Riding space-borne data with the TLS prototype.

M. G. Lattanzi (presenter), R. Morbidelli, D. Busonero, B. BucciarelliM. Crosta, A. Vecchiato, M. Carnerero, C. Raiteri

INAF-OATo

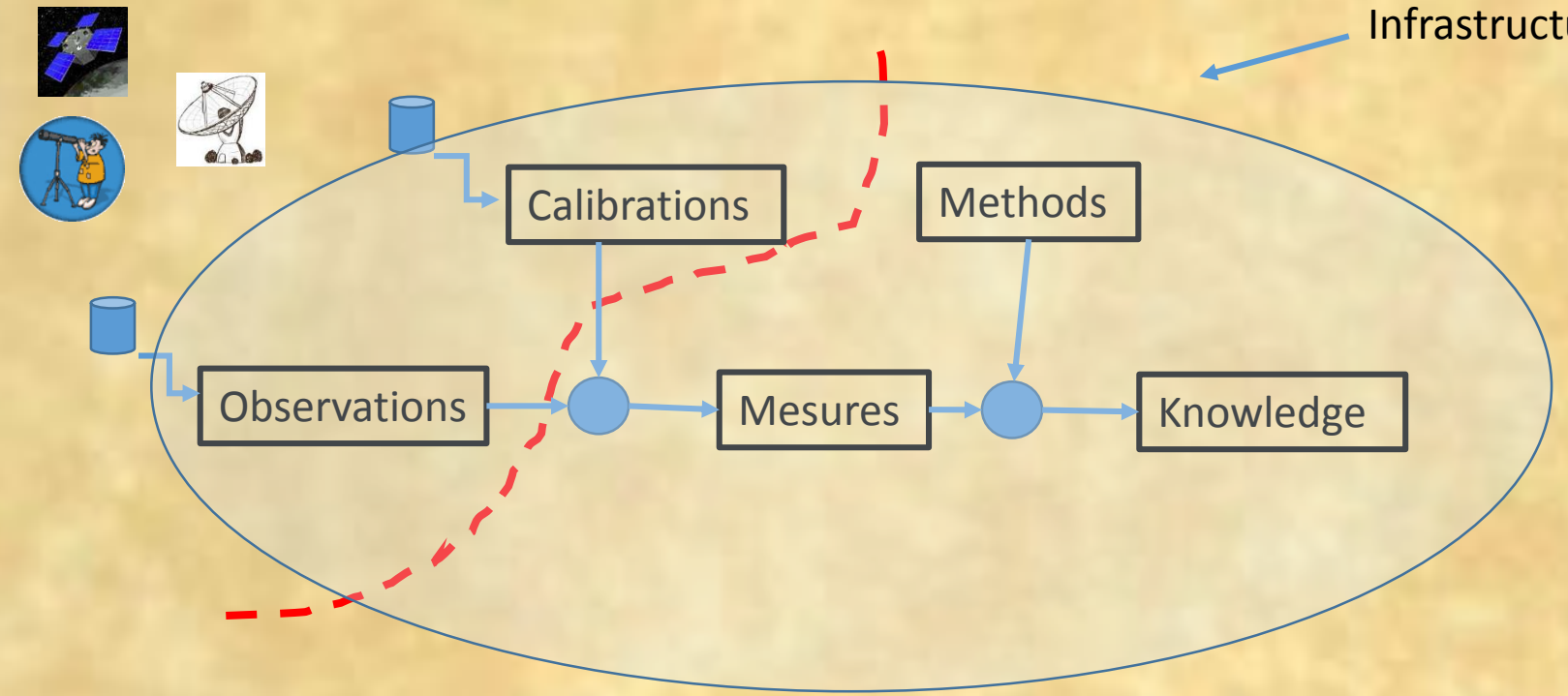
&

R. Messineo, A. Mulone

ALTEC

Observations (event acquisition), Measurements, Knowledge (interpretation)

«Knowledge comes from measurements, ……..»



Note: '**Calibrations**' includes methods & data (telemetry, house keeping data) & '**Methods**' includes simulations!!

Update on the status of the implementation of the TLS prototype project, part of the MITIC "premiale 2015" initiative (2017), at the Gaia DPCT.

(INAF, ASI & ALTEC)

General motivations behind prototyping the TLS concept:

“The sky is never the same even when an observation can be repeated ”

- **Probing/challenging the discovery space of data taken for a given program/mission means to exploit that data for issues (scientific and/or technological) not in the primary science case of that program/mission to:**
  - **objectively assess the true value of the associated archive for the different agencies/institutions potentially interested in its long-term preservation.**

Specific motivations behind prototyping the TLS concept on the Gaia archive @ DPCT:

- Addressing the intrinsic “discovery space” of the Gaia absolute all-sky survey, as DB-ed at the DPCT through experiments on the first PB-class (still highly non-static) archive of the optical celestial sphere available in Italy.
- Same as above, but in combination with other present and future archives (i.e., deep ground-based photometric and spectroscopic surveys, EUCLID,…)
- Other uses/applications (i.e., mission planning, space environments and instrumentation,….)

Implementation principles:

- Select experiments that best represent the motivations behind the TLS idea (cases requiring recalibrations or special data treatment, including cases requesting daily data as well): proving (or disproving) the paradigma of long-term preservation of data and (some/all ?) calibrating pipelines.

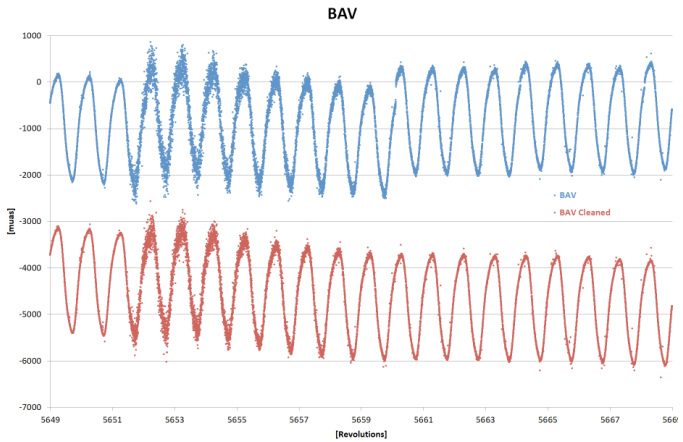
- Because of budget limitations, delay HW decisions until the very detailed planning of the representative experiments adopted (e.g., Data Requests) and detailed design of SW architecture finalized.

Expected results:

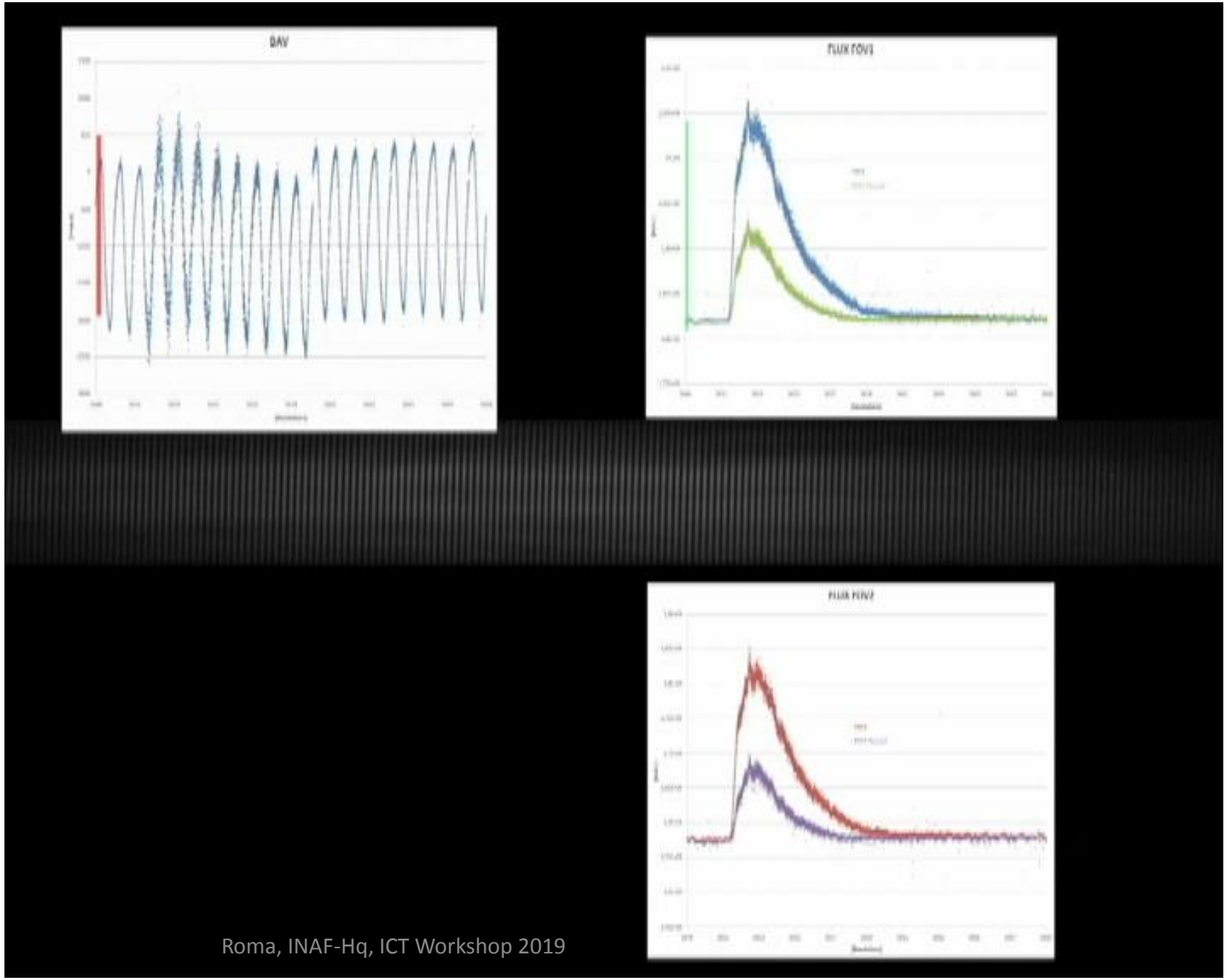
- Show successful non-standard scientific and technical exploitation of Gaia RepDB

- Final Report on curation and preservation of data and methods, and on relevance of the human factor (e.g., data scientists)

Sep 2017 CME as seen by Gaia metrology (BAM) system



(Courtesy of A. Riva)



The selected scientific experiments

(Science is engineering investigations as well!)

1. **Discovery and characterization of AGNs (Carnerero, Raiteri, Busonero, Lattanzi, Morbidelli). In particular of Blazars, i.e., AGN's with unpredictable variability scales of hours-days. Requests: single-CCD photometry (both instrumental and calibrated fluxes) in G-band (AstroElementary) and Bp, Rp bands (PhotoElementary). Special data processing possible (TBC).**

2. **Support multi-wavelength multi-messenger astronomy: search/characterization of Gravitational Wave sources (Crosta, Branchesi, Brocato, Bucciarelli, Busonero, Lattanzi, Morbidelli, Vecchiato).**
   Unique requirement: fast reaction times.
   Retrieve all Astrometric and Photometric epoch data over square degrees. Special processing of astrometric data including images (AstroObservation = there are 1,200 billions observations, i.e. single CCD transits, in 5 years. These are small images, like pixel "cutouts", centered on detected objects)

3. **Space science & engineering: effects of L2 environment on digital devices aboard (ALTEC+INAF-OATo, TECSEL2).**
   Utilizes also telemetry (house-keeping) data that have not entered the Gaia DPAC processing.

2. Data Extraction

(what is essential for science exploitation?)

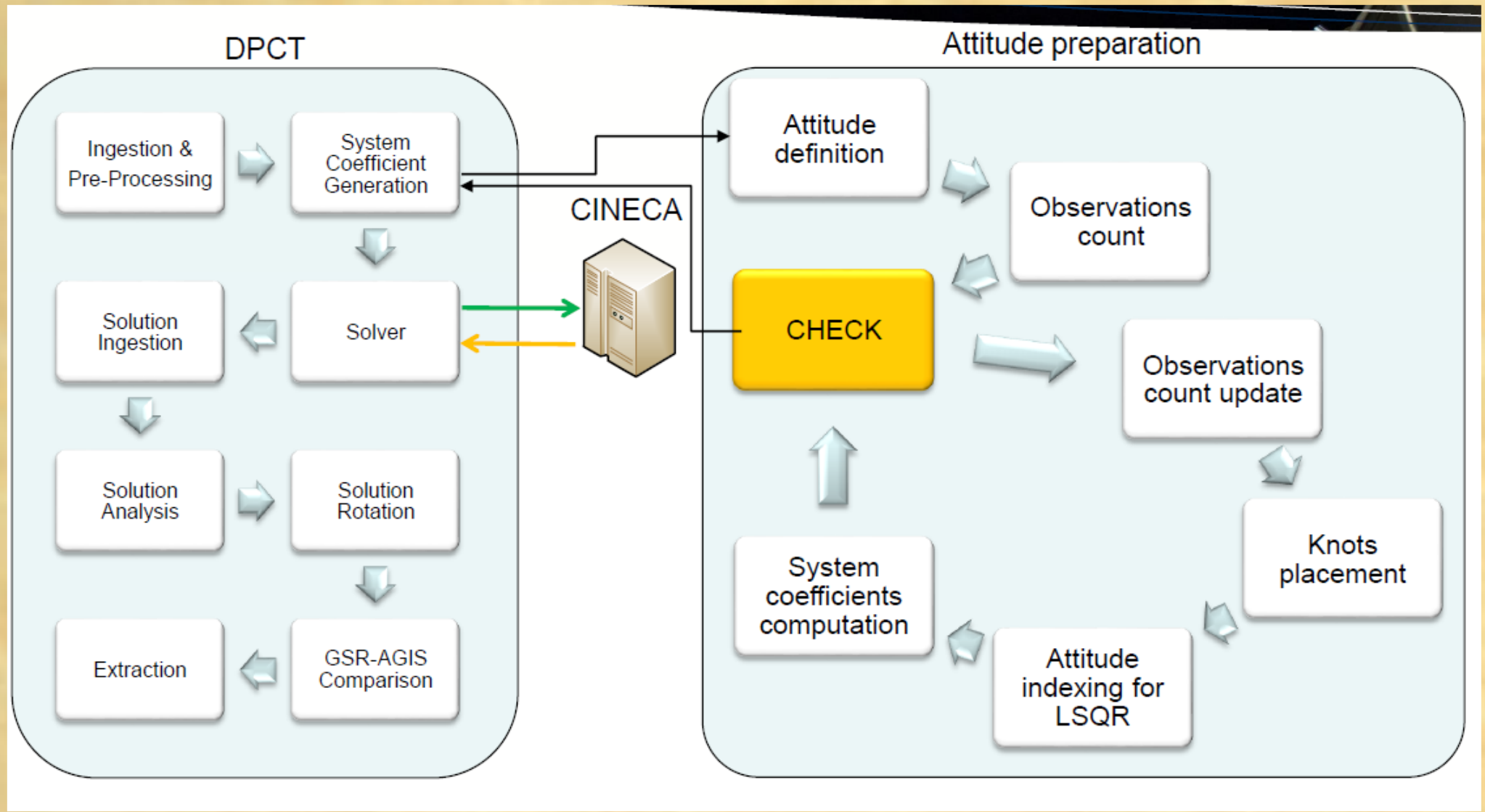
- **Queen queries for the 3 experiments through "Data Requests", that are not user-friendly and will not be such for the prototype**. Besides **CompleteSource** (much like the catalog delivered at Data Releases), extract:
  - Transit data (AstroObservation, PhotoObservation, SpectroObservation, making up most of the Gaia raw data) and their corresponding direct "meta-data" (AstroElementary, PhotoElementary, SpectroElementary)
  - Calibration data (CDB)
  - Orbital and satellite attitude info
  - ……

3. Data Transformation

- **Transformation to FITS and/or CSV (ASCII standard) files**
- **Export methods (?) for specific needs (e.g., calibrated fluxes; more accurate astrometry)**

The items above already call for considerations on curation and preservation of data and, possibly, methods.

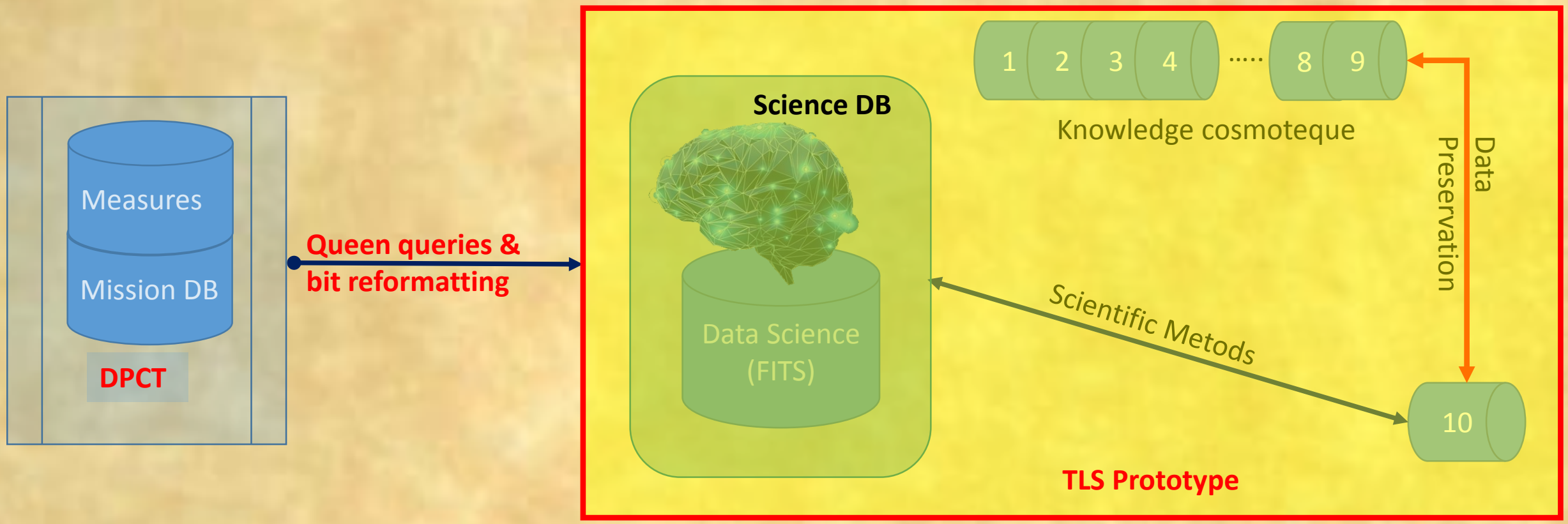
Example. Global (Cyclic) astrometric calibration (GSