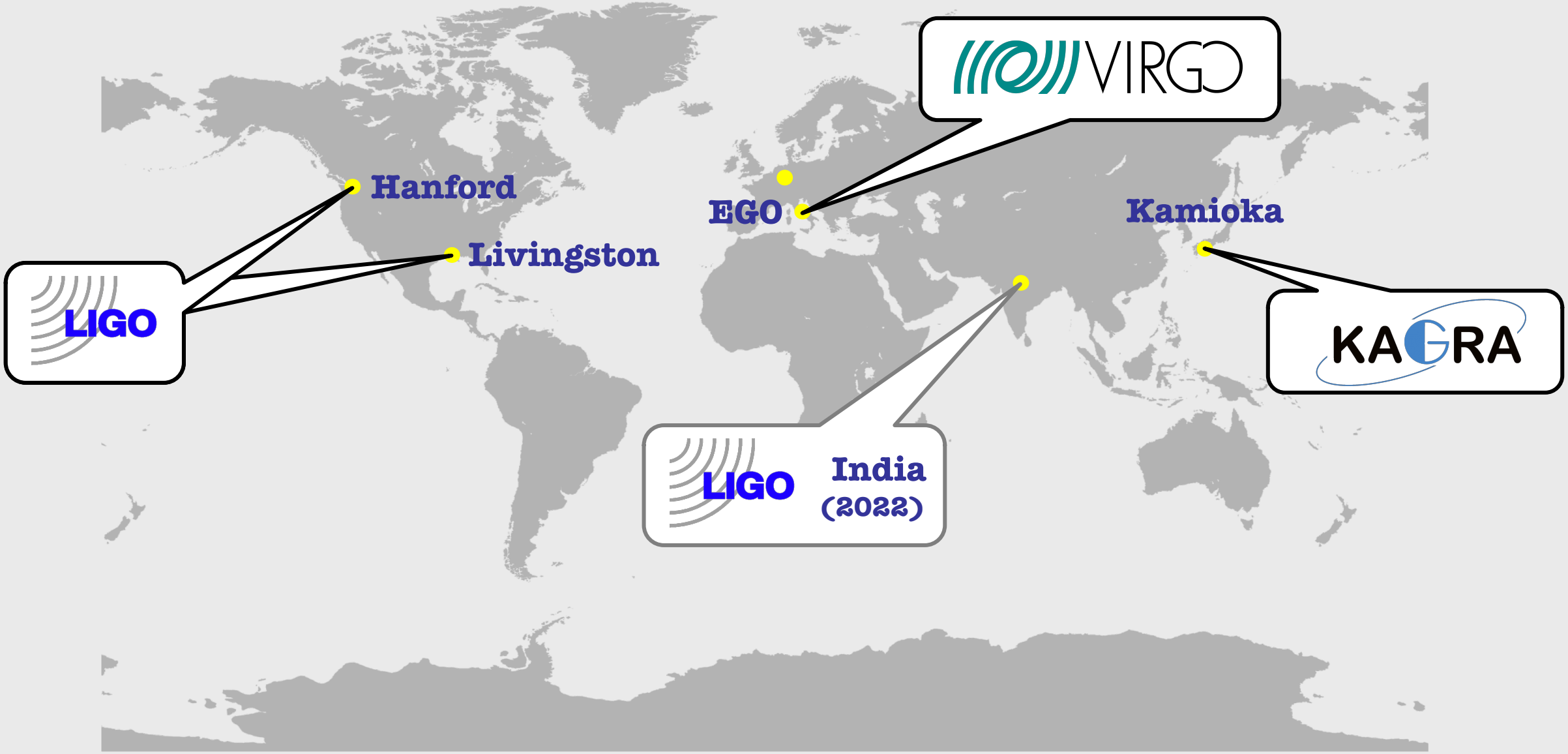
VIRGO AND GRAVITATIONAL WAVES COMPUTING IN EUROPE

Stefano Bagnasco | INFN Torino

WORLDWIDE NETWORK



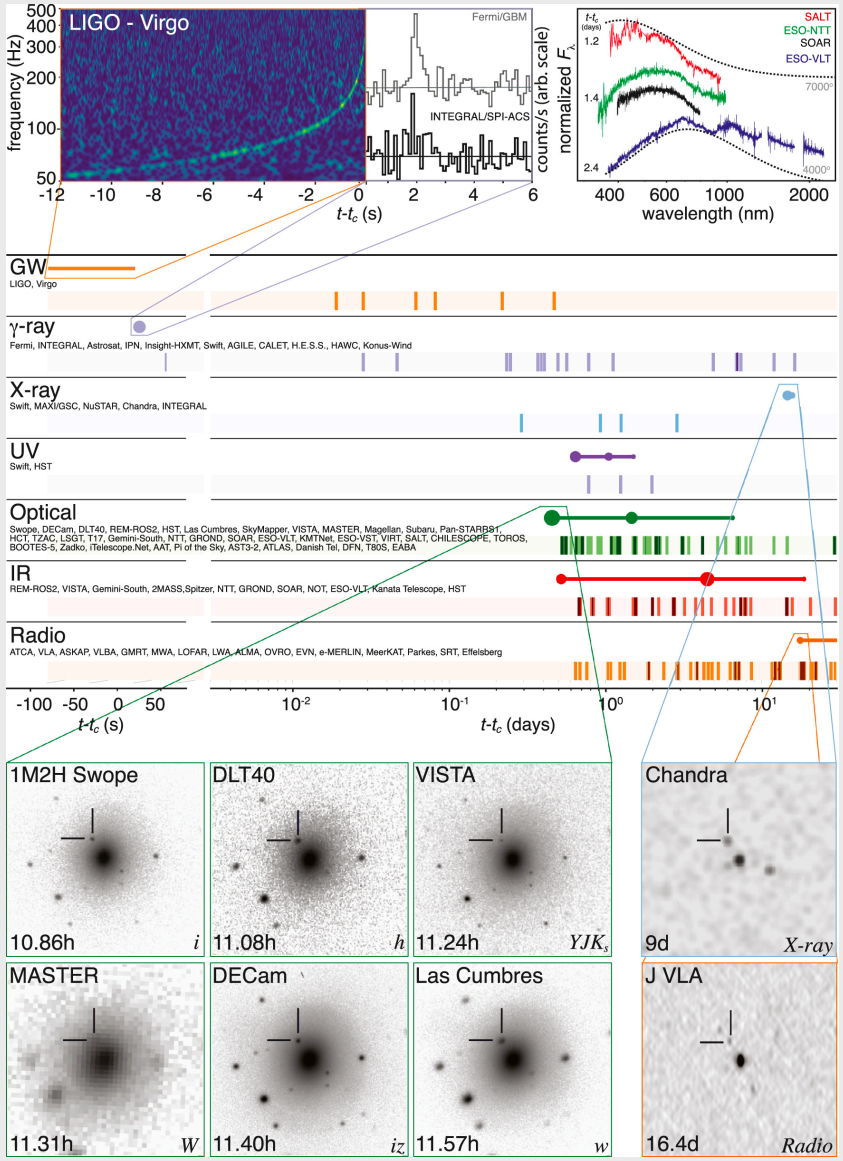
Burst sources:

- Compact Binary Coalescence
  - Coalescing Compact Binary Systems (Neutron Star- NS, Black Hole-NS, BH-BH): Strong emitters, well modelled
- Unmodeled transient bursts
  - Asymmetric Core Collapse Supernovae: weak emitters, not well-modelled ('bursts'), transient
  - Cosmic strings, soft gamma repeaters, pulsar glitches

Continuous sources:

- Continuous stochastic background
  - Cosmological stochastic background (residue of the Big Bang, cosmic GW background, long duration)
  - Astrophysical stochastic background
- Continuous waves
  - Spinning neutron stars (known waveform, long/continuous duration)

MULTIMESSENGER ASTRONOMY



"Multi-messenger Observations of a Binary Neutron Star Merger"
B. P. Abbott *et al.* 2017 *ApJL* 848 L12
doi:10.3847/2041-8213/aa91c9

LOW LATENCY SEARCHES

Low-latency searches

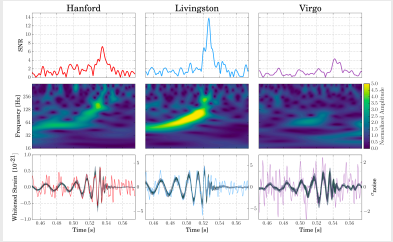
Detector sanity, Data Quality, localization,...

