Pushing the boundaries
of Spectroscopic Surveys

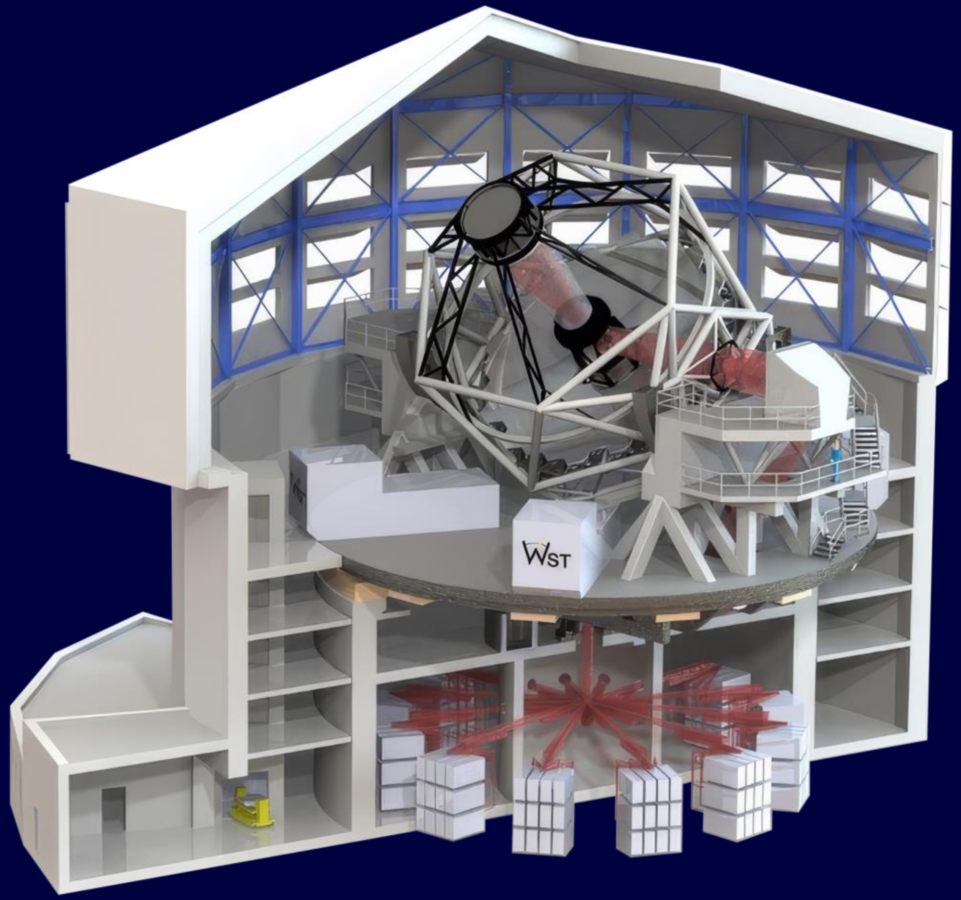
Funded by
the European Union

Integrating sustainability in the design of the Wide-field Spectroscopic Telescope's data processing and storage

Laurane Fréour

University of Vienna

In collaboration with Arlette Pecontal, Peter Weilbacher, Bodo Ziegler, and the WST Project Office

