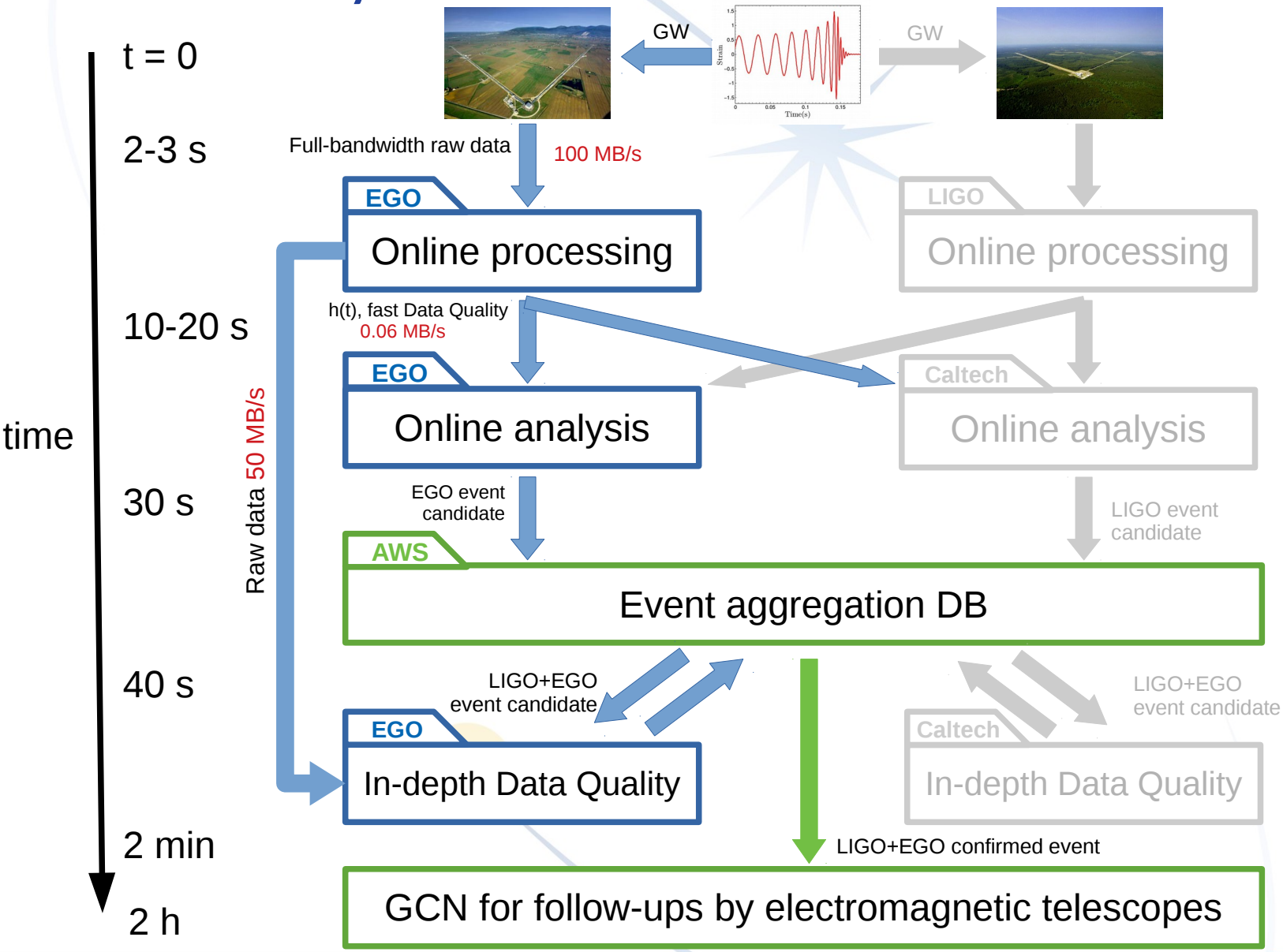
EGO use case

Pierre Chanial, EGO

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

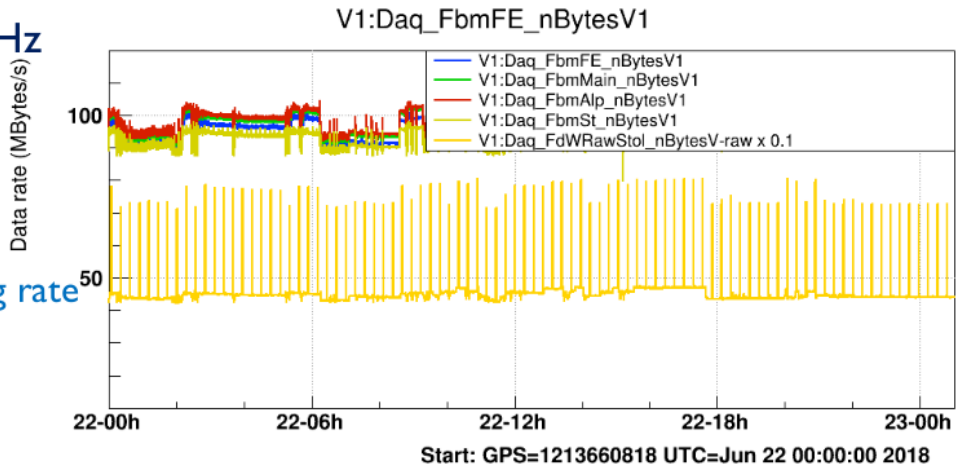
Data low-latency workflow



Funded by the European Union's Horizon 2020 - Grant N° 824064

Data production

- Raw data: 32700 channels from 1 Hz to 100 kHz
  - + samples @ 400 MHz, images…
  - Archived at the computing centers
  - Now: around 50 MB/s 100 s. long files
- Full raw data
  - Raw data + debugging channels at higher sampling rate
  - Keep at the site on a circular buffer of few days
  - Rate = 80-90 MB/s 100 s. long files
- Reduced data stream: most channels at 50 Hz
  - Convenient for commissioning studies
  - Rate = 0.8 MB/s 43 files of 1.5 GB per day
- Trend data stream: min/mean/max/rms @ 1 Hz
  - Convenient for long term studies or the search
  - Rate = 0.06 MB/s 1 file of 4.6 GB per day to archive
- h(t) frames:
  - The useful stream for event searches
  - Rate = 0.06 MB/s 43 files archived of 120 MB per day

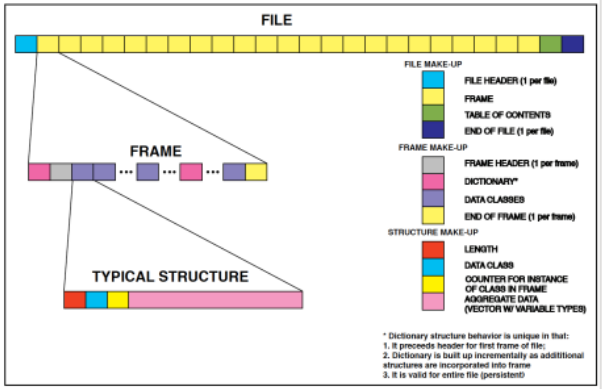


(slide by A. Bozzi)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Data format : GWF

- Virgo is closer to an accelerator rather than a high energy detector:
  - The primary data stream stream is a continuous stream of channels not events
- Data are organized in frames
  - Frames are a time window (from 1s to hours), containing multiple parallel channels
  - A Virgo frame files
    - Could contains multiple frames
    - Has a Table Of Content for fast single channel access
