Table ronde sur la publication des données

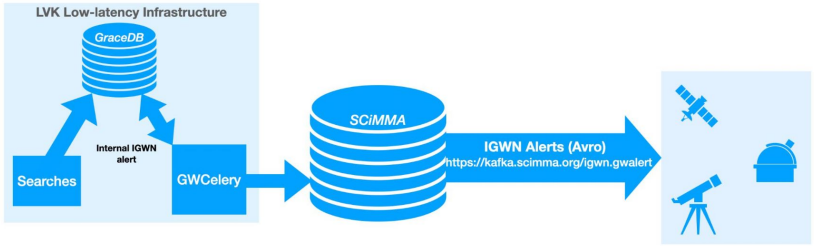
Eric Chassande-Mottin, Antoine Petiteau, Ed Porter

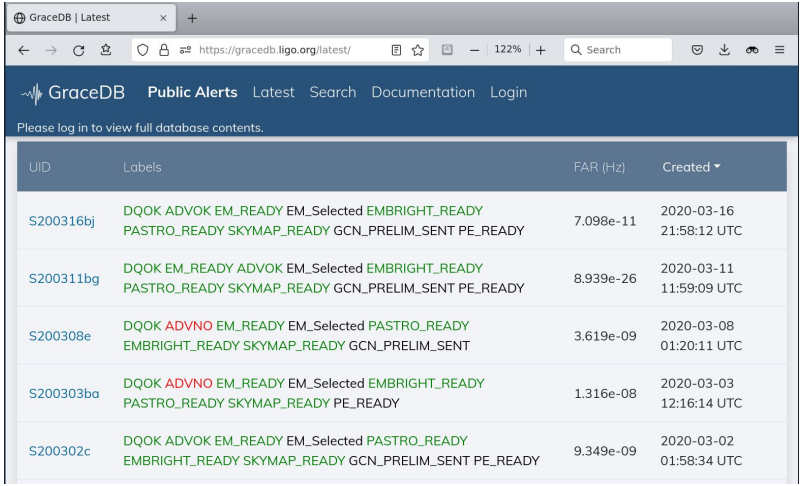
6ème Assemblée Générale du GdR Ondes Gravitationnelles
Oct 2022, Toulouse

Virgo – Open science: status and future

- **GW Open Science Center – GWOSC**
  - Started in 2011 by Caltech under NSF impulse
- Release policy – Cadence & proprietary period
  - Releases will occur every 6 months, in blocks of 6 months of data, with a **latency of 18 months** from the end of acquisition of each observing block
- So far data from LIGO, Virgo, GEO and KAGRA have been released according to this schedule
  - O1: 2018 - O2: Feb 2019 - O3: Apr & Oct 2021 O3GK: Mar 2022
- **Typical traffic: 100-200 users/day**
  - 330 papers published with GWOSC data
- Scientists (in and outside LVK)
  - **Searches**: “bulk” data, DQ, calib systematics
  - **Astro population**: event catalog with param estimates
  - **Test of GR, waveform**: GW event with data snippet around the event
- University and high-school students
  - **Hands on:** data analysis software and tutos

Virgo – Open science: status and future



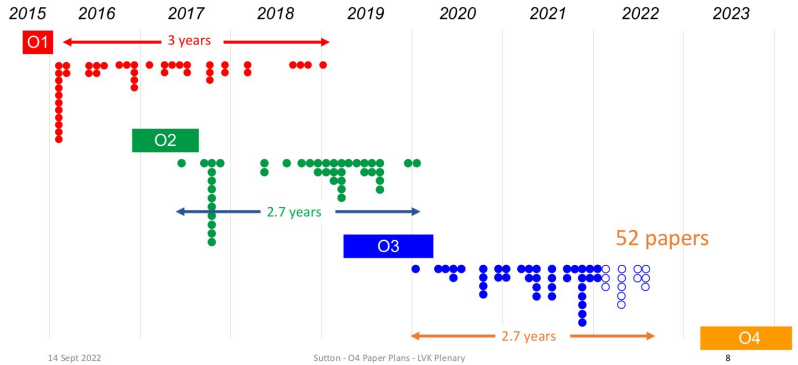


Low latency alerts

- Expect 1 alert / day during O4
- Few on-going upgrades
  - New distribution channel: SCiMMA kafka broker
  - Early warning (pre-merger) alerts
  - Preliminary alert with < 30 sec latency
  - Skymap : more compact multi-scale format