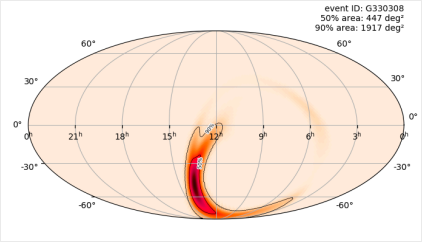
Event validation

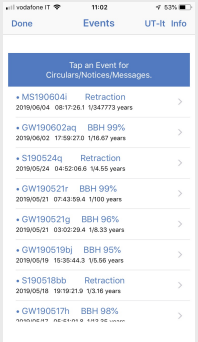
GW candidate

Sky localization

Public Alert







On-site

A few minutes

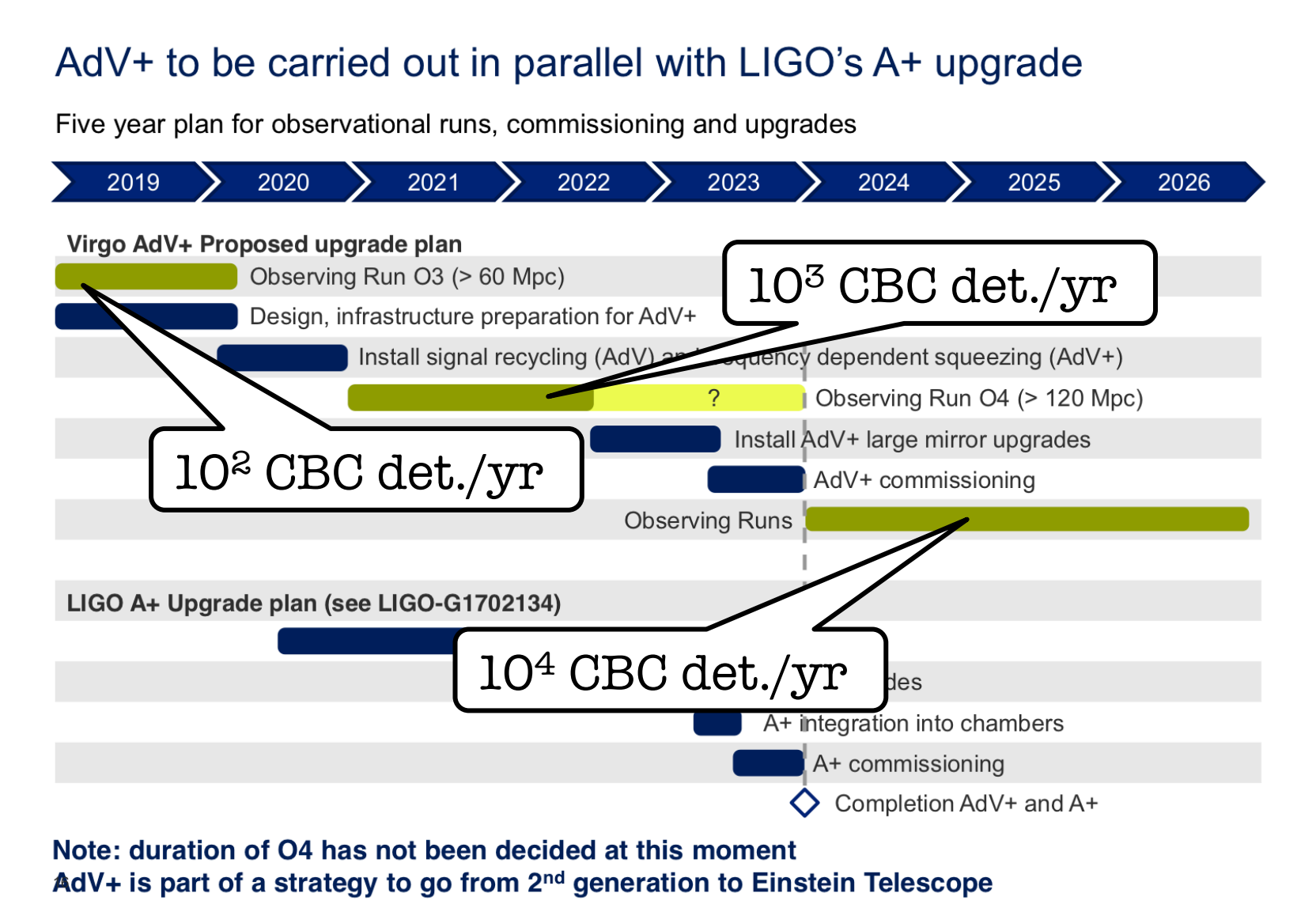
½ hour

Off-site

- Parameter estimation
- GW Candidate Update

Hours, days

LIGO - VIRGO TIMELINE



Jo van den Brand – jo@nikhef.nl

Raw Data, ~1PB/yr:

- **Full Bandwidth Raw**, not exported (7.5-10 TB/day)
- **Raw Data:** downsampled, calibrated and uncalibrated h(t) (3-4 TB/day)
- **A few levels of reduced data sets** for various uses

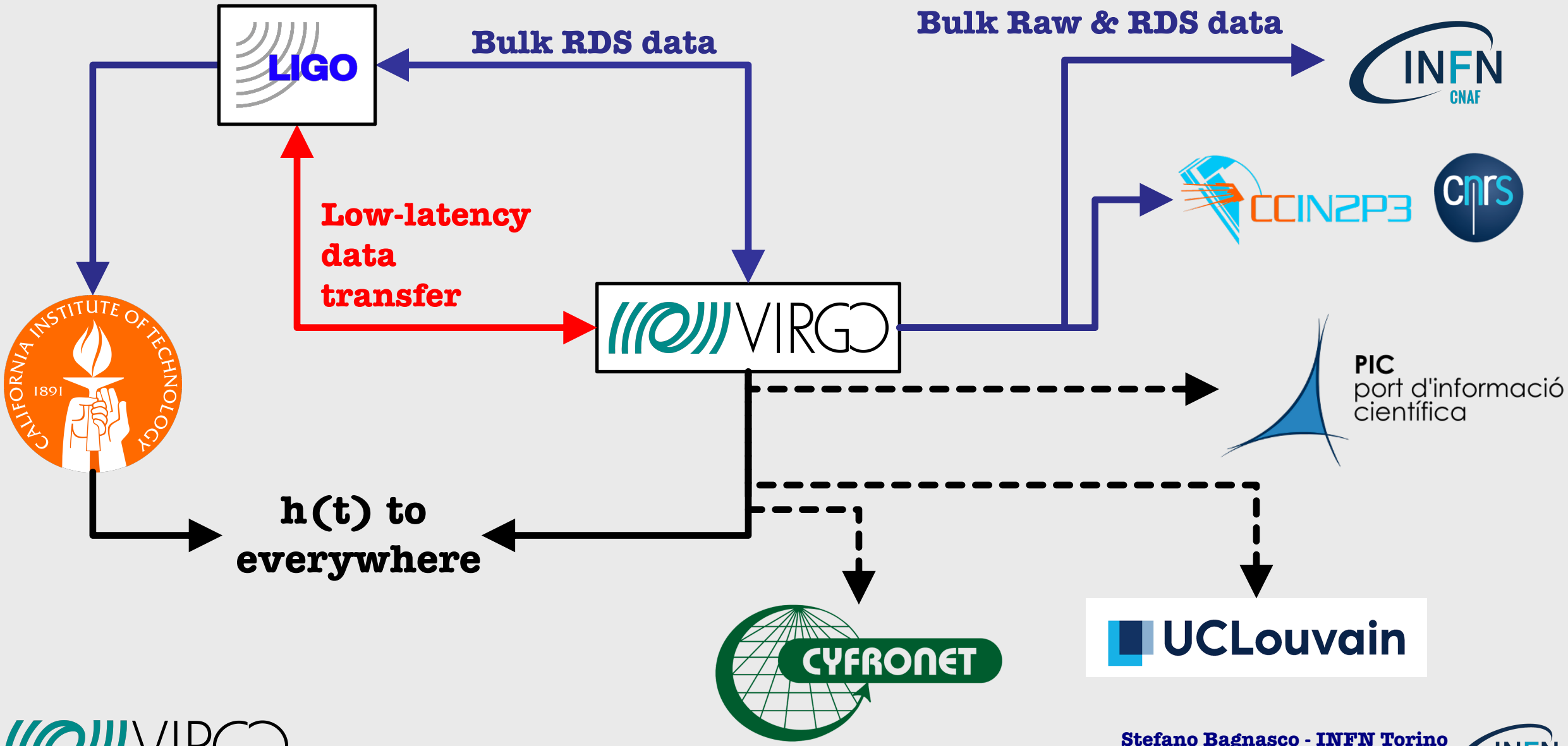
Data for physics, ~ 5TB/yr/detector:

- Virgo h(t): calibrated «strain» data
  - sampled at 10 KHz, as ~1kSec frame files
  - Includes state vector (data quality flags, vetoes,...)

$$h(t) = \frac{1}{R}\frac{2G}{c^4}\ddot{I}(t)$$

- LIGO h(t)
  - Copied online to EGO for low-latency searches and exported to CCs for offline analysis
- KAGRA h(t)
  - Coming soon...

CCS AND SIMPLIFIED DATA FLOWS



BOUNDARY CONDITIONS

- Many activities still based on legacy architectures, heterogeneous environment, custom tools that increase the management and maintenance burden on a small community
- Analysis pipelines are very heterogeneous (Matlab, Phython, C++,...) and often tightly coupled to a specific running environment (Grid submission, HTCondor cluster, shared disk,...)
- As a consequence of this, there is no clear boundary between «middleware» and analysis code proper and little or no common framework
- Single-interferometer data have no or very little scientific use

Need to define and deploy a common and sustainable GW computing environment

- Provide a uniform runtime environment for offline pipelines
- Full interoperability between Virgo, LIGO (and KAGRA)
- Provide scalability and the opportunity to exploit heterogeneous resources
- Adopt mainstream, widely used tools, leveraging upon HEP experience

Enter **IGWN** – the International Gravitational Waves observatories Network

- A coordination effort aimed at jointly discussing the computing policy, management, and architecture issues of LIGO, Virgo, and KAGRA.
- **Caveat:** this (partial) summary sees things from the Virgo point of view!

SHOPPING LIST

Bulk data transfer

Safely transfer all data to custodial storage in CCs

Software packaging and distribution

Make pipeline software available ubiquitously

Data distribution

Make h(t) data available to worker nodes anywhere

Data cataloguing and bookkeeping

Organize all data and metadata and provide querying capabilities

Workload management

Provide a uniform job submission and runtime environment

High-level workload orchestration

Keep a database of all jobs and allow the enforcement of priorities

Monitoring and accounting

Monitor distributed computing and provide reliable accounting

CVMFS AND STASHCACHE FOR DATA DISTRIBUTION

- Currently no «data catalogue» like the ones most HEP experiments use
  - Metadata encoded in PFNs
  - Data discovery through custom tool scanning filesystems
- CVMFS used to export a filesystem-like namespace and provide uniform access to h(t) data
  - Authenticated access trough OSG-developed x.509 plugin
  - «External data» feature + client plugin used to read actual bytes from…
- StashCache as Content Delivery Network
  - xCache-based cache hierarchy, http protocol
  - see e.g. I. Sfiligoi's presentation on Tuesday
  - Currently two instances in the EU: UvA (production), CNAF (under test)