    - Includes checksums for standalone file verification
    - Include lossless compression algorithms
    - The frame format is common to all (ground based) GW detectors
  - We have different types of frames, see later
  - Tools to merge frames, make them longer, shorter…
- Events are produced downstream from a small number of channels
  - Small data rate compared to the raw data stream

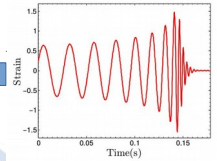


(slide by A. Bozzi)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Data offline workflow

GW



GW

Full-bandwidth raw data

EGO

Online processing

Raw data

h(t), fast Data Quality

CC

Offline pipelines:
Continuous waves,
Compact binary
coalescence

LIGO

Online processing

EGO

Online analysis

Caltech

Online analysis

EGO event candidate

LIGO event candidate

AWS

Event aggregation DB

LIGO+EGO event candidate

LIGO+EGO event candidate

EGO

In-depth Data Quality

Caltech

In-depth Data Quality

LIGO+EGO confirmed event

GCN for follow-ups by electromagnetic telescopes

Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Lifecycle

- Hot (frequent access) : for low-latency processing & analysis
  also h(t) for offline pipeline analysis
- Cold (once a month): Raw data < 2 yr (detector characterization)
- Archive (less that once a month): Raw data > 2 yr (archive)

Reprocessed 2-3 times per year for h(t) recalibration

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Most representative use case : data distribution