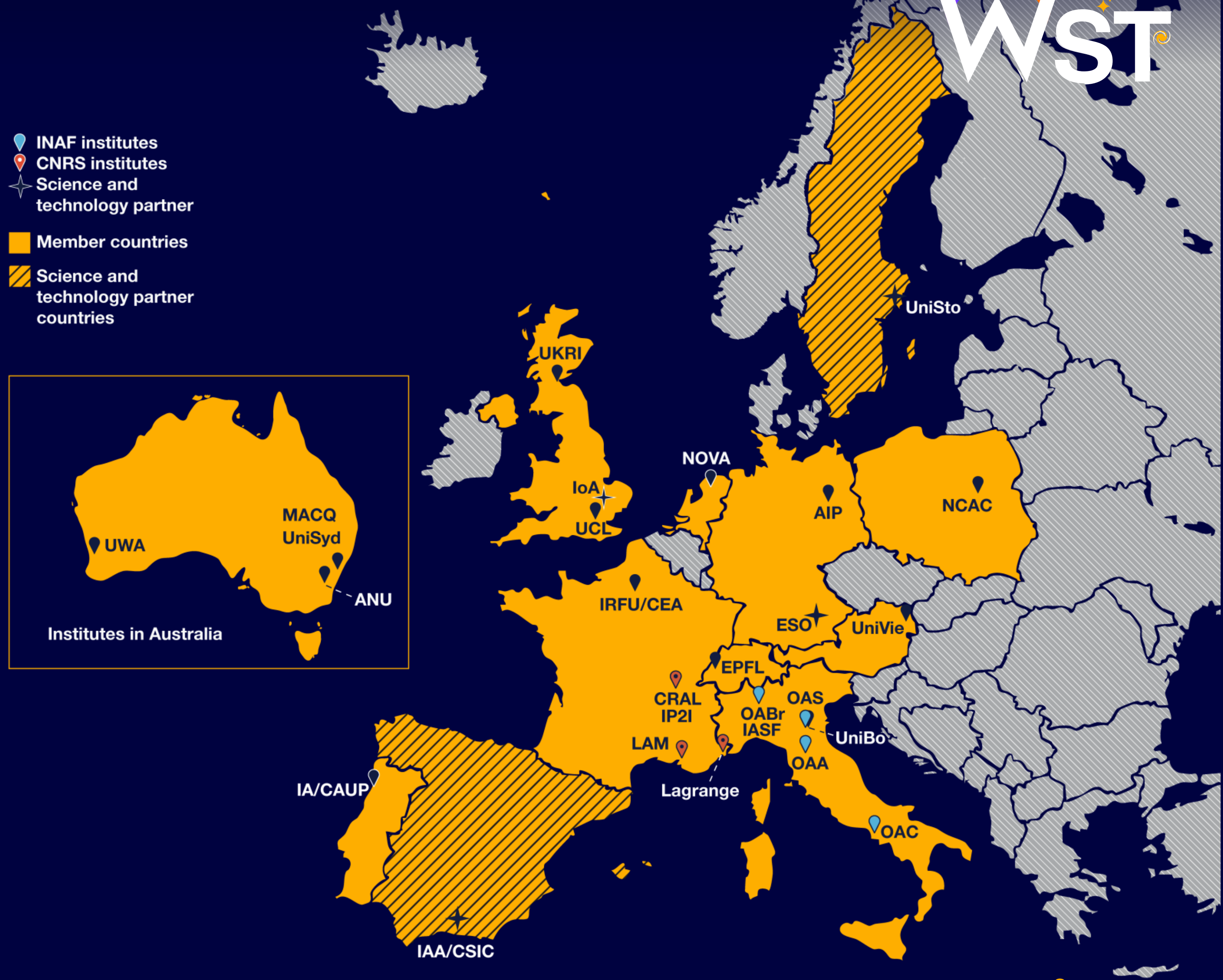


Outline

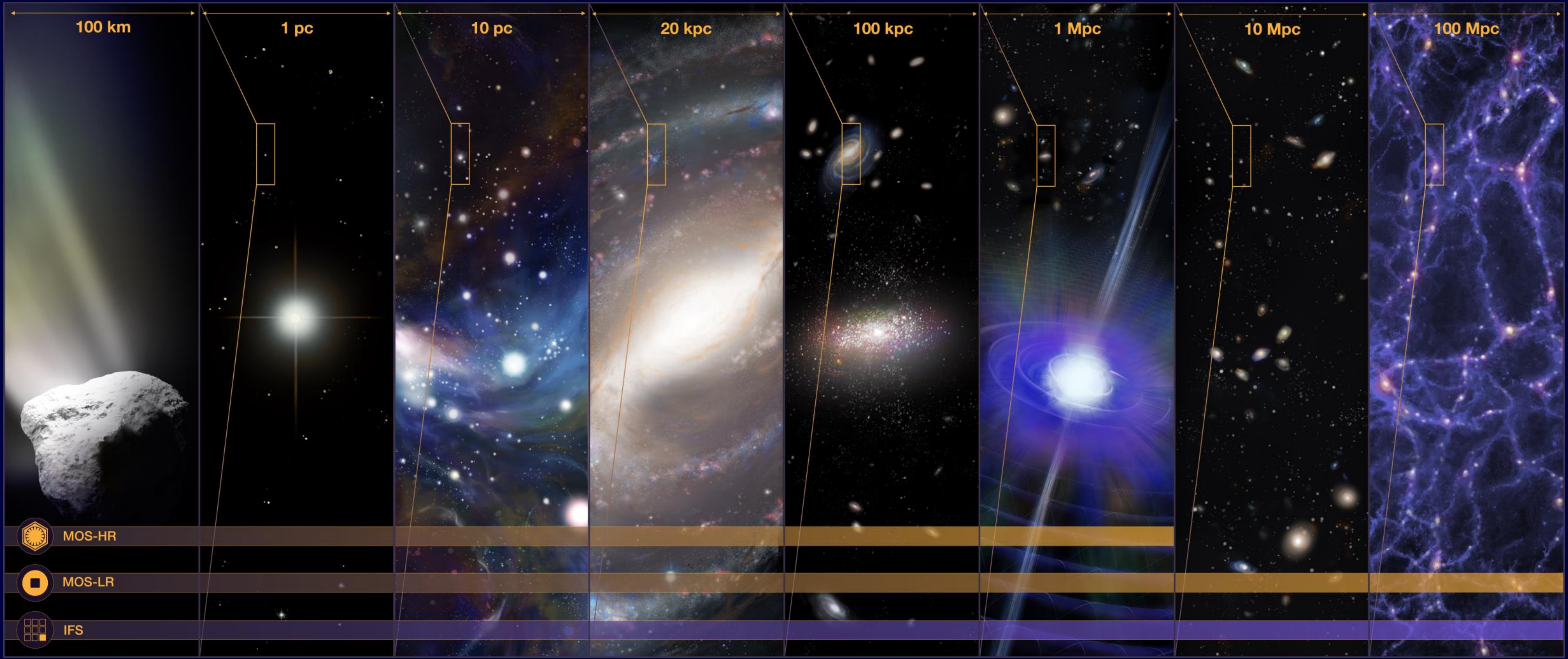
- The Wide-field Spectroscopic Telescope
- Time-domain astronomy & the era of big data
- Environmental considerations on WST's data processing and storage
- Comparison with other environmental impacts

Consortium

- **26 research institutes or universities** spread over 12 different countries
- **~800 science team members** from 34 countries

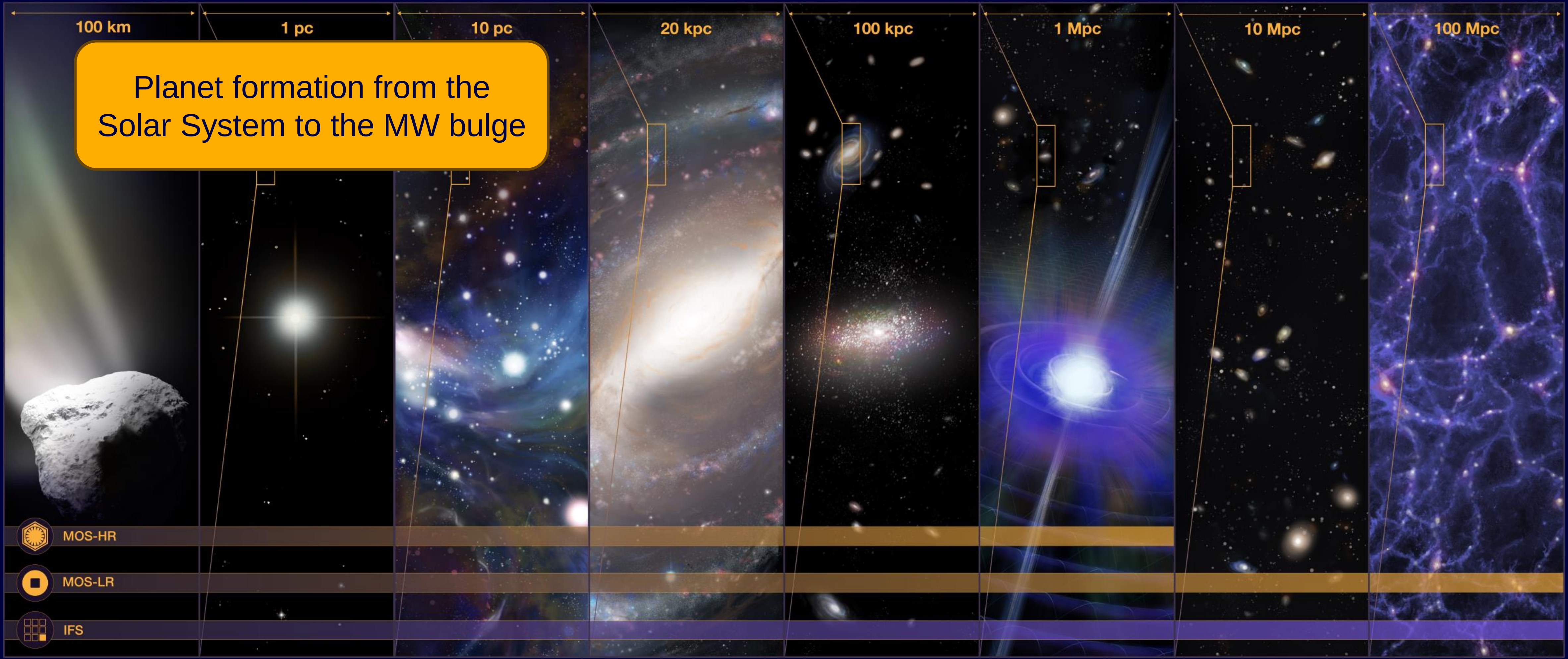


Addressing key scientific questions across all astrophysical scales



Credit to Maria Cristina Fortuna (INAF)

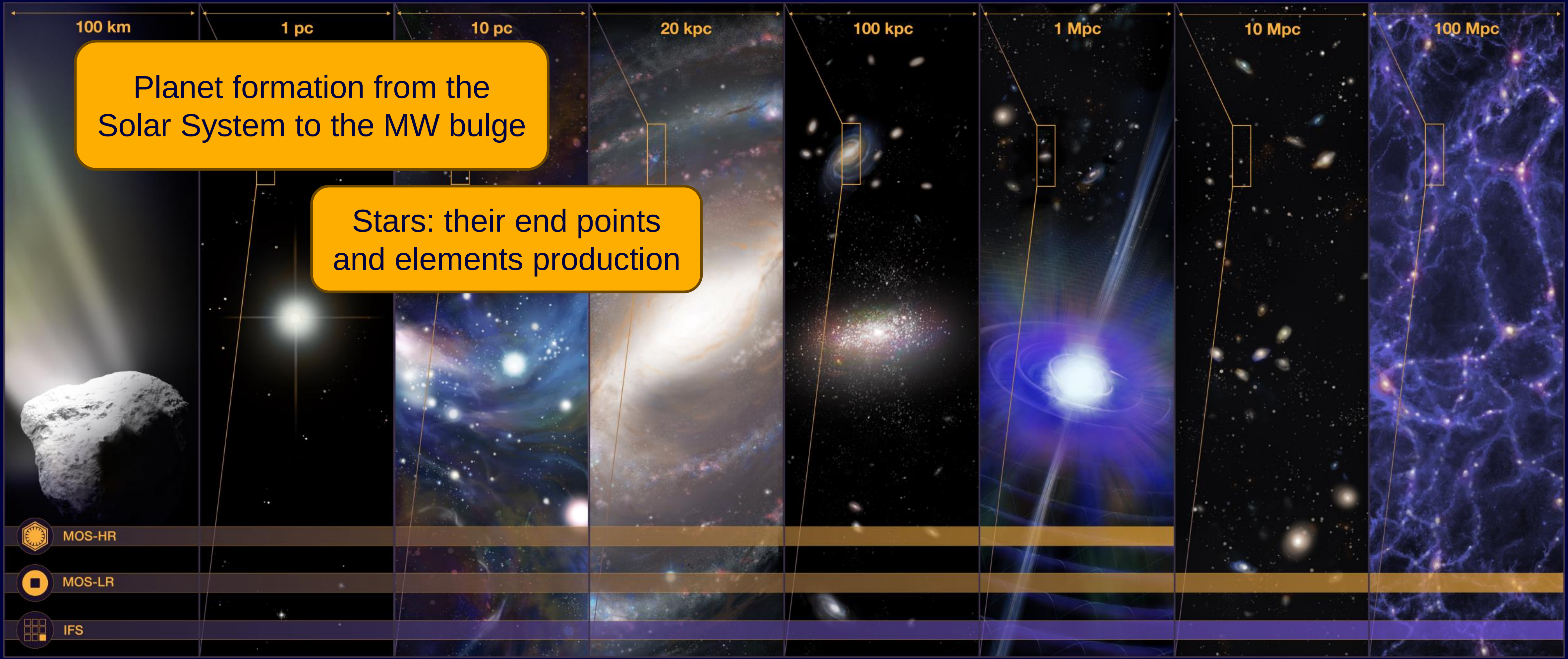
Addressing key scientific questions across all astrophysical scales