Virgo – Open science: status and future

Timeline of runs and paper releases (as of Feb 2022)

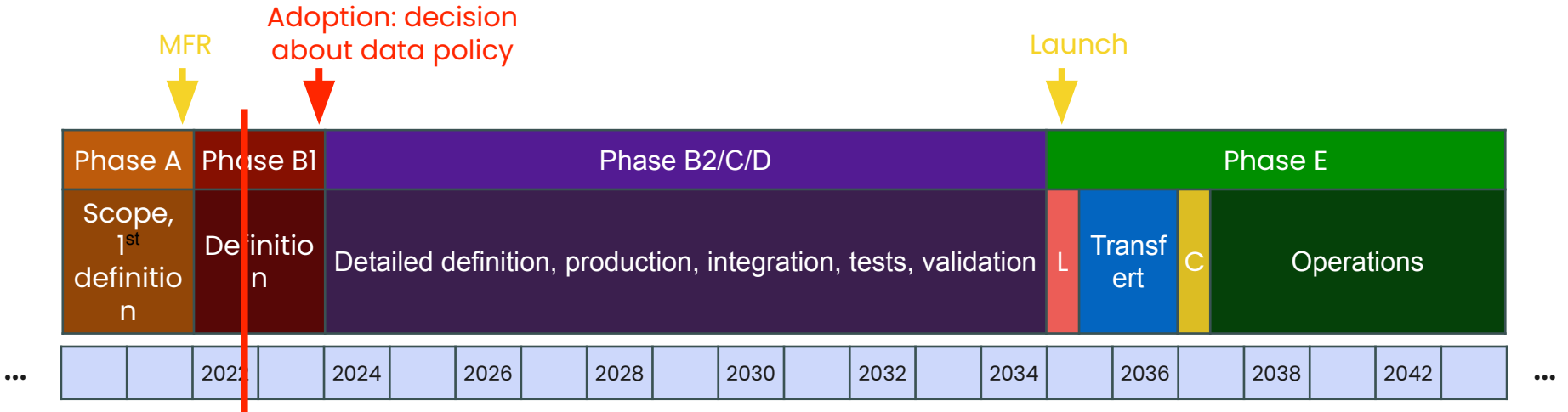


Release plan for O4

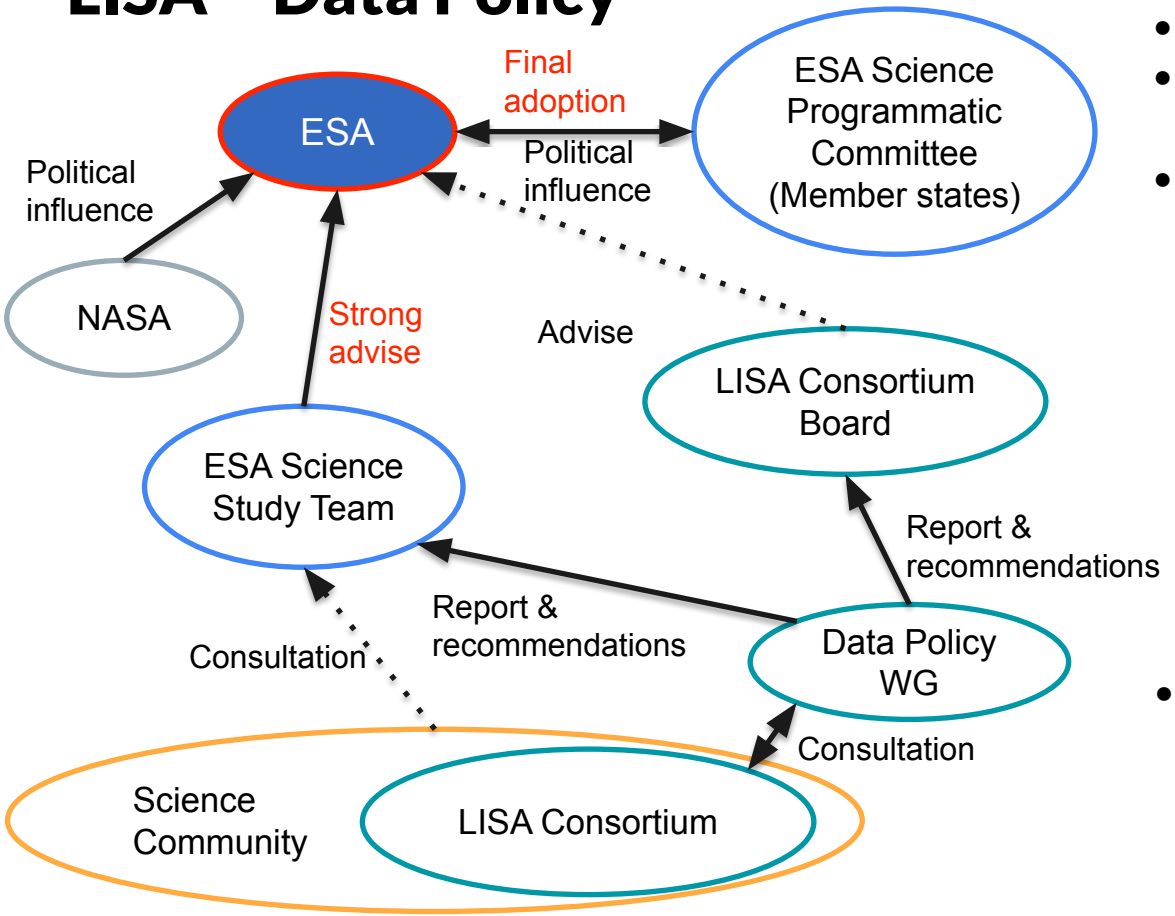
- Extensive discussion about the evolution of the proprietary period
  - Dedicated committee – 40 page internal report
  - Conclusion: remain with the same public release policy as O3 [*+18 months latency*]
  - Publication plan for O4 led to reconsideration of alternative scenario – Under discussion
- This question is key for the stability of the collaboration
  - Provide sufficient time for reaping academic reward in return on investments/efforts to produce the data
  - Ensure a high standard of quality
  - Each run starts with a ‘new machine’
  - Connect to work condition (stress and pressure)

LISA – Data Policy

- Data policy is in the Science Management Plan validated at adoption (Nov 23) => "decision" in the next months
- Decision taken 15 years before the first real data arrives and it's the first mission of such kind => some (limited) flexibility is needed.
- Different data levels; two groups:
  - L0 (raw data) / L1 (TDI data: data where dominant noises have been reduced)
  - L2 (results from multiple pipelines extracting GWs) / L3 (final catalogs and other science products)



LISA – Data Policy



- ESA will decide at the end
- But needs the agreement of the ESA member states
- Elements to consider:
  - Scientific consideration from ESA Science Study Team, based on the vision of the community at large (including Consortium)
  - Political vision of ESA member states
  - Political vision of NASA (partner)
  - Vision of the LISA Consortium (Science Ground Segment provider, instrument providers and majority of the scientists supporting the mission)
- In addition discussion at national level: in France dedicated group to identify the French vision