Bulk dataTransfer was born (~ 2007) like a collection of bash scripts and procedures (manually triggered) used for sending some Virgo set of rawdata files related to a specific engineer or commissioning run. We had some constraints to satisfy:

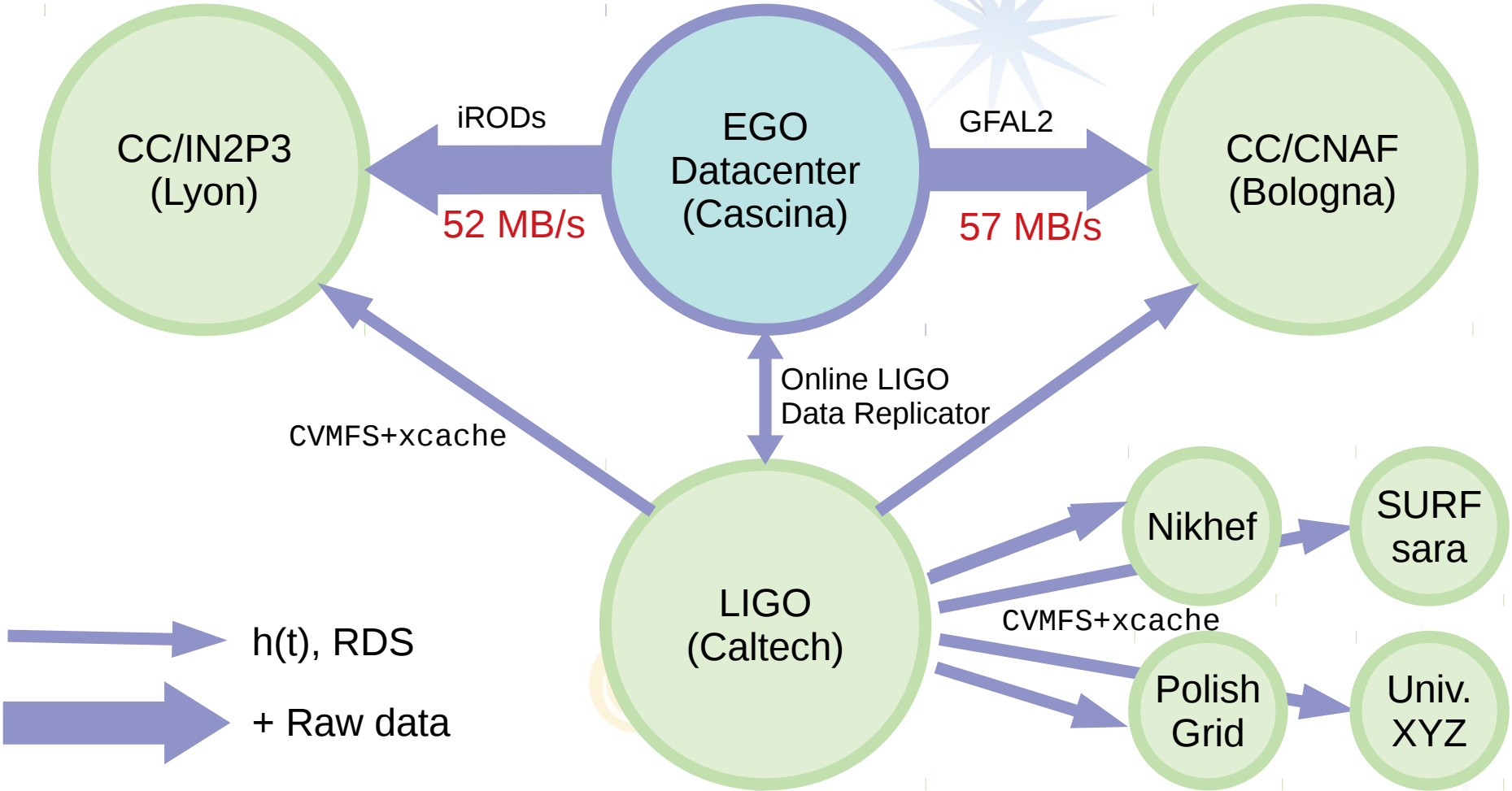
- the time ordering file transfer was a must (due to disk file allocation – no more valid constraint);
- were no possibility to install local daemons/servers;
- different checksum controls (md5sum for CCIN2P3 and adler32 for CNAF);
- the data transfer tools used for sending data were different for each CC and they changes periodically:
    - CCIN2P3: bbftp, SRB, iRODS;
    - CNAF: ftp, castor, LCG, GFAL2;

Experience gained in data transfer has led over time to the implementation of increasingly sophisticated procedures up to the current solution (using threads, sockets, SQL, robust exception handling, CLI), used for O2 in 2017 and O3 in 2019.