Planet formation from the Solar System to the MW bulge

Credit to Maria Cristina Fortuna (INAF)

Addressing key scientific questions across all astrophysical scales

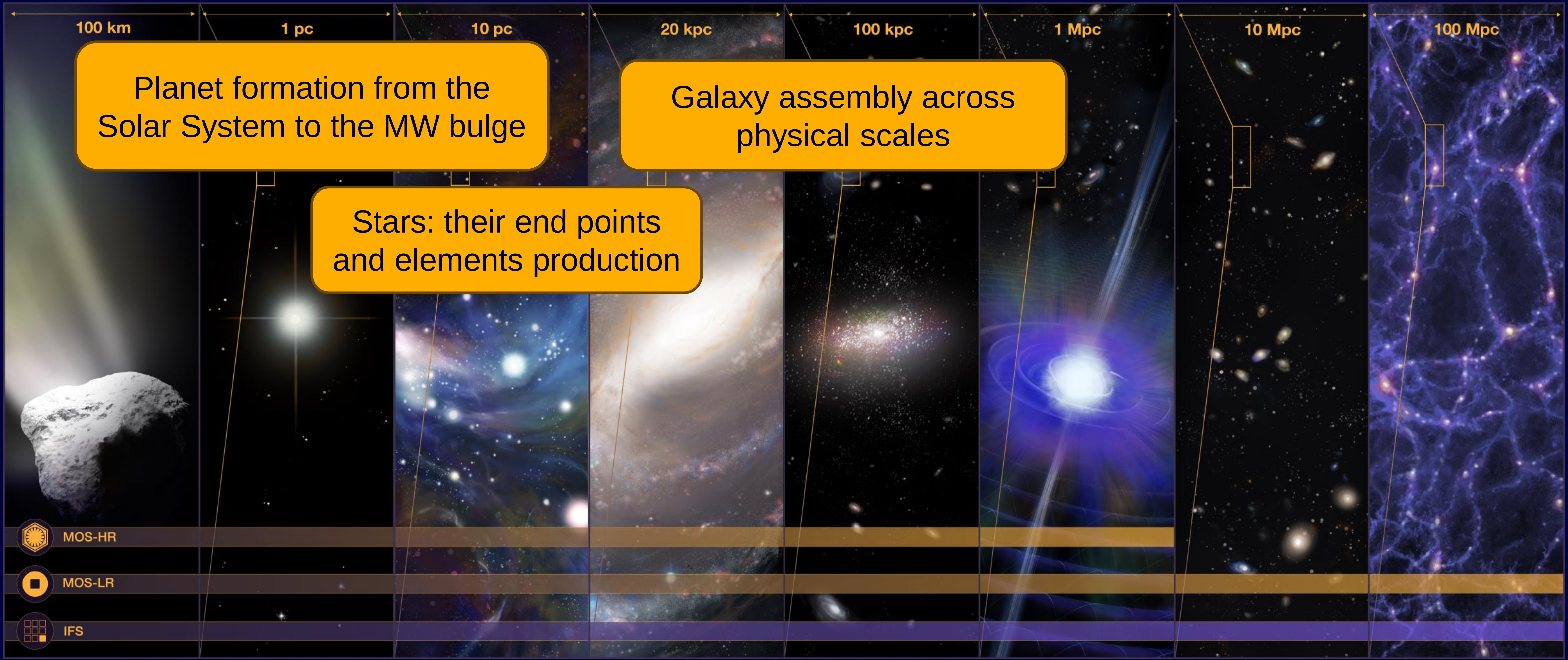


Planet formation from the Solar System to the MW bulge

Stars: their end points and elements production

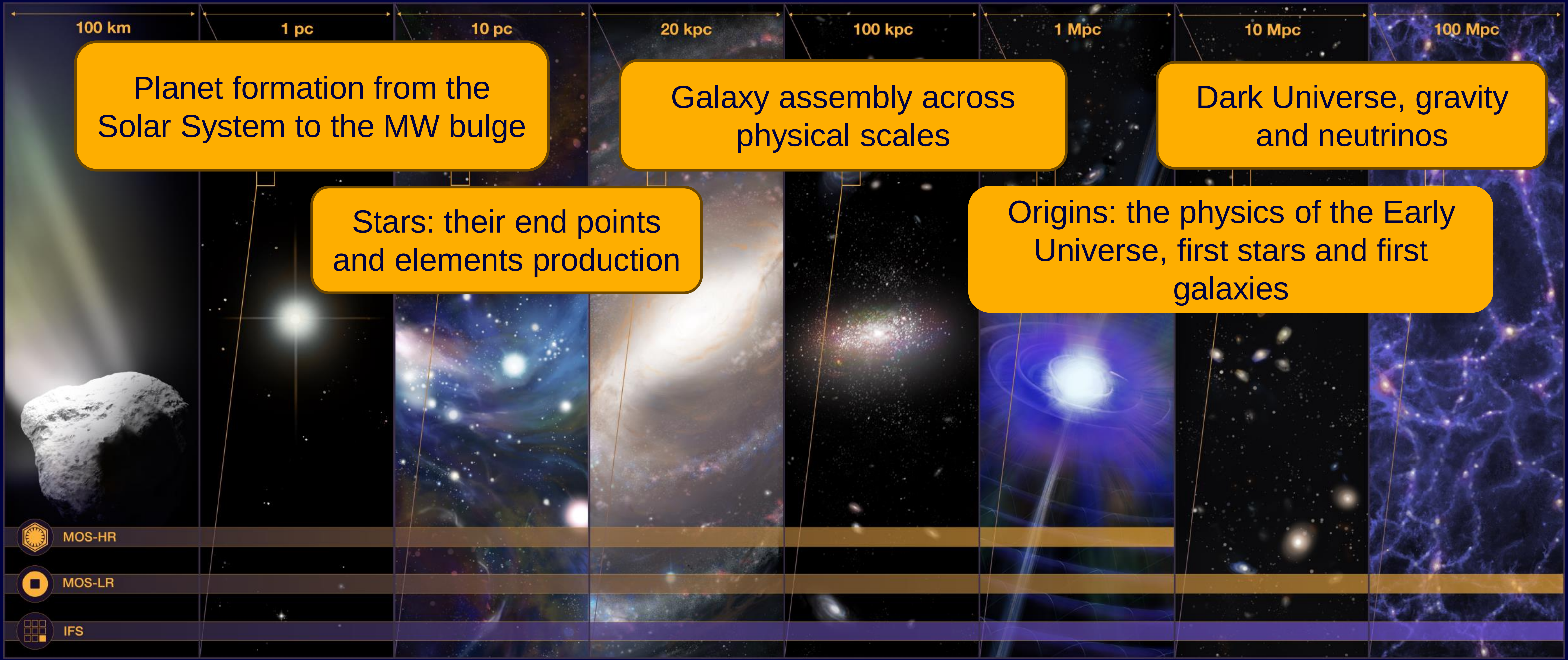
Credit to Maria Cristina Fortuna (INAF)

Addressing key scientific questions across all astrophysical scales



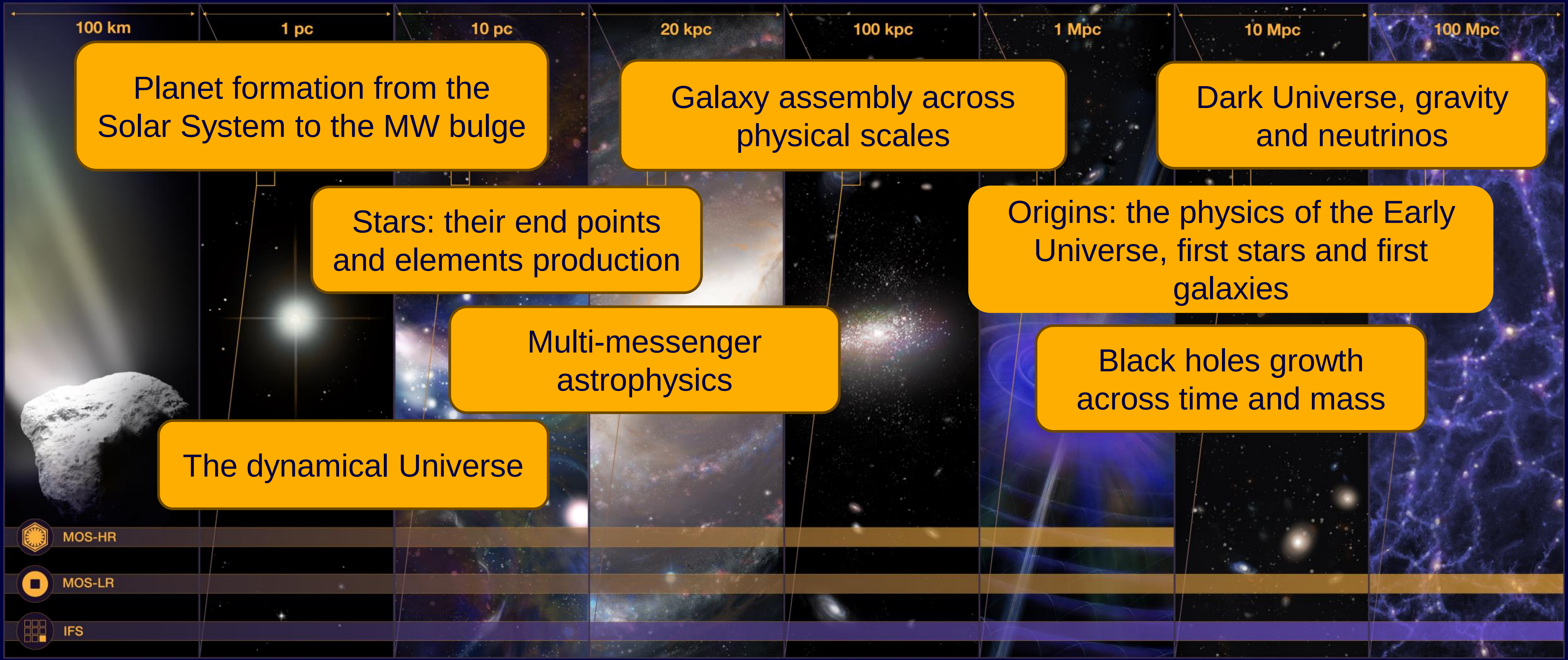
Credit to Maria Cristina Fortuna (INAF)

Addressing key scientific questions across all astrophysical scales

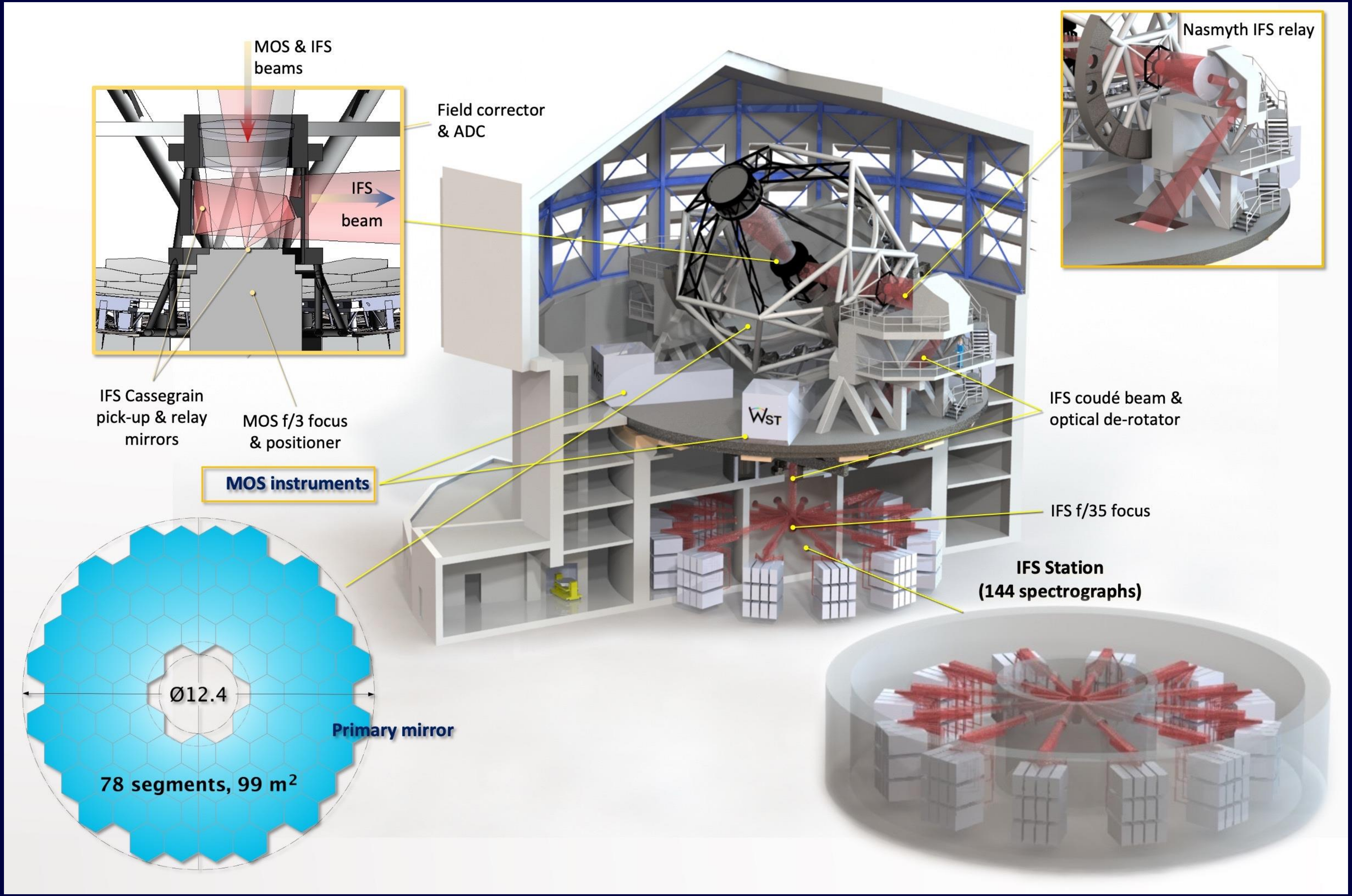


Credit to Maria Cristina Fortuna (INAF)