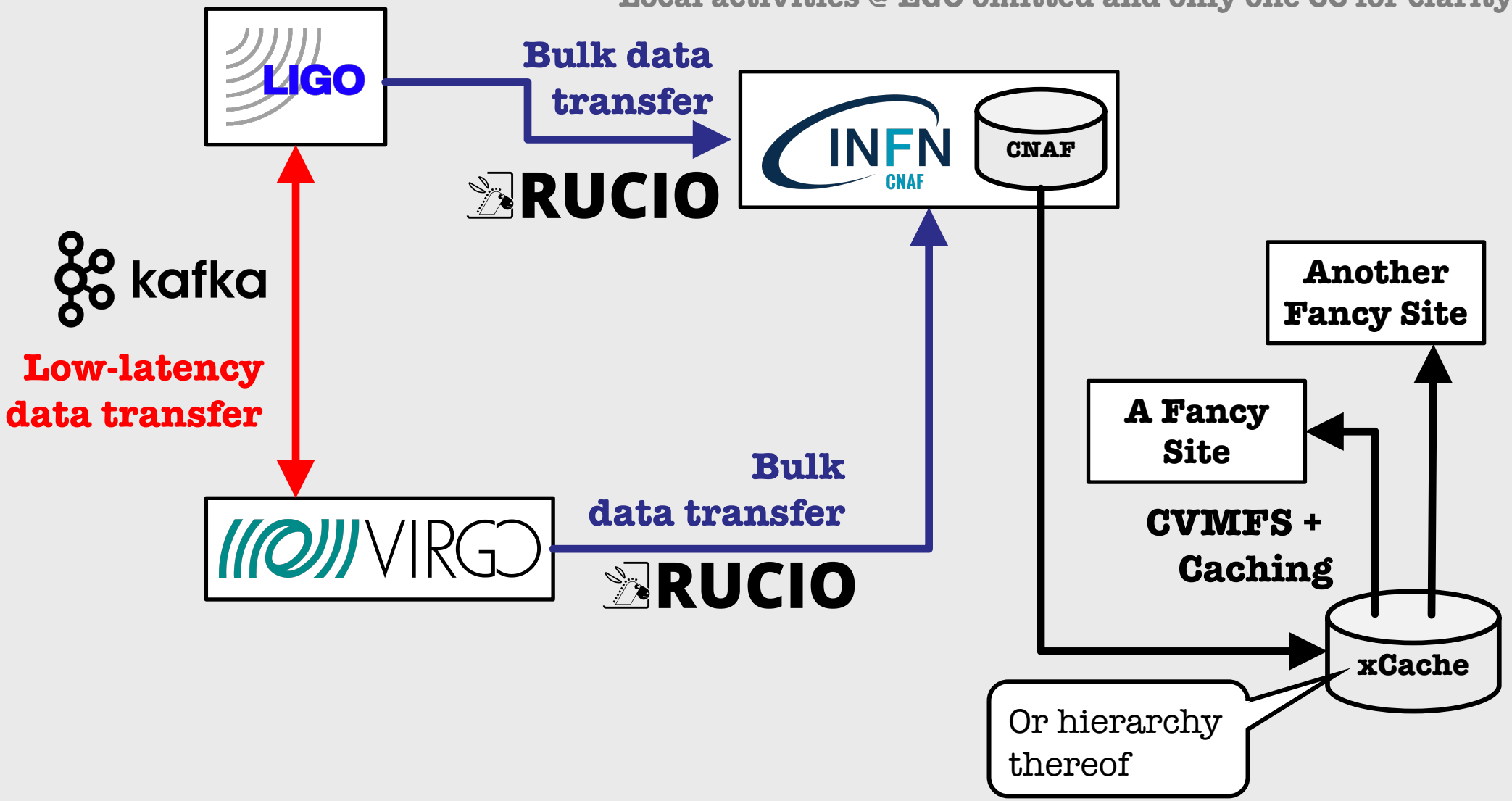
RUCIO

- **Step 1:** use Rucio to manage bulk raw data transfer to custodial storage
  - See also M. Lassnig's presentation on Monday
  - HA deployment studies
  - Multiple checksum algorithms fix

- **Step 2:** use it as a full-fledged Data Catalogue (under discussion)
  - Possibly with a federated architecture
  - Another possible development: POSIX mount to replace CVMFS (kicked off at last Rucio Code Camp at CERN)