(slide by A. Bozzi)

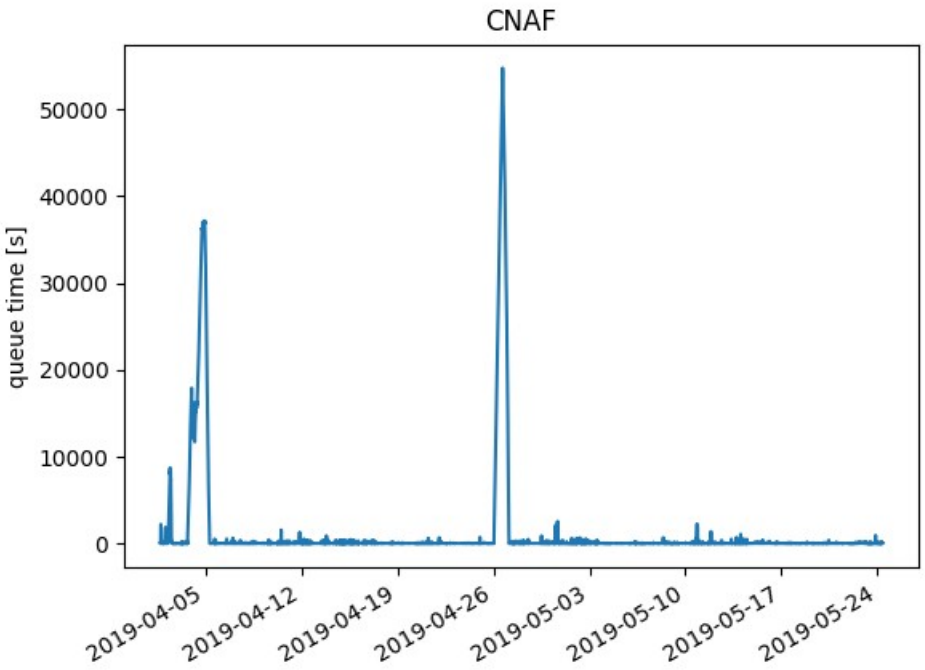
Funded by the European Union's Horizon 2020 - Grant N° 824064

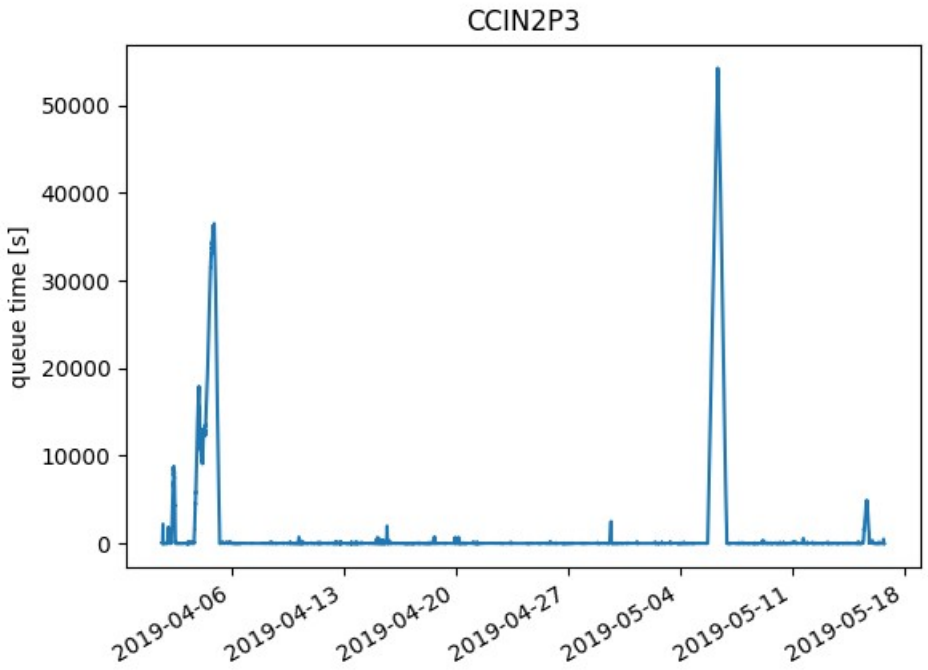
Data distribution : LIGO (low-latency) & Tier-1 CC (offline analysis)