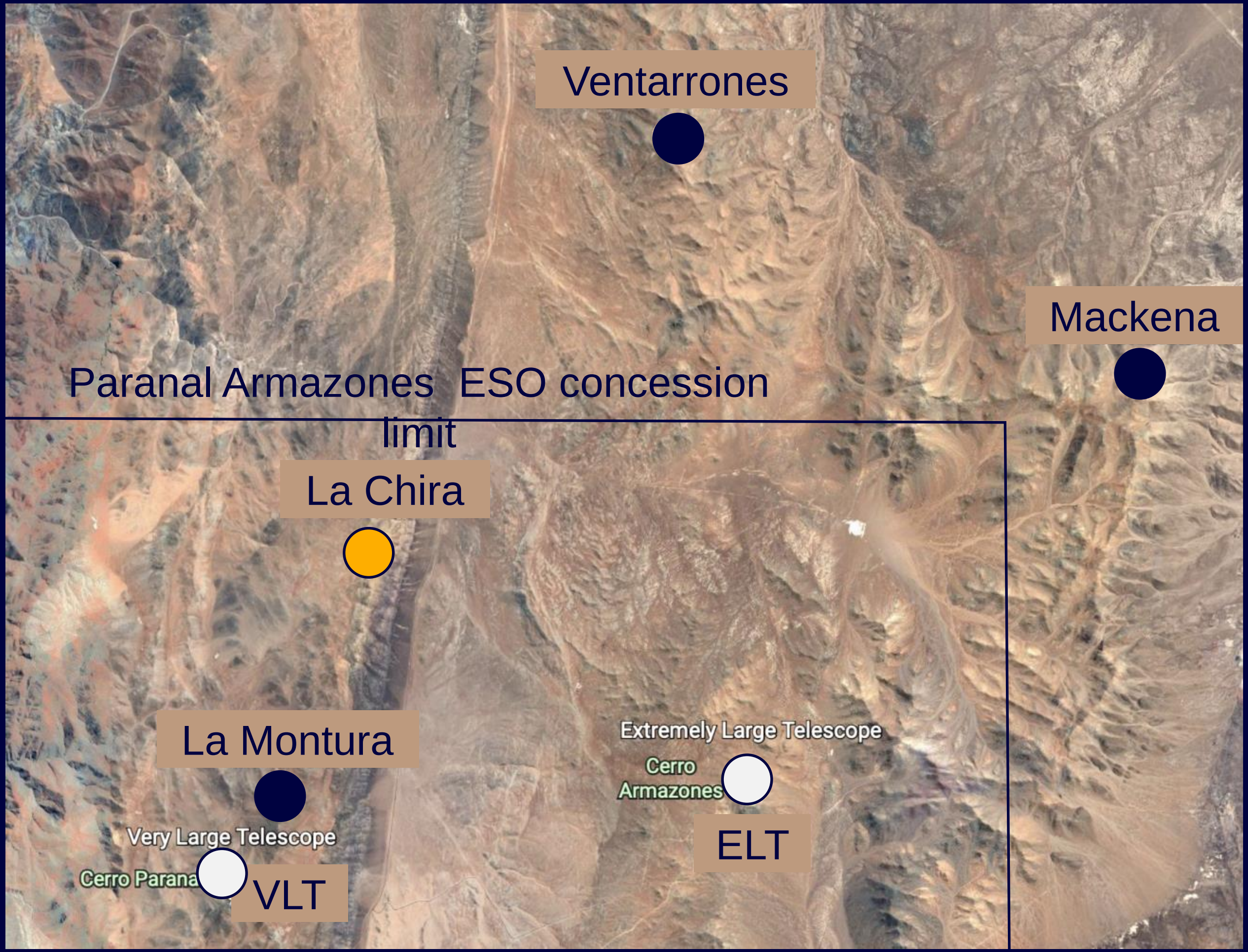
Addressing key scientific questions across all astrophysical scales



Credit to Maria Cristina Fortuna (INAF)


MOS & IFS
beams
IFS
beam
Field corrector
& ADC
Nasmyth IFS relay
IFS Cassegrain
pick-up & relay
mirrors
MOS f/3 focus
& positioner
MOS instruments
IFS coudé beam &
optical de-rotator
IFS f/35 focus
IFS Station
(144 spectrographs)
Ø12.4
Primary mirror
78 segments, 99 m²
WST

Potential sites in ESO areas



Paranal solar plant - 9 MW (Filippi et al, 2022)

1700 t of CO_2 equivalent

◀ Potential sites in and around Paranal-Armazones ESO area (Angel Otarola, ESO)

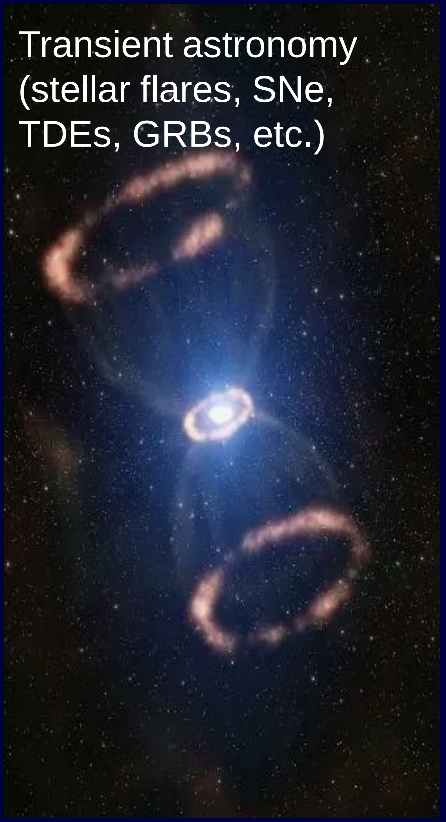
Time-Domain: Exciting science on all scales

MOS-LR

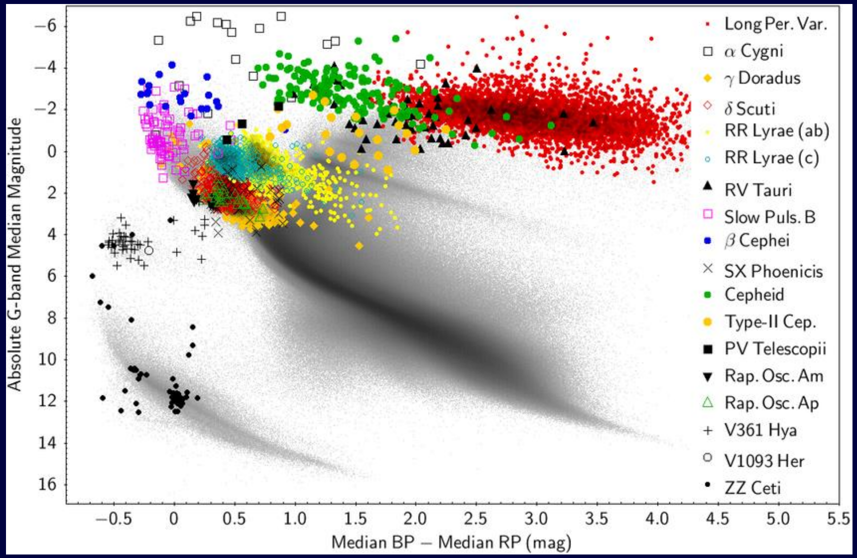
MOS-HR

IFS

Supermassive BHs (AGN, QSOs)



Compact objects (WDs, stellar-M BHs)



Stellar Variability (Gaia, LSST)