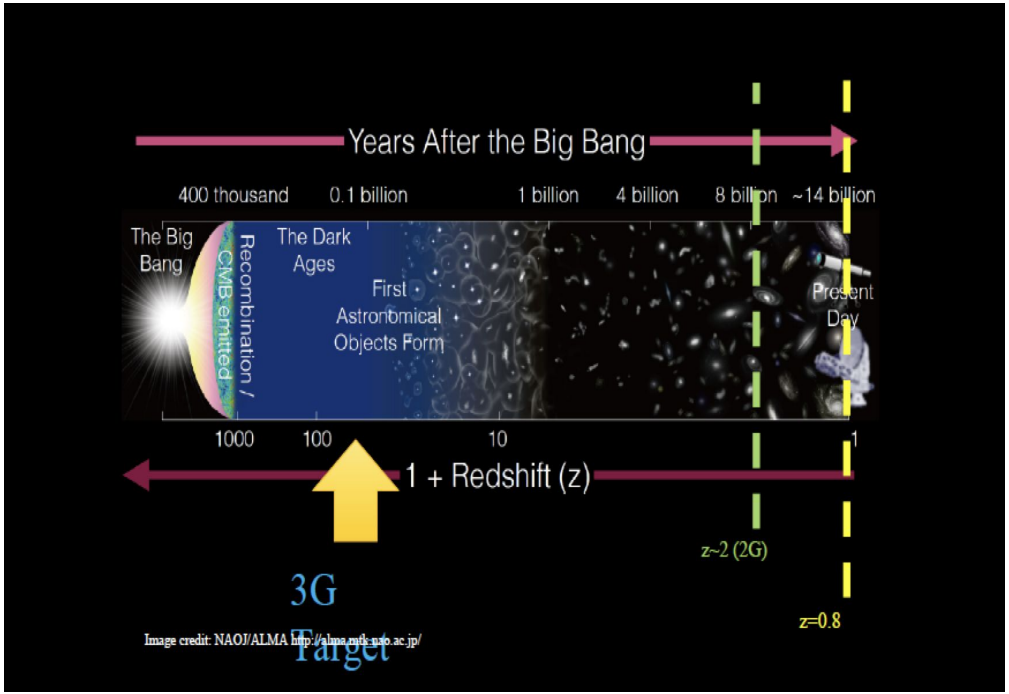
LISA – Data Policy

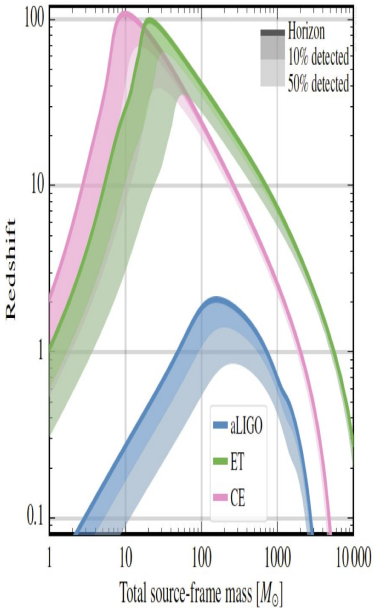
- Push for open data: LISA data will be opened for sure but the question is when
- Key points from discussion sessions in the LISA Consortium
  - Data quality
    - early release => poor quality data => faulty science
    - difference of opinion on the definition of quality data and on the analysis time required to produce quality data
  - Credit and career advancement for the core contributors to the mission
  - Impact on the Consortium: no proprietary period => risk of many people leaving the Consortium and not having ressource for doing deep analysis.
- Agreed points and commonalities
  - Alerts will be released as fast as possible; the question is more on what is in the release.
  - Data validation is done on the measurement of GW from the strongest verification galactic binaries (VGBs).
  - Data release in chunks:
    - First release 6 to 9 months (2 to 4 VGBs with SNR = 20)
    - Later releases may have shorter chunks
  - L0/L1 data release documentation is ESA's responsibility, L2/L3 is more in the hand of the Consortium
  - Re-analysis of all data at each release so each data release is not the final "best" analysis of that data