Funded by the European Union's
Horizon 2020 - Grant N° 824064

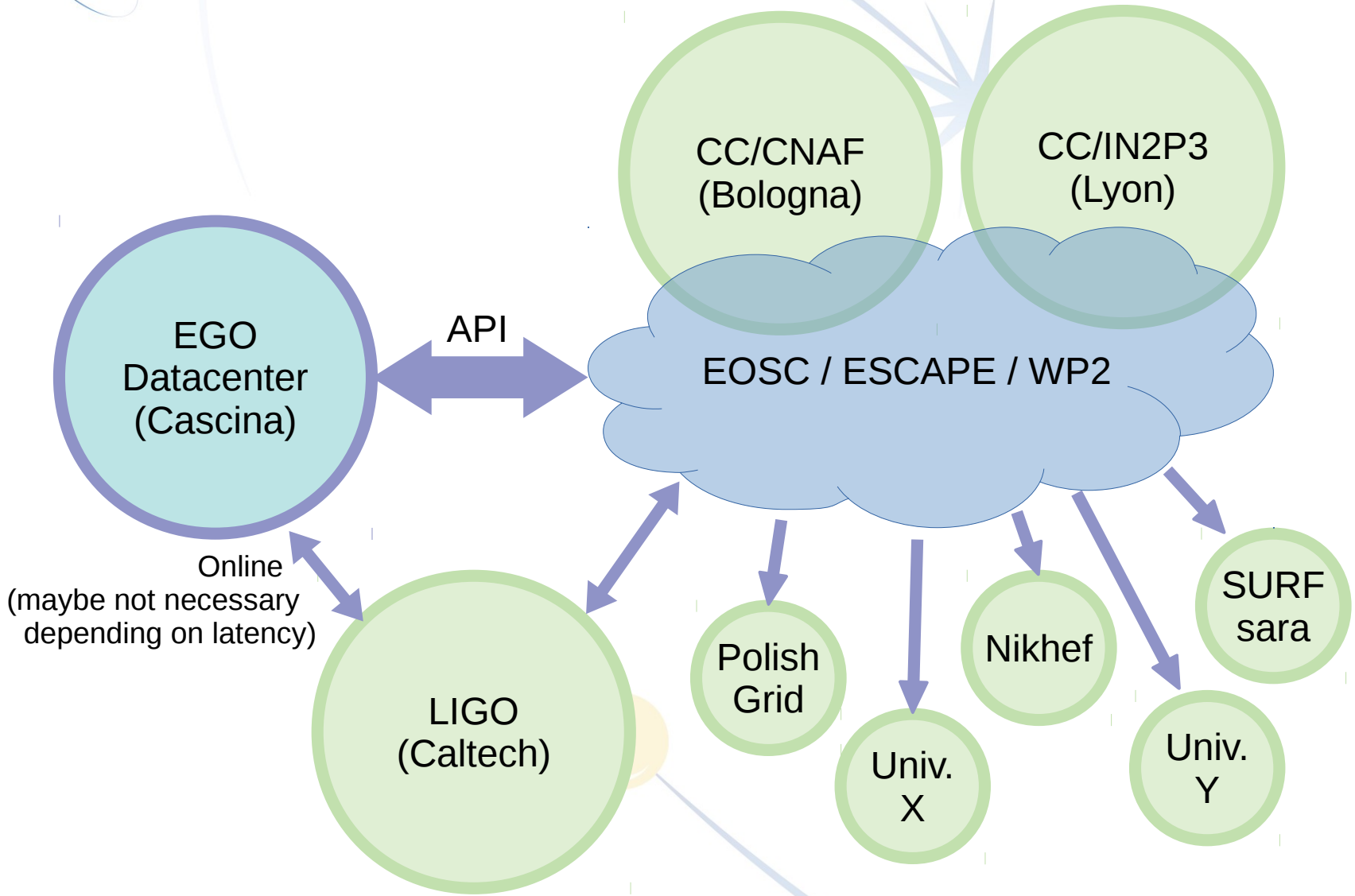
Raw data transfer (O3 run)





Funded by the European Union's Horizon 2020 - Grant N° 824064

Data distribution : Wouldn't that be nice?



Funded by the European Union's
Horizon 2020 - Grant N° 824064

Thank you !

Funded by the European Union's Horizon 2020 - Grant N° 824064