OVERALL PICTURE: DATA

Local activities @ EGO omitted and only one CC for clarity



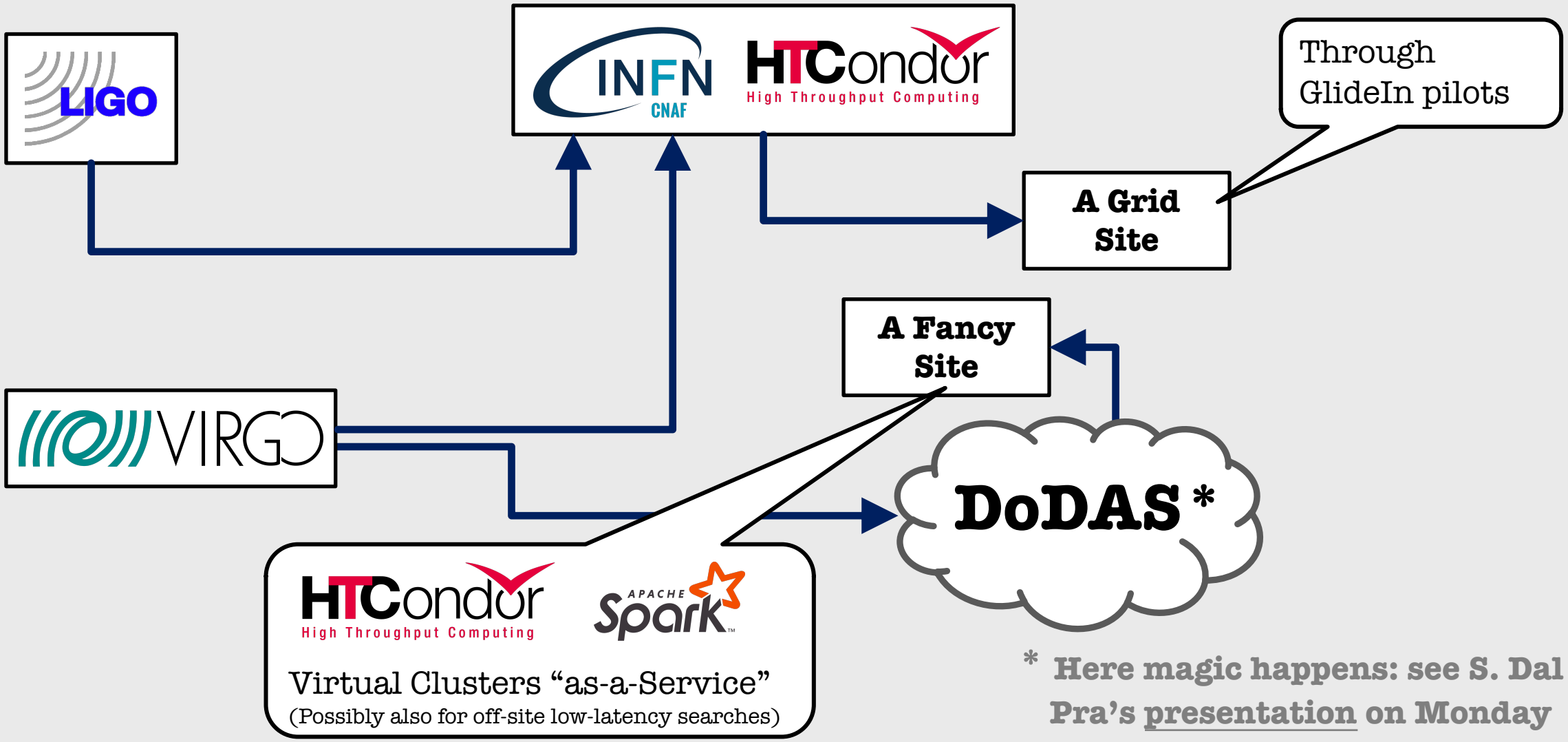
HTCondor
High Throughput Computing

Use HTCondor as a uniform job submission layer

- To local HTCondor clusters if needed
  - Some pipelines already rely on HTCondor DAGs for dependencies
- To common Condor Pools
  - Exploit Grid resources through GlideIn pilots
- To on-demand HTCondor Virtual Clusters
  - Through DoDAS Distributed on-Demand Analysis Services
  - Exploit heterogeneous or opportunistic resources
  - Cater to special needs like off-site low-latency searches

OVERALL PICTURE: JOBS

Local activities @ EGO omitted and only one CC for clarity



* Here magic happens: see S. Dal Pra's presentation on Monday

WHERE ARE WE?

Bulk data transfer

Rucio

Software packaging and distribution

CVMFS (+ Conda + Singularity)

Data distribution

CVMFS (but see above) + StashCache

Workload management

HTCondor and, possibly, DoDAS

Mostly decided except implementation details, under deployment or active development

Up next?

Data cataloguing and bookkeeping

Probably Rucio again

High-level workload orchestration

Possibly DIRAC, custom tools if it's an overkill

Monitoring and accounting

TBD, maybe DIRAC again?

Not an immediate priority

CONCLUSIONS

- GW computing (in EU and elsewhere) is making a significant step forward towards organized computing
- IGWN means a coordinated and common infrastructure between the collaborations
- The adoption of common mainstream solutions like HTCondor or Rucio also means GW joins the larger physics computing community
- ...which bodes well for the proposed 3G interferometers computing effort in a decade from now.