Solar system Comets & Exocomets

Operation challenges

How to

Run simultaneously multiple surveys in parallel,

with three instruments MOS-LR, MOS-HR & IFS,

allowing time-domain science (ToO and multiple visits)

and the ability to adapt to new science case in a reasonable time-scale

How to

Manage the data quantity

and complexity (spectra and datacubes),

with the need of rapid processing to issue alerts

and to maintain homogeneous results across the surveys,

while minimizing environmental impact

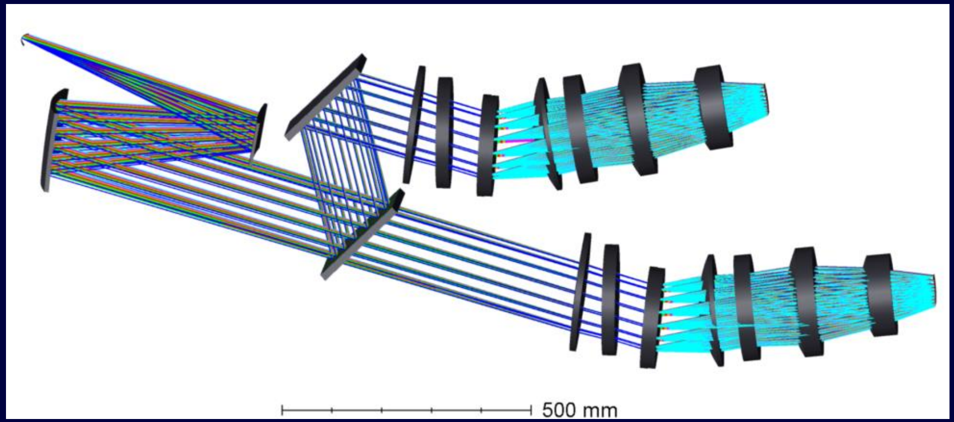
Step 1: Data collection

- Light is collected by the ~600 detectors continuously 10h/night

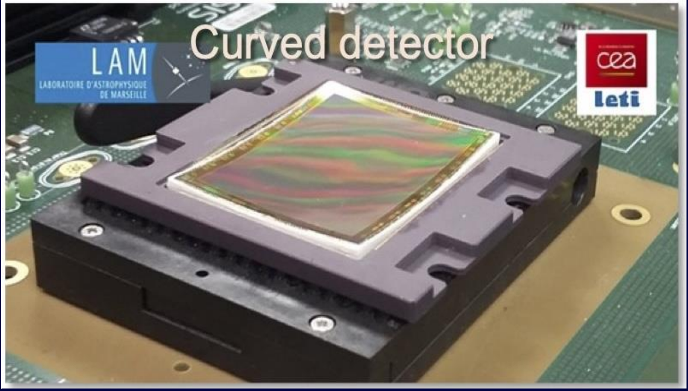
  → "Raw data"

- Volume of raw data ∝ Total number of pixels ∝ Number of detectors

- First estimate: **2.5 TB** of raw data per night



IFS spectrograph design (CRAL)

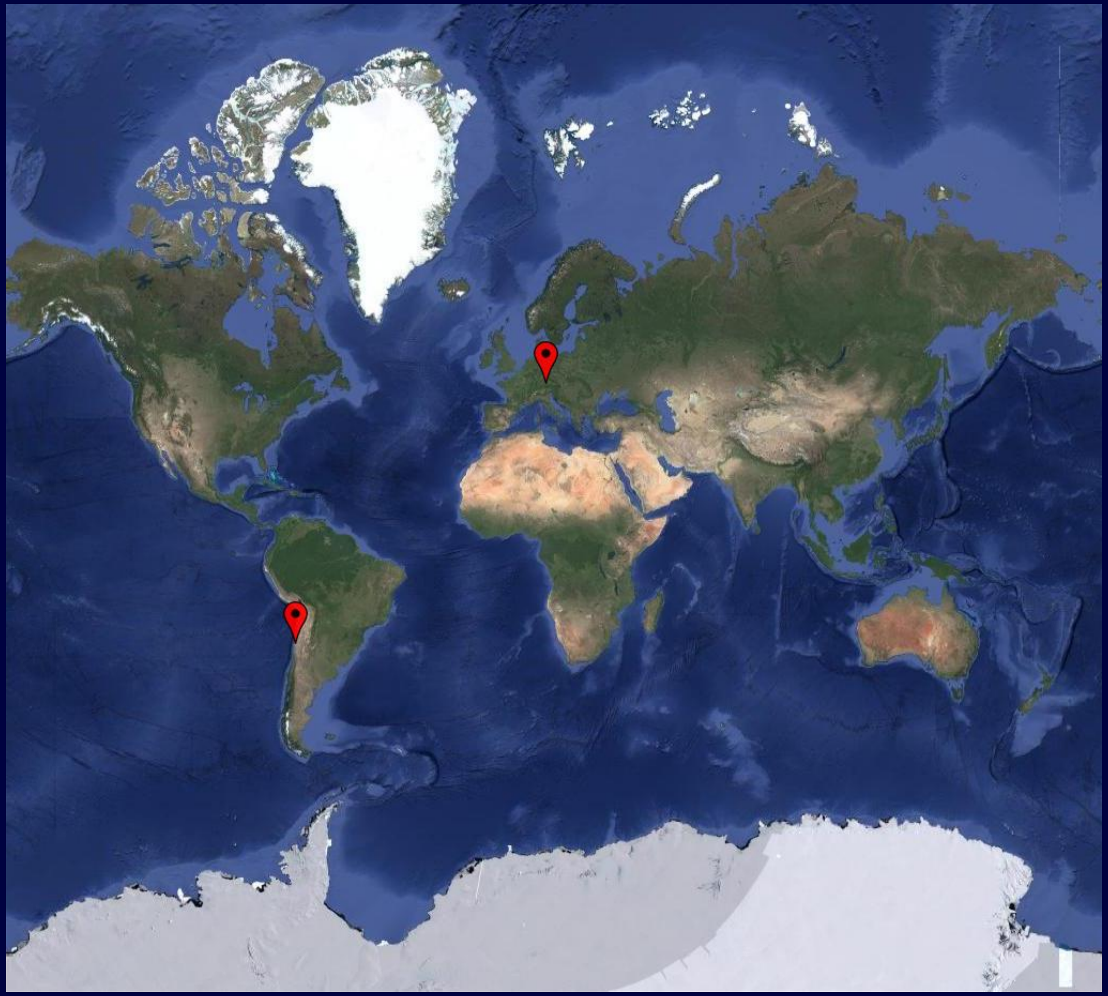


Curved CMOS (LAM)

Step 2: Data transfer

- The ~2.5 TB raw data should be transferred to Europe, where the data reduction and storage will take place.

- Geographical distance: ~11400 km



Step 3: Data reduction

- Raw data must be calibrated and reduced to be used by the community

- The data reduction process involves three distinct levels, with intermediate products, leading to the final data catalogue (e.g., stellar positions, radial velocities of stars) serving as the main resource for scientific analysis

Step 4: Data storage

- All the data (raw, calibration and reduced) should be stored

→ Increase over time

Research questions

- What is the carbon footprint of the different data processing steps?
- How does it compare with other aspects of the telescope (e.g., cooling of instruments)?
- How can we minimize the carbon footprint of data?

Carbon footprint of the steps

- **Step 2: Data transfer**
  Embedded impact of the optical fibers and shelters, using formulas & emission factors from Loygue et al. 2024

$$\text{Carbon footprint of a route } (kg\ CO_2\ eq./GB) = \sum_{equipement} \text{Emissions of use + manufacturing } (kg\ CO_2\ eq./GB) + \text{Emissions of fiber-optic } (kg\ CO_2\ eq./GB)$$

Carbon footprint of the steps

- **Step 2: Data transfer**
  Embedded impact of the optical fibers and shelters, using formulas & emission factors from Loygue et al. 2024
- **Step 3: Data reduction pipeline**
  Footprint quantified using the green algorithm (Lannelongue et al. 2021)

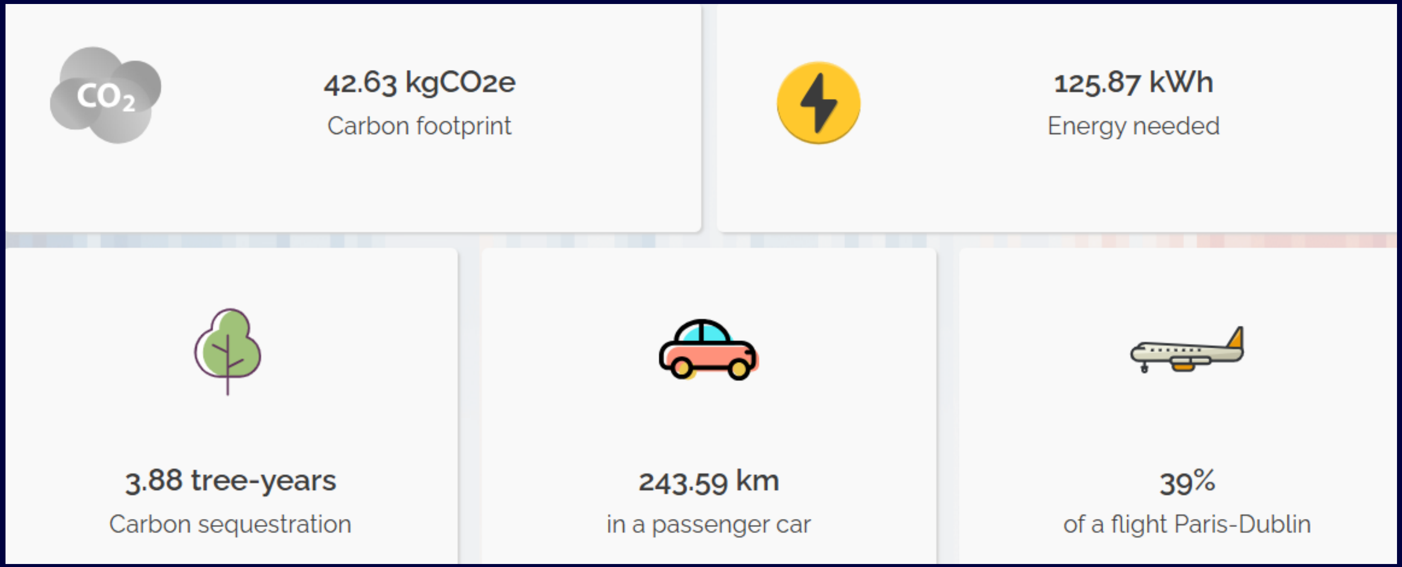


Image from: https://www.green-algorithms.org/

Carbon footprint of the steps

- **Step 2: Data transfer**
  Embedded impact of the optical fibers and shelters, using formulas & emission factors from Loygue et al. 2024
- **Step 3: Data reduction pipeline**
  Footprint quantified using the green algorithm (Lannelongue et al. 2021)
- **Step 4: Data storage**
  Different storage types with kWh/TB reported (Seagate)

Corvault 4U106 Data System with EXOS 24TB HDDs

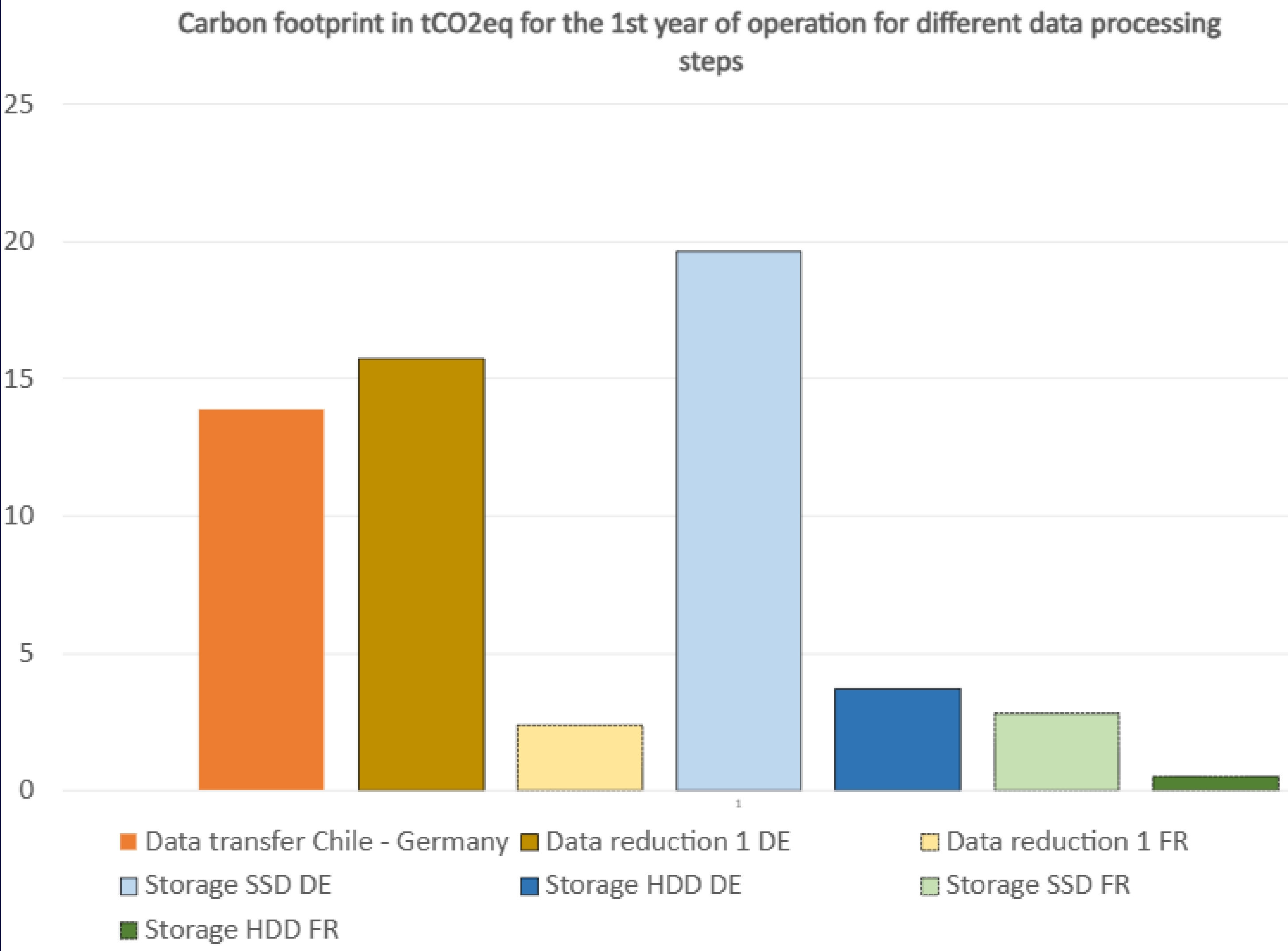
Power Consumption	Per Unit	Per TB
Average Idle Power (W)	1,560	0.61
Operating (W)	1,750	0.69
Average Annual (kWh)	12,943	5.09

Nytro 3000 SSD

Power Consumption	Per Unit	Per TB
Operating, Average (W)	9.0W	5.6W
Idle, Average (W)	4.4W	2.8W
Average Annual (kWh)	43.4kWh	27.1kWh

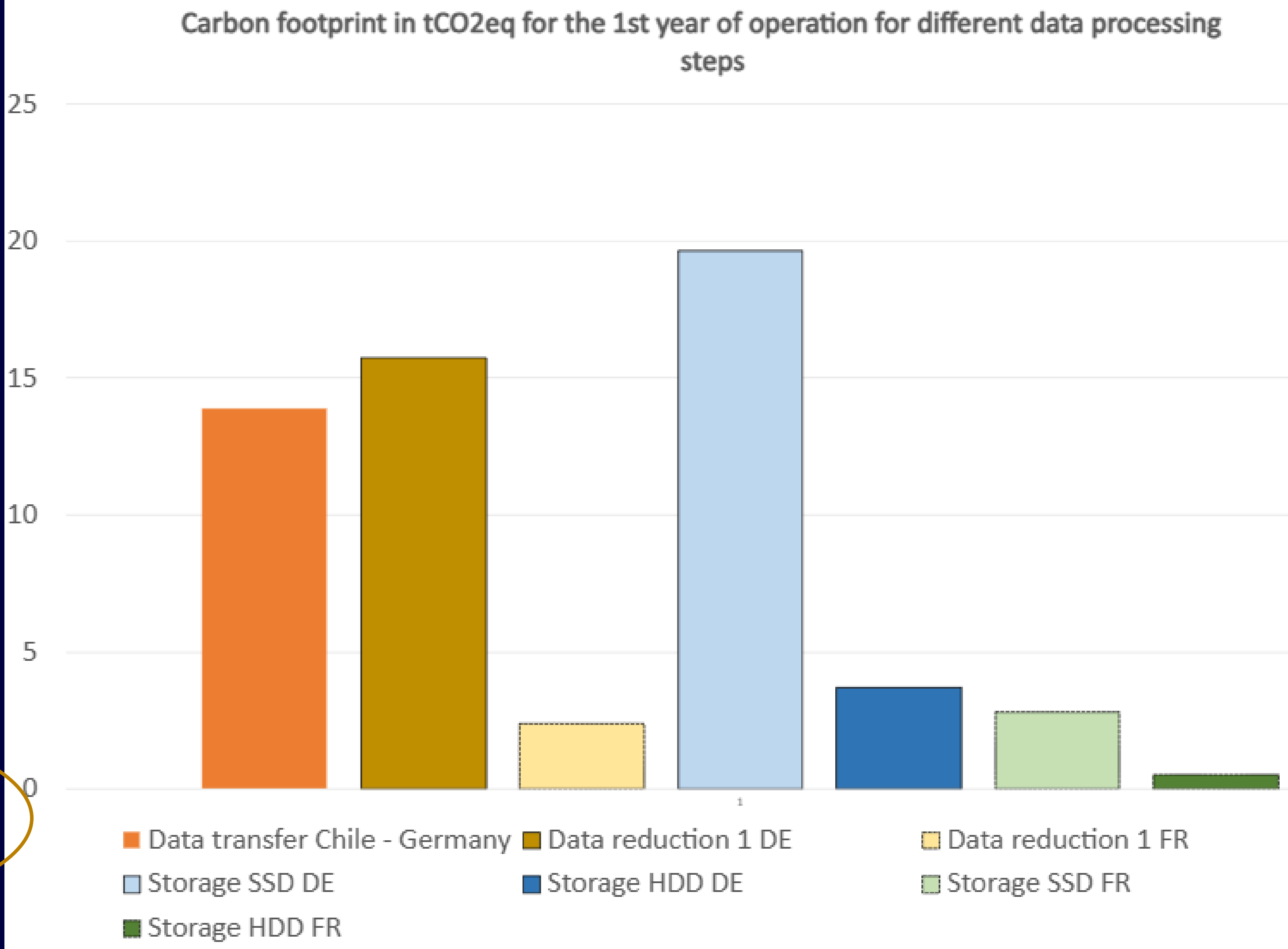
Carbon footprint of the steps

- **Step 2: Data transfer**
  Embedded impact of the optical fibers and shelters, using formulas & emission factors from Loygue et al. 2024
- **Step 3: Data reduction pipeline**
  Footprint quantified using the green algorithm (Lannelongue et al. 2021)
- **Step 4: Data storage**
  Different storage types with kWh/TB reported (Seagate)



Carbon footprint of the steps

- **Step 2: Data transfer**
  Embedded impact of the optical fibers and shelters, using formulas & emission factors from Loygue et al. 2024
- **Step 3: Data reduction pipeline**
  Footprint quantified using the green algorithm (Lannelongue et al. 2021)
- **Step 4: Data storage**
  Different storage types with emission factors from Seagate



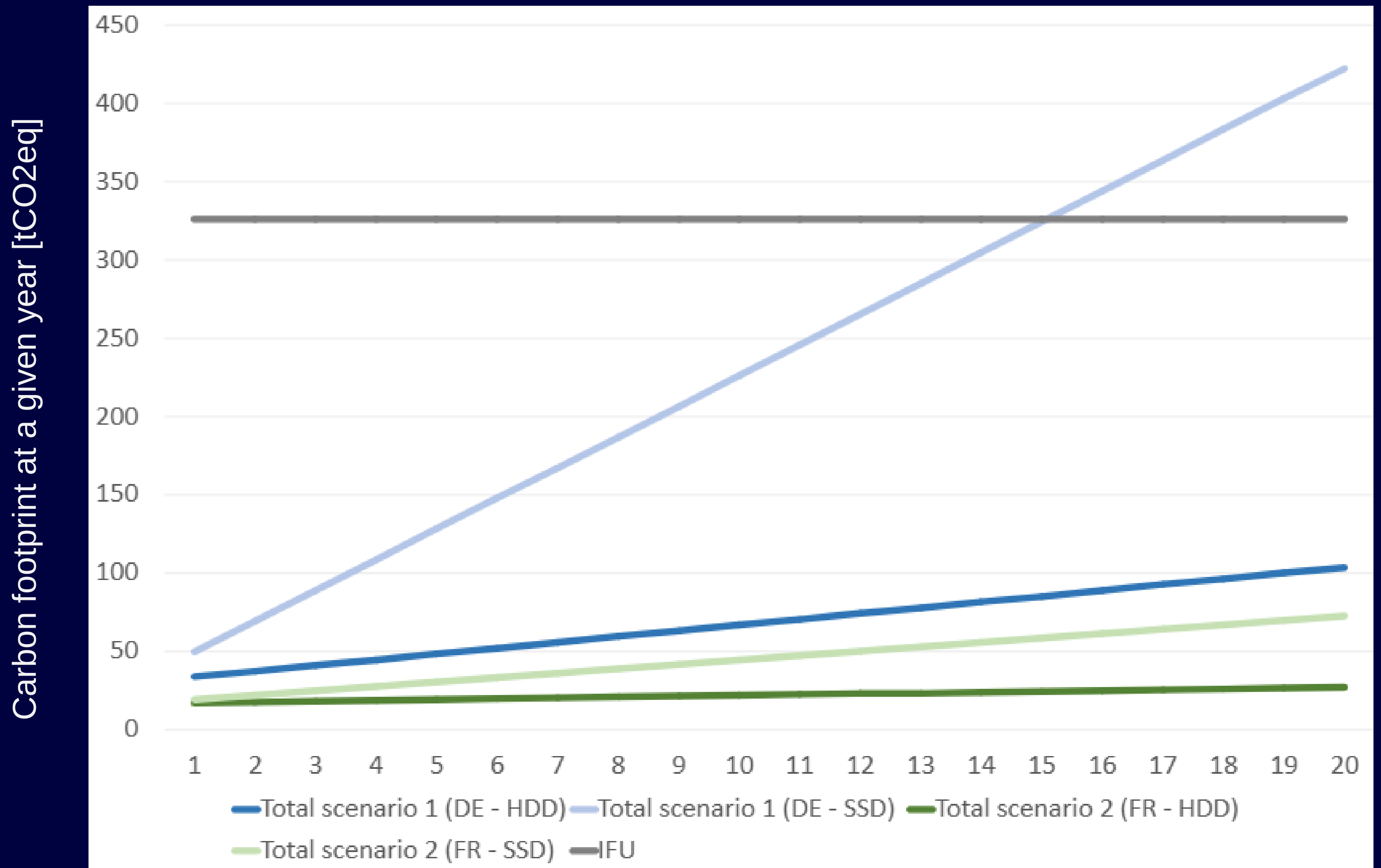
Carbon footprint of different data processing scenarios

- **Scenario 1**: Data stored in Germany
  - **HDD**
  - **SSD**
- **Scenario 2:** Data stored in France
  - **HDD**
  - **SSD**

Carbon footprint of different data processing scenarios

- **Scenario 1**: Data stored in Germany
  - **HDD**
  - **SSD**
- **Scenario 2:** Data stored in France
  - **HDD**
  - **SSD**

Carbon footprint at a given year of different scenarios for data processing compared to the cooling of the Integral Field Unit in tCO2eq



First results

- The carbon footprint linked to the cooling of instruments dominates over the one from data
- Data storage quickly dominates the carbon footprint of data processing over the data reduction and transfer
- Both location & type of hardware have a big impact and should be selected carefully

Future considerations & Open questions

- Only the first level of the data reduction pipeline has been included → underestimation of the carbon footprint linked to the data reduction
- Discussing rules on “hot” and “cold” archival will enable us to better select the hardware
- Should we decrease the data storage, even if it means increasing the computing power?
- What will be the landscape of data storage in the 2040s?

Conclusion

Funded by
the European Union

- Integrating the carbon footprint as a parameter in the design of future facilities is crucial to ensure they are conceived and built while minimizing the impact on the environment

https://wstelescope.eu/