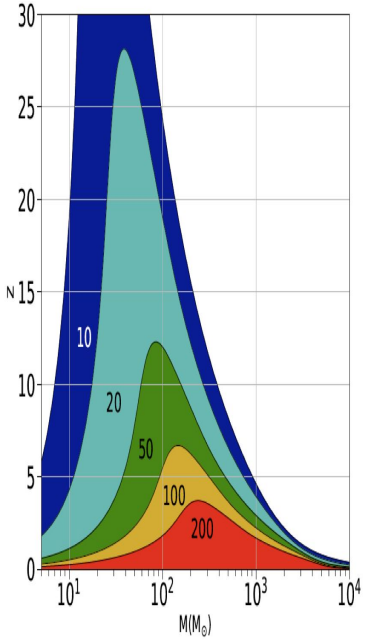
LISA – Data Policy

- Scenario 1
  - L0-L1 data are released as soon as ESA has confidence about their quality (detection of some VGBs by the Consortium); then almost continuous release
  - L2-L3 data produced by consortium are a property of consortium which decides when to release
- Scenario 2
  - Every data release includes L0/L1 data. In addition, the first 1 or 2 data release(s) also include L2/L3 data.
  - The first chunk of data has a proprietary period of 1 year (data validation and papers).
  - If and only if in the first chunk there is no MBH merger & no EMRIs are found, the second chunk also has a proprietary period.
- Scenario 3
  - Every data release includes both L0.5/L1 and L2/L3 data (no fundamental difference in data management).
  - Data is released in chunks, with a lag, to allow construction of L2/L3 catalogues.
  - Releases include associated technical and scientific documentation produced by the consortium:
    - The first release occurs with a relatively long delay after data taking (min 6 months, max 1 year) to allow processing, validation and paper writing (description of the instrument performance, data processing & catalogues, but also scientific interpretation of the 1st catalogue).
    - Subsequent releases occur with lower latency, e.g., in 3 month chunks, with a lag of 3 months.
- Current agreement is more in the direction of scenario 2 (mixed with some elements of scenario 3)

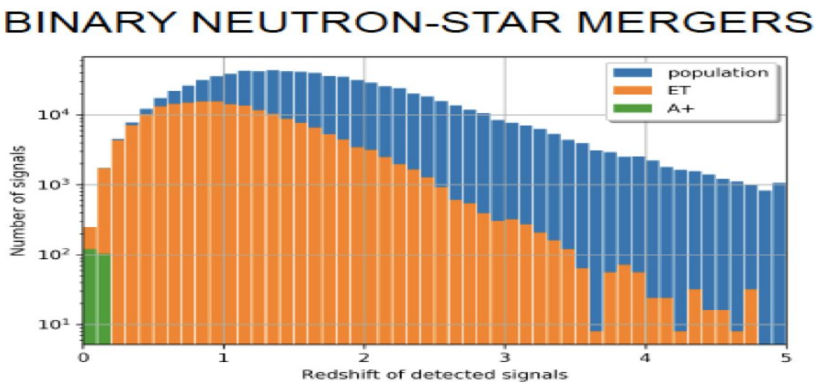
Einstein Telescope

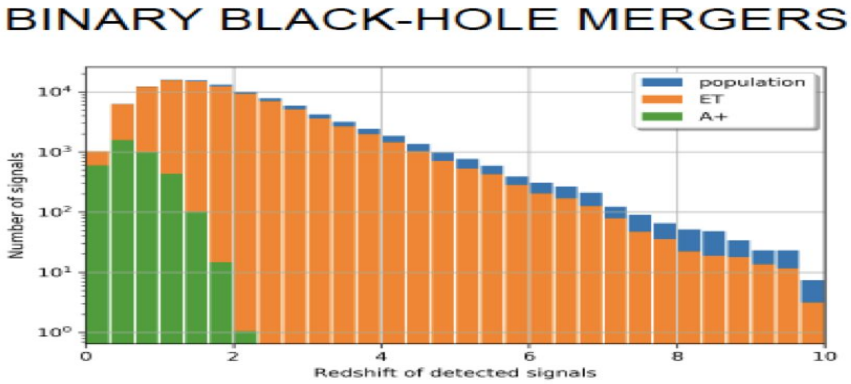






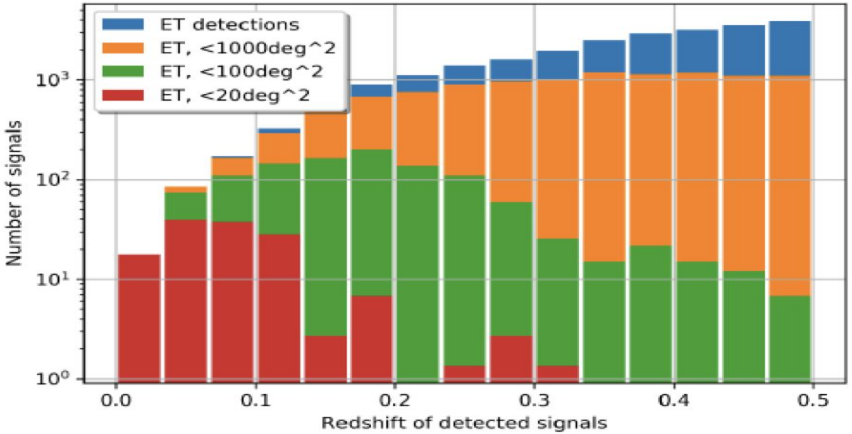
Einstein Telescope

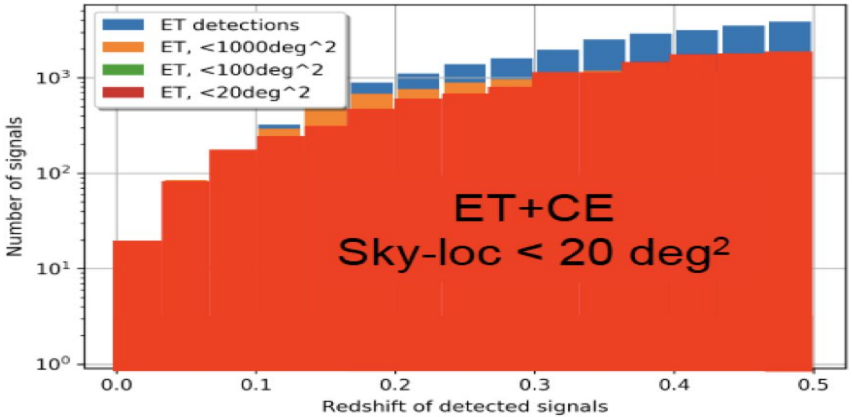




10^6 BBH mergers/yr up to z = 50
10^5 BNS mergers / yr up to z = 2
10-100 possible EM counterparts / year
High SNR events

MMA Observations





In 1 year of observation:
100 detections/yr with sky error < 20 sq. degrees
Pre-merger alerts of hours - minutes

Open questions on ET data policy

- No discussion has taken place on the collaboration level yet
  - N.B. these are my personal opinions
- Currently adopting the LVK model
- Future collaboration with CE (NSF) will be necessary
  - Potential conflict over proprietary period
- Need a community to build and scientifically exploit ET
  - Community needs a scientific/career benefit return for the effort
- Impossible to see how this can be done without a proprietary period

Issues of open science for gravitational wave astronomy

- Objectives/benefits of opening the data
  - **Reproducibility** of the analyses
    - Enhance *credibility* of the result
    - Accelerate *dissemination*
  - **Accessibility** to a wider/larger scientific community
    - More results - Better *return on investment* for agencies
    - Share with scientists from developing countries
  - Give access to **general public** (“tax payer”)
- Side benefits from the actions required to open the data
  - **Long term preservation** of the data (make sure the data are readable)
  - Tracing and book-keeping: document **provenance**
  - Improve **internal accessibility** (to collaboration members)
    - Useful for interns and students
  - Provide incentive for **free software**

Issues of open science for gravitational wave astronomy

- Opening the data takes **time and energy (→money)**
  - Significant manpower to curate, document and review the release
  - Reward can be an issue for early career scientist investing time in this activity
- Affect the **group dynamics and cohesion** of large collaborations
  - Incentive for small group projects rather than collaboration core projects
  - Duration of the proprietary period fixes a deadline. If too short:
    - Risk of scooping when publication not in time
    - Not able to sustain high-standards for the final results (time needed for internal review)
    - Affect quality of working conditions (pressure and stress on vulnerable staff members)

Topics for the discussion

- What minimal requirements should satisfy a good policy for open science ?
- How to determine the right duration for the proprietary period ?
- How do we address the clear disparity in policy between Europe and the US?