AdvancedVirgo e il calcolo: stato e prospettive

Stefano Bagnasco | INFN Torino

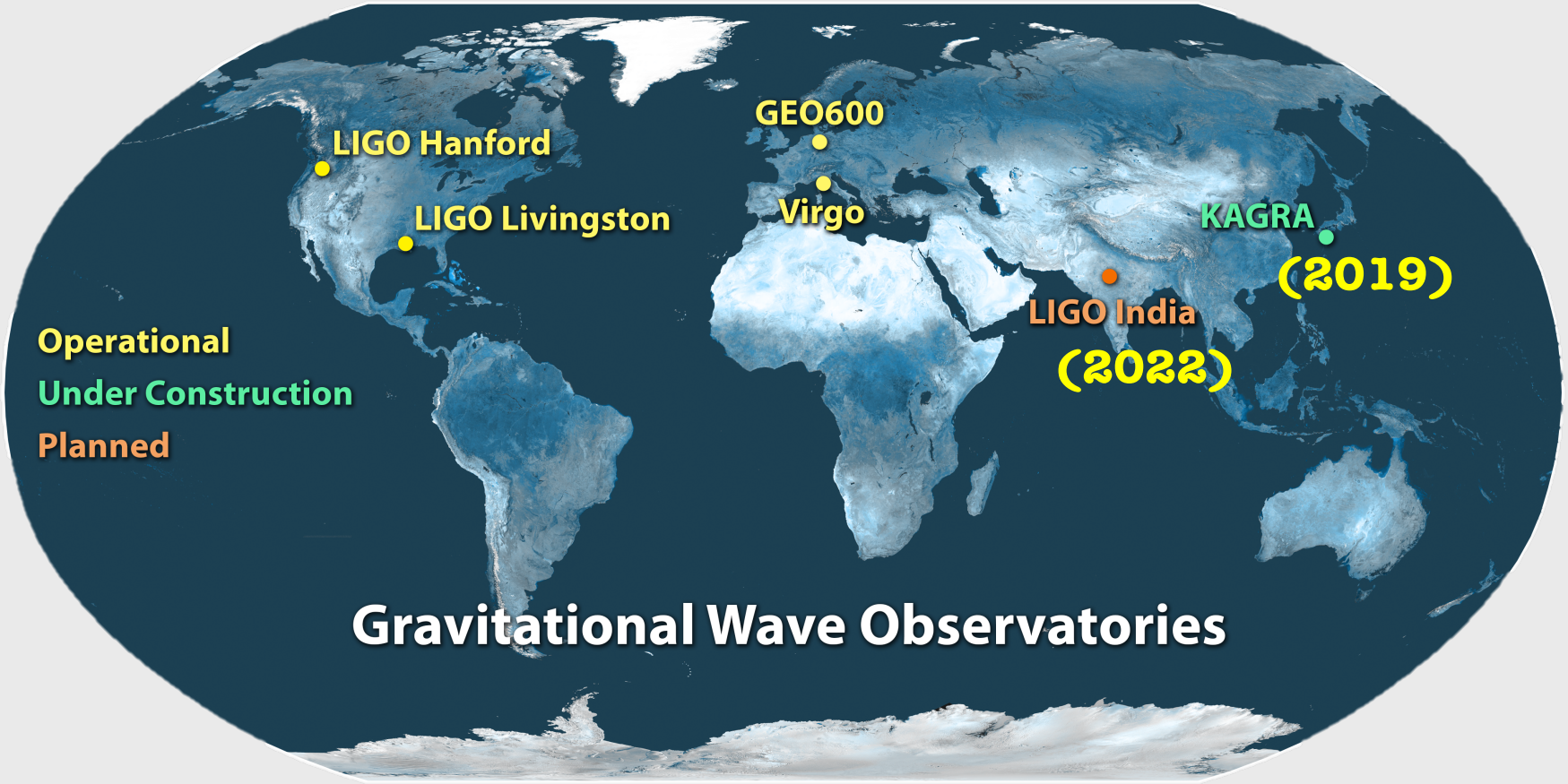
OUTLINE

- Recap on GW, AdV & aLIGO
- What happened since Michele's talk in Rimini?
- Data flows and processing
- Computing model revision
- AoB

Biases & caveats:

- Talk from a VIRGO newbie
- Personal bias towards distributed computing infrastructure
- So some activities left out (feel free to ask)
- Contributions from several people (see last slide)

WORLDWIDE NETWORK



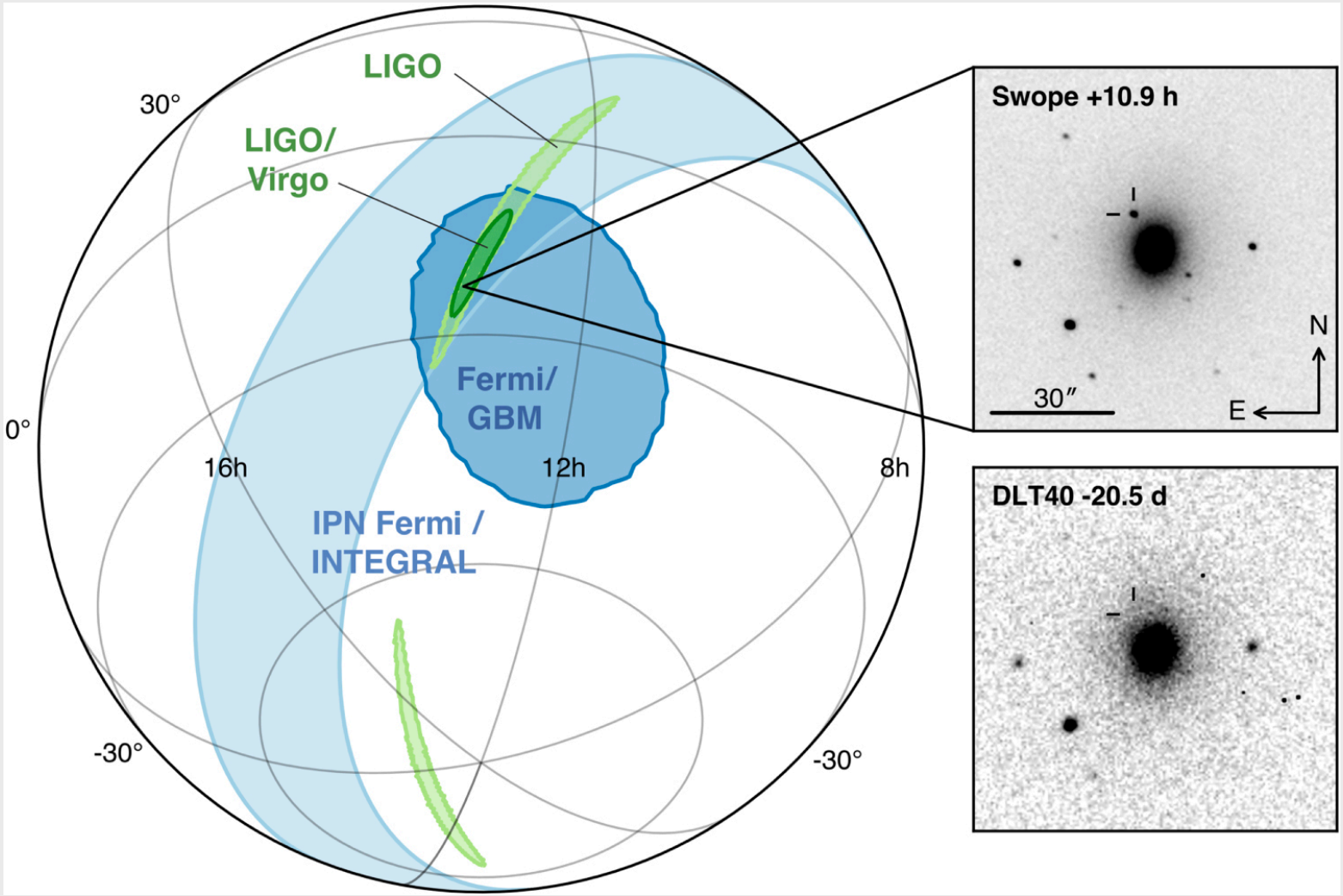
Burst sources:

- 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:

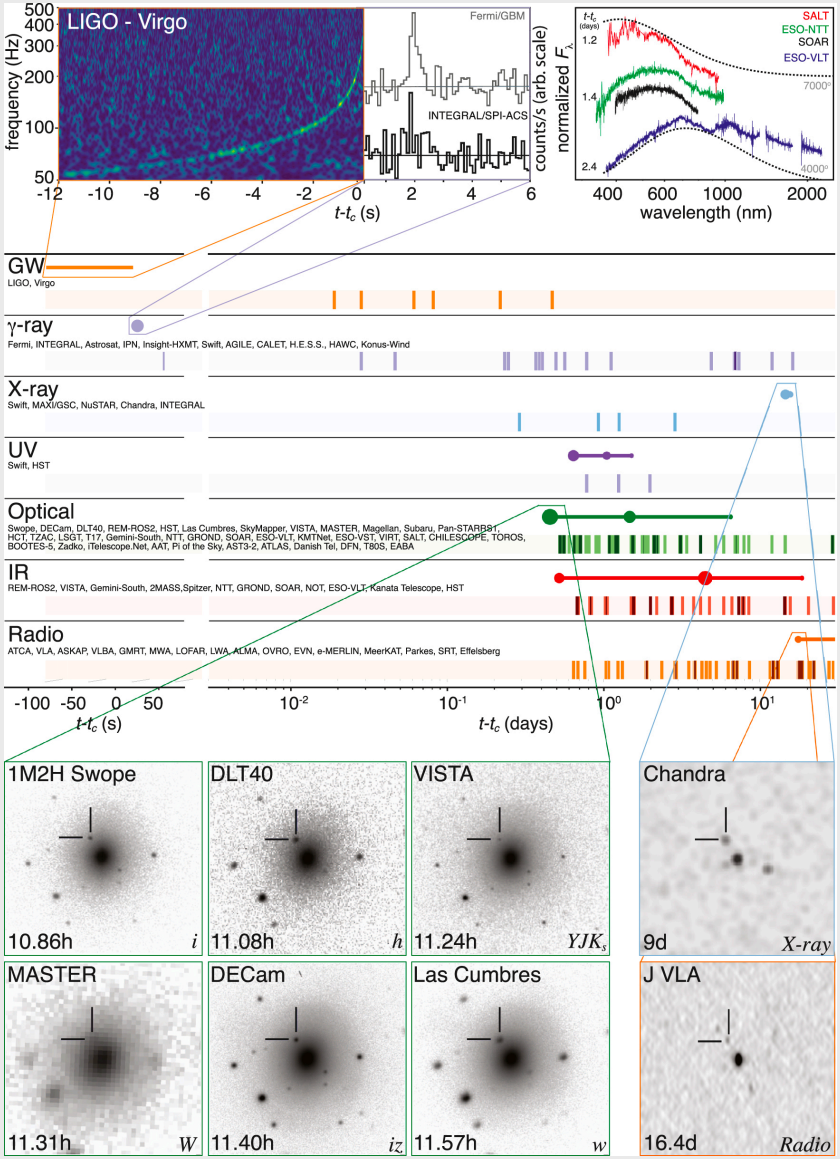
- Continuous stochastic background
  - Cosmological stochastic background (residue of the Big Bang, cosmic GW background, long duration)
  - Astrophysical stochastic background
- Continuous waves
  - Spinning neutron stars (known waveform, long/continuous duration)

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

MULTIMESSENGER ASTRONOMY



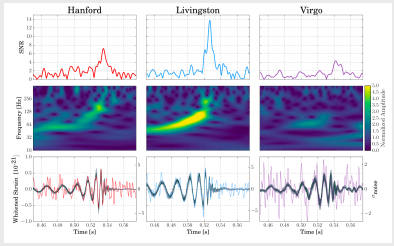
"Multi-messenger Observations of a Binary Neutron Star Merger"
B. P. Abbott *et al.* 2017 *ApJL* 848 L12
doi:10.3847/2041-8213/aa91c9

TRANSIENT SEARCHES

Low-latency searches

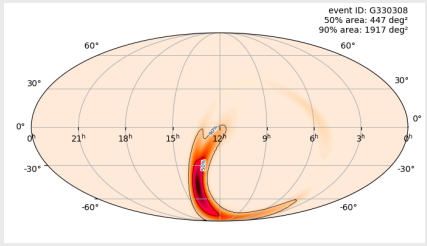
Detector sanity, Data Quality, localization,...

GW candidate

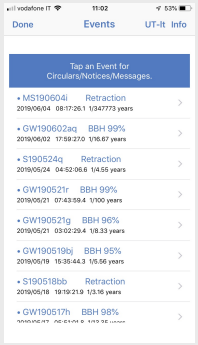


Event validation

Sky localization



Public Alert



A few minutes

½ hour

Hours, days

- Parameter estimation
- GW Candidate Update

PUBLIC ALERTS

GraceDb | Latest

https://gracedb.ligo.org/latest/

GraceDB — Gravitational Wave Candidate Event Database

HOME | SEARCH | CREATE | REPORTS | RSS | LATEST | ALERTS | DOCUMENTATION | LOGOUT

AUTHENTICATED AS: STEFANO BAGNASCO

Latest — as of 3 June 2019 16:57:09 UTC

Test and MDC events and superevents are not included in the search results by default; see the query help fo
and superevents in those categories.

Query:

Search for: Superevent

Search

UID	Labels	Preferred Event	GW Events	t_start	t_0	
S190603be	DQOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY	G335137	G335137	1243615414.234300	1243615415.234300	124
S190603bd	DQOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY	G335136	G335136	1243615164.168098	1243615165.168098	124
S190603bc	DQOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY	G335132	G335134, G335133, G335132	1243614786.923033	1243614787.924558	124
S190603bb	DQOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY	G335130	G335131, G335130	1243614350.439874	1243614351.439874	124
S190603ba	DQOK SKYMAP_READY PASTRO_READY EMBRIGHT_READY	G335129	G335129	1243614344.811259	1243614345.811259	124
S190603az	DQOK SKYMAP_READY PASTRO_READY EMBRIGHT_READY	G335128	G335128	1243614223.542193	1243614224.542193	124

GCN Circulars Archive

https://gcn.gsfc.nasa.gov/gcn3_archive.html#...

GCN Circulars Archive

This page changes after each Circular submission, so hit the <reload> button NOW.

1. The Latest Circulars
2. Older Circulars
3. Tarfile of all Circulars
4. Circulars grouped by GRB/transient

This is the archive of the GCN Observation Report Circulars.

To learn more about the GCN Circulars and how to subscribe. and to unsubscribe here.

(Listed in reverse serial number order -- newest first)

- 24729 GRB 190530A: R-band observation from HCT
- 24728 LIGO/Virgo S190602aq: No counterpart candidates in INTEGRAL SPI-ACS and IBIS prompt observation
- 24727 LIGO/Virgo S190602aq: upper limit in INTEGRAL SPI-ACS prompt observations
- 24726 LIGO/Virgo S190602aq: No counterpart candidates in Fermi-LAT observations
- 24725 LIGO/Virgo S190602aq: No counterpart candidates in the Swift/BAT Observations
- 24724 Subject: LIGO/Virgo S190602aq: Upper limits from Fermi-GBM Observations
- 24723 LIGO/Virgo S190602aq: Upper limits from AGILE-GRID observations
- 24722 LIGO/Virgo S190602aq: No counterpart candidate in AGILE-MCAL observations
- 24721 LIGO/Virgo S190602aq Global MASTER-Net observations report
- 24720 LIGO/Virgo S190602aq: Coverage and upper limits from MAXI/GSC observations
- 24719 LIGO/Virgo S190602aq: No neutrino counterpart candidates in ANTARES search
- 24718 LIGO/Virgo S190602aq: No counterpart candidates in HAWC observations
- 24717 LIGO/Virgo S190602aq: Identification of a GW compact binary merger candidate
- 24716 LIGO/Virgo S190602aq: Upper limits from IceCube neutrino searches
- 24715 Konus-Wind observation of extremely bright GRB 190530A
- 24714 GRB 190530A: Insight-HXMT/HE detection
- 24713 GRB 190531B: Insight-HXMT/HE detection
- 24712 GRB 190530A: Mondy and AbAO optical observations
- 24711 GRB 190531B: VLT optical upper limit
- 24710 Fermi trigger No 581068049 Global MASTER-Net observations report
- 24709 GRB 190530A: Konkoly optical observations of the afterglow

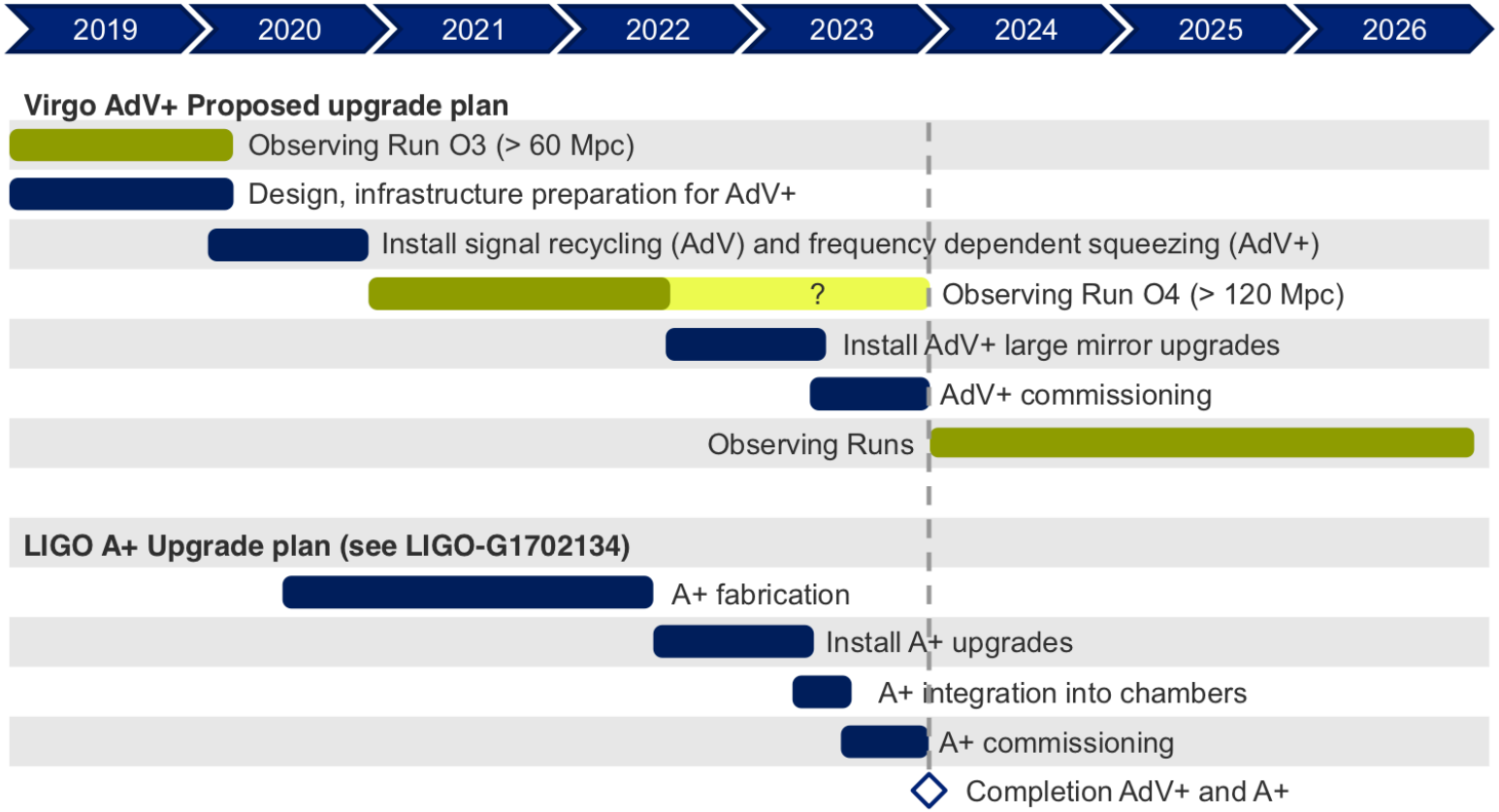
Third Observational Period

- Started April 1, 2019
  - 1 year + possible extensions
  - Detection rate $\gtrsim$1/week (expected: 10^2/year)

LIGO - VIRGO TIMELINE

AdV+ to be carried out in parallel with LIGO's A+ upgrade

Five year plan for observational runs, commissioning and upgrades



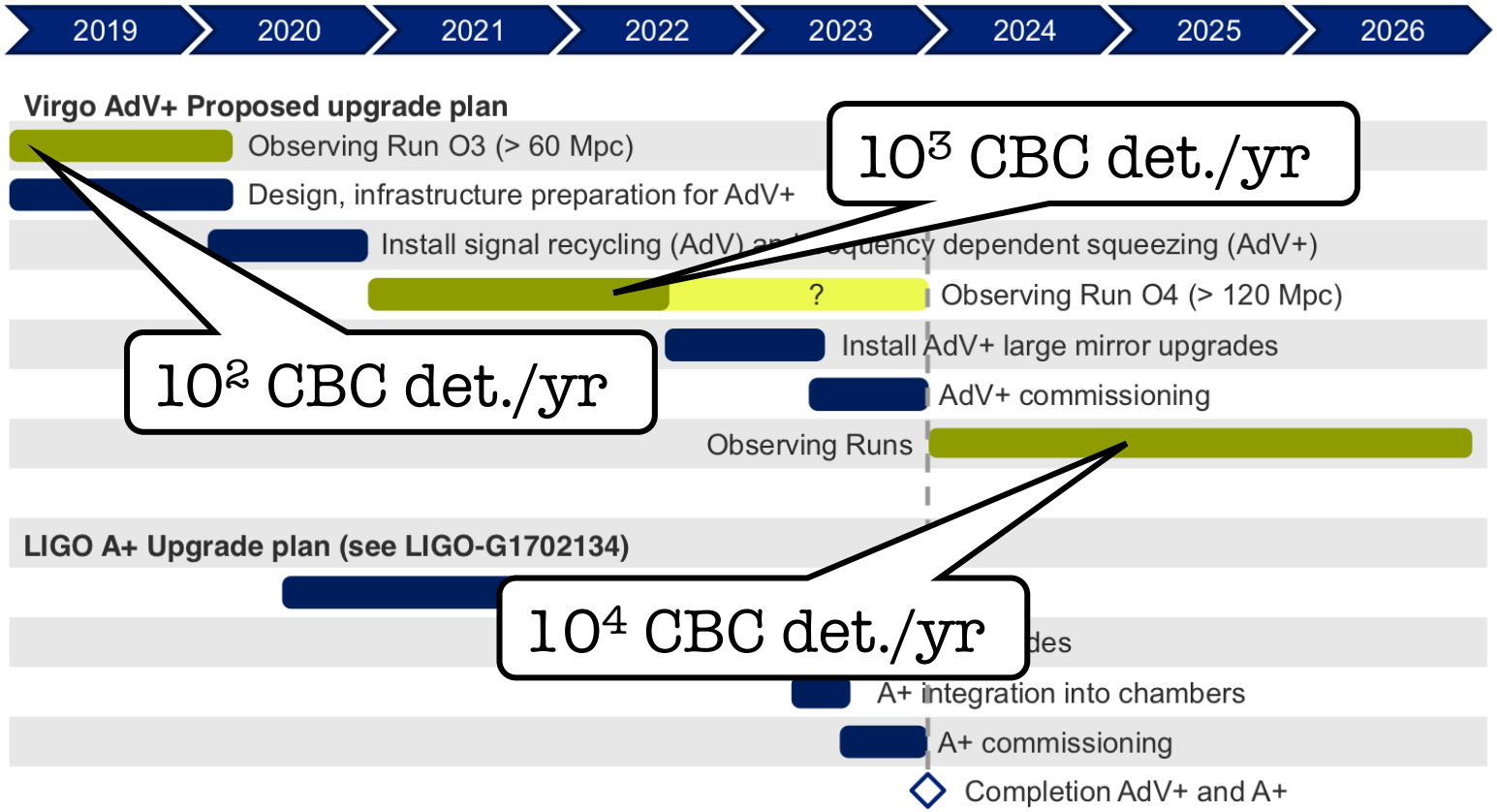
Note: duration of O4 has not been decided at this moment
AdV+ is part of a strategy to go from 2nd generation to Einstein Telescope

Jo van den Brand – jo@nikhef.nl

LIGO - VIRGO TIMELINE

AdV+ to be carried out in parallel with LIGO's A+ upgrade

Five year plan for observational runs, commissioning and upgrades



Note: duration of O4 has not been decided at this moment
AdV+ is part of a strategy to go from 2nd generation to Einstein Telescope

Jo van den Brand – jo@nikhef.nl

«Raw Data», ~1PB/yr:

- **Full Bandwidth Raw** (7.5-10 TB/day) are not exported
  - 7 days (40 TB) buffer @ EGO
- **Raw Data** (downsampled) includes calibrated and uncalibrated h(t) (3-4 TB/day)
  - 12 months circular buffer @ EGO
  - copied to Tier-1s tapes for significant runs* (and a small subset of commissioning runs)
- **Reduced Data Sets** (RDS) are highly downsampled raw data for DetChar and data analysis (180GB/day)
  - Stored in the EGO buffer for one year and permanently copied to at least two CCs
- **aLIGO RDS** is the same from aLIGO (60 GB/day)
  - Directly copied to CCs
- **Trend data** are 1-sec & 1-min min, max, mean and RMS values of all channels, 5-6 GB/day
  - Stored in the EGO circular buffer and copied to all the AdV CCs permanently.

* These include science, astrowatch and calibration runs

Processed data for physics:

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

- AdV h(t)
  - sampled at 10 KHz, as ~1kSec frame files
  - Includes state vector (data quality flags, vetoes,...)
  - Stored in buffer @ EGO for low-latency searches and exported to all CCs for offline analysis for significant runs
- aLIGO h(t)
  - Copied online to EGO for low-latency searches and exported to CCs for offline analysis
- Mock Data Challenge h(t)
  - 20 TB/yr of simulated data

Please note:

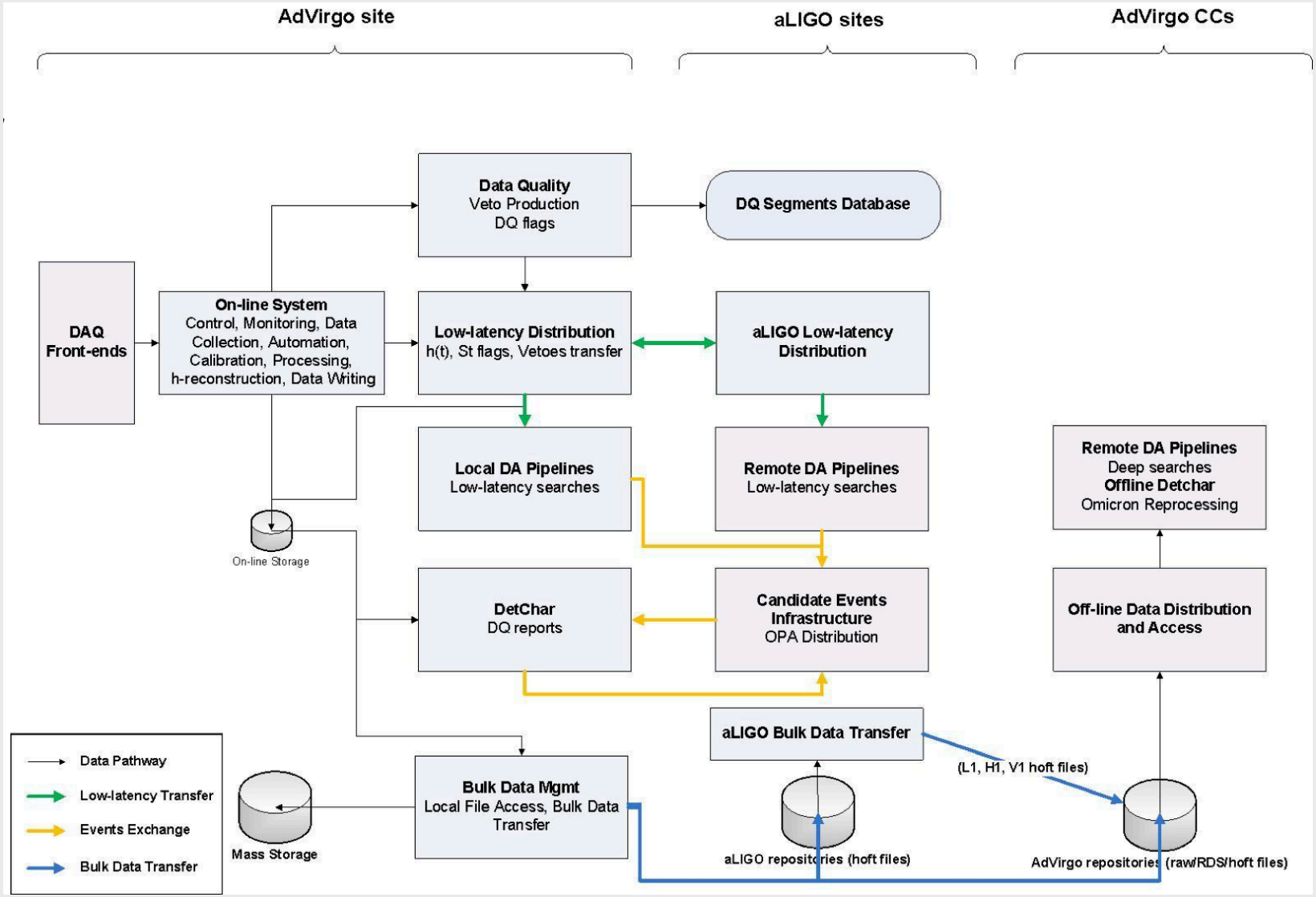
- Single-interferometer data have little scientific use
- Some research groups started asking for extra channels (e.g. magnetometers)

Additional space can be needed for interesting segments of data to be stored for longer time.

O3 DATA SIZES

Data	Data flow [GB/day]	Target buffer length @ EGO	Buffer @ EGO [TB]	Offline storage [TB per year]
Full Bandwidth	7500	~ 1 week	41	–
Raw data	3000	1 year	~ 1000	745
AdV RDS	180	1	11	11
aLIGO RDS	60	0	0	22
aLIGO h(t) and status flags	15	0	0	6
Trend data	5	3	5	1.5
AdV h(t) and status flags	10	1	3	3
aLIGO h(t) and status flags	15	0.5	3	–
MDC h(t)	–	–	–	20

DATAFLOW SCHEMA



Franco.Carbognani@ego-gw.it

O3 ONLINE ACTIVITIES

- DetChar and Low Latency searches
  - Running on plain VMs or through HTCondor
- Bulk data transfer
  - Legacy solution based on GFAL2 (CNAF) and iRODS (CCIN2P3) + custom scripts
- h(t) files transfer
  - Legacy solution based on LDR (Ligo Data Replicator)
- Low latency data distribution
  - Legacy solution based on Cm (custom tool)

✔ Everything is working reasonably well, no immediate problems for O3, but...

O3 OFFLINE HIGHLIGHTS

- New MoU requires AdV to provide ~25% of LVC computing resources
  - Progresses, but still not reached
- Mostly CW pipelines running offline
  - Both from AdV and aLigo groups
- Derived data expectations for O3 (3 detectors, 1 year observation): 28 TB

✔ In general, groups are able to run satisfactorily enough, but...

...BUT:

- Legacy architectures, heterogeneous environment, custom tools increase the management and maintenance burden
- Analysis pipelines are very heterogeneous and often tightly coupled to a specific running environment (Grid submission, HTCondor cluster, shared disk,...)
- No clear boundary between «middleware» and analysis code
- (Almost) no common framework
- Resource contention is likely to happen at some point in O3; tools for prioritization, sharing and accounting will be needed

WAY FORWARD 1

- New groups joining Virgo with computing infrastructure expertise
  - INFN-Torino, ICC Universitat de Barcelona, Louvain Tier-2
- Several activities ongoing
  - See next slides
  - **Disclaimer:** the following are to be considered evaluation activities. No final decision by the collaboration is yet taken about tools or architectures beyond O3.
  - Intense coordination activity

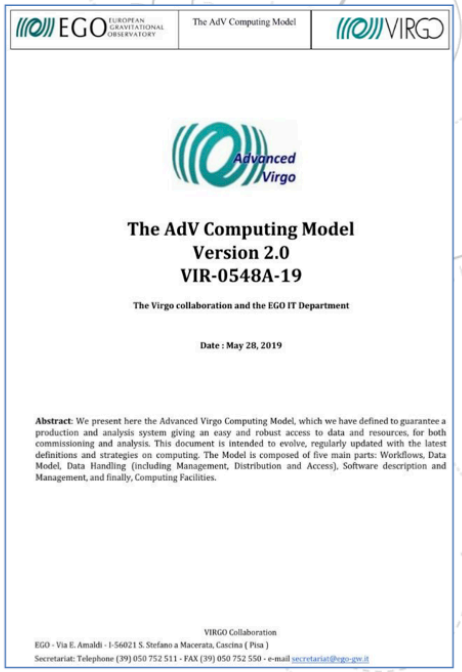
WAY FORWARD 2

The collaboration is testing a number of technologies to evolve the computing infrastructure towards O4 and AdV+

- Provide a uniform runtime environment for offline pipelines
- Help pipelines adopt common submission and data access tools
- Interoperability with aLigo (and aLigo+)
- Use (a small number of) mainstream, widely used tools to reduce support burden
- Data size increase not a big technical issue
- In perspective, try to transparently run on heterogeneous resources (OSG, wLCG-like, HPC, dedicated, opportunistic, whatever)

Recent milestones

- **DIRAC workshop** in Virgo Week (Cascina, Nov'18)
- **AdV Computing Kick-Off Meeting** (Barcelona, Feb'19), VIR-0258A-19
  - Further meetings will follow
- **Computing Model revision**, VIR-0548A-19
  - Low-latency & offline pipelines (the ones *actually* being used, and with an updated description)
  - Online systems, DetChar, DataQuality...
  - Computing resources
  - Data management
  - Software management
  - Outlook and planned improvements
- Some significant technical milestones (see later)


EGO EUROPEAN GRAVITATIONAL OBSERVATORY | The AdV Computing Model | VIRGO

Advanced Virgo

The AdV Computing Model
Version 2.0
VIR-0548A-19

The Virgo collaboration and the EGO IT Department

Date : May 28, 2019

Abstract: We present here the Advanced Virgo Computing Model, which we have defined to guarantee a production and analysis system giving an easy and robust access to data and resources, for both commissioning and analysis. This document is intended to evolve, regularly updated with the latest definitions and strategies on computing. The Model is composed of five main parts: Workflows, Data Model, Data Handling (including Management, Distribution and Access), Software description and Management, and finally, Computing Facilities.

VIRGO Collaboration
EGO - Via E. Amaldi - I-56021 S. Stefano a Macerata, Cascina (Pisa)
Secretariat: Telephone (39) 050 752 511 - FAX (39) 050 752 550 - e-mail secretariat@ego-gw.it

Jordi Portell i de Mora – jportell@fqa.ub.es


RUCIO for organized bulk data transfer

Uniform runtime environment in US centres:

- HTCondor for workload management
- CVMFS for software and data namespace distribution
- StashCache (xcache) for distribution of both aLigo and AdV h(t) data

https://arxiv.org/pdf/1705.06202.pdf

CVMFS AND XCACHE FOR DATA DISTRIBUTION

- There is currently no «file catalogue» like the ones most HEP experiments use
- CVMFS can be used to export a filesystem-like namespace and provide uniform access
  - Authenticated access trough OSG-developed x.509 plugin
  - «External data» feature + client plugin used to access data
  - xCache hierarchy provide data distribution

- StashCache instance exists at UvA
  - So CVMFS data access works at CNAF (but poor performance)
  - Another StashCache instance at CNAF, or...
  - ...reuse work by D. Ciangottini (CMS, see Diego's talk)
  - Under investigation at CNAF (L. Morganti et al.)

RUCIO FOR BULK DATA TRANSFER

- Data management tool originally from ATLAS
  - (you heard about it)
  - Rucio provides policy-based replication and interfaces to most storage systems
  - Already chosen by aLIGO
  - Test activities already undergoing for transfers to CNAF and CCIN2P3
  - Fixing minor issues (e.g. checksums for different GridFTP implementations)
- Also, it can be used as a fully-featured Data Catalogue!

VIRGO

HTCONDOR FOR WORKLOAD MANAGEMENT

- HTCondor is gaining momentum in Grid sites
  - CNAF for one!
  - And is the standard running environment in aLIGO sites
- Some pipelines currently depend on HTCondor
  - Difficult to decouple actual physics analysis code from pipeline management (that inlcudes HTCondor API calls)
- Obvious strategy would be to “use HTCondor for all submission”
  - To provide uniform job submission and management
  - To minimize pipeline migration effort
- However...

HTCONDOR FOR WORKLOAD MANAGEMENT

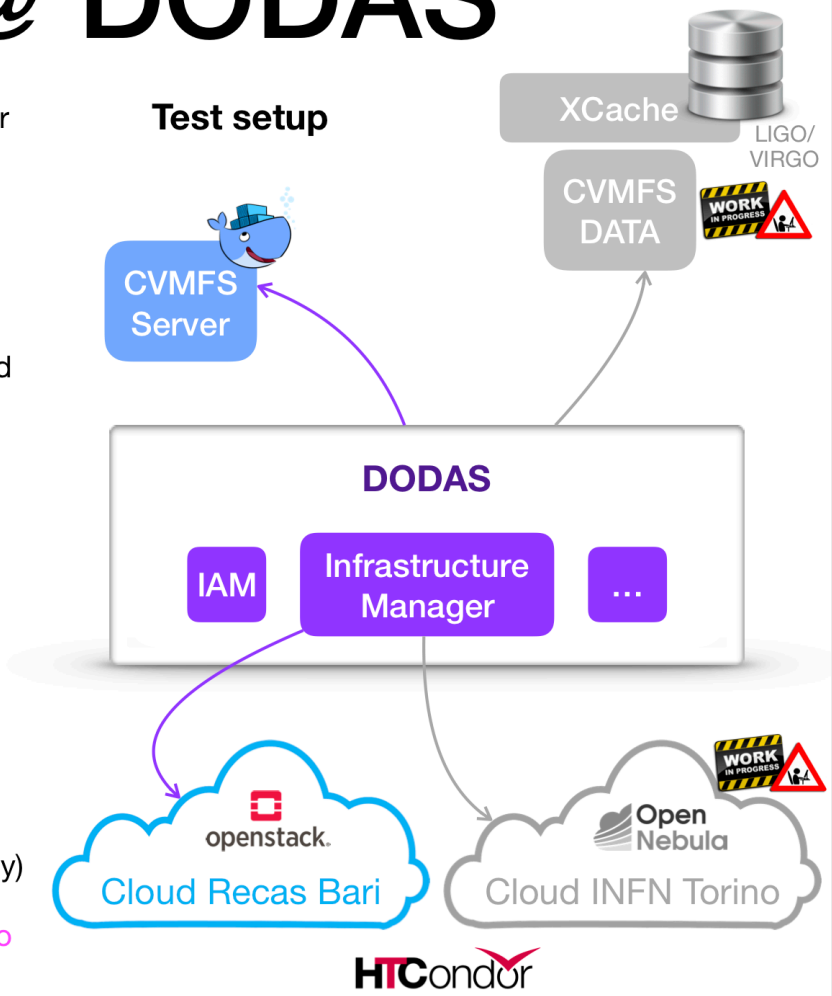
- However, HTCondor can be many things...

  - a cluster's local workload management and submission system (the cluster may well be virtual);
  - a distributed computing platform for a network of sites, via pooling and flocking mechanisms;
  - a grid site's underlying LRMS;
  - a component of a Grid site's CE implementation, the CondorCE;
  - a linear combination of the above.

- AdV strategy is still not completely clear
  - And some ML pipelines use Apache Spark
  - But please see next slide

Virgo @ DODAS

- Configuration of a DODAS HTCondor cluster for Virgo workloads is being defined
- **What we have so far:**
  - Test cluster deployed at Recas Bari
  - Software packaged in Singularity images and distributed via CVMFS
  - *Dockerized* CVMFS server (stratum0/1) deployed in Torino
  - Workflows with auto-generated data being tested
- **What is coming:**
  - Data access through custom CVMFS acting as file catalog to remote data @ CNAF
  - Cache layer @ CNAF (see Virgo talk on Friday)
  - Scale out the HTCondor cluster on the Torino OpenNebula Cloud



Sara.Vallero@to.infn.it

AND DIRAC?

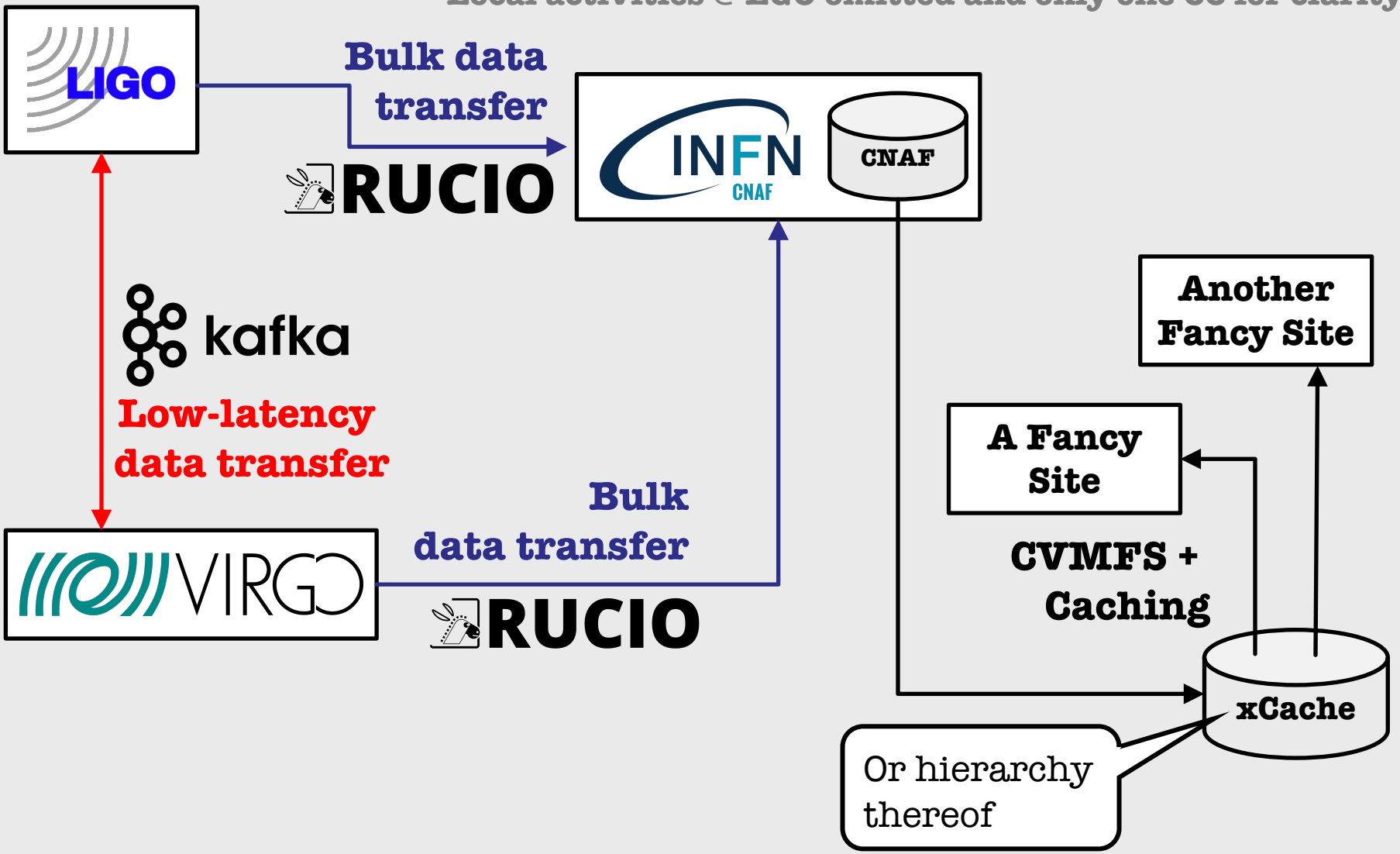
- No comprehensive workload management platform/framework
  - DIRAC a good candidate, but not an immediate priority
- DIRAC was a good candidate also for data management one year ago
  - However, aLIGO chose Rucio (which is a good solution anyway)
  - Several ongoing activities to use Rucio and DIRAC together
  - e.g. see talks by Belle2 and RAL at last DIRAC workshop [https://indico.cern.ch/event/756635]
  - Preliminary exploration in Virgo (G. Fronzé, L. Rei)

OTHER ONGOING ACTIVITIES

- CVMFS infrastructure for code distribution
  - Stratum-0 and Stratum-1 @ EGO, Stratum-1 @ CNAF?
- Apache KAFKA for low-latency data transfer
  - Replacing legacy tools
- Git + CONDA + cmake for code and build management
  - Replacing SVN + CMT
- Docker/Singularity evaluation
  - For a number of uses
- ...

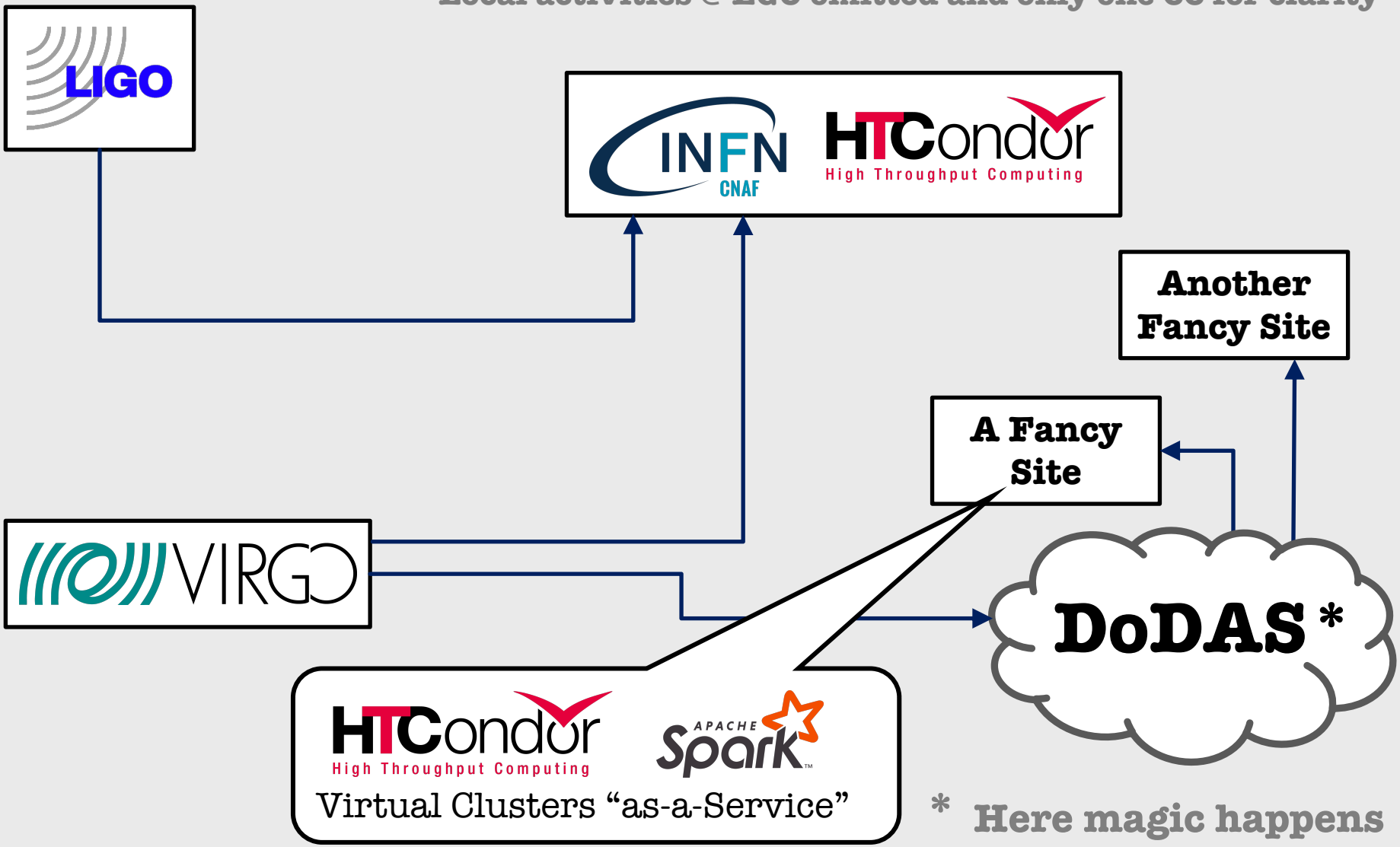
A POSSIBLE OVERALL PICTURE: DATA

Local activities @ EGO omitted and only one CC for clarity



A POSSIBLE OVERALL PICTURE: JOBS

Local activities @ EGO omitted and only one CC for clarity



GW OPEN DATA

- Relatively simple data
  - 10 kHz time series + data quality metadata
- Dataset not huge
- Very strong case for interoperability
  - Network of interferometers
  - Multimessenger astronomy
- NSF requirement for LIGO
  - Also Virgo had to comply
- Policy in place
  - Release after 24 months latency

GW OPEN DATA

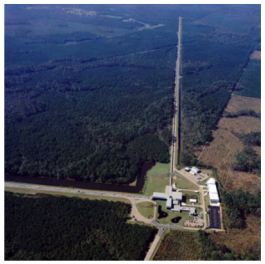
INFN Identity Check | ALICE Offline week (5-7 Decen | Gravitational Wave Open Scien

https://www.gw-openscience.org/about/

Gravitational Wave Open Science Center

- Getting Started
- Data
  - Catalogs
  - Bulk Data
- Tutorials
- Software
- Detector Status
- Timelines
- My Sources
- GPS ↔ UTC
- About the detectors
- Projects
- Acknowledge GWOSC

LIGO Hanford Observatory, Washington
(Credits: C. Gray)



LIGO Livingston Observatory, Louisiana
(Credits: J. Giaime)

Virgo detector, Italy
(Credits: Virgo Collaboration)

The Gravitational Wave Open Science Center provides data from gravitational-wave observatories, along with access to tutorials and software tools.

Get started!

Download data

GWTC-1: Catalog of Compact Binary Mergers

Join the email list

Data become public after 24 months

VIRGO

THANKS

Sarah Caudill (Nikhef), Franco Carbognani (EGO)

INFN Torino: Gabriele Fronzé, Stefano Lusso, Sara Vallero

INFN Genova: Luca Rei

EGO: Stefano Cortese, Giuseppe di Biase

ICCUB: Jordi Portell i de Mora et al.

CNAF: Lucia Morganti, Stefano Dal Pra, Diego Michelotto

...and of course the staff at CNAF Tier-1 and EGO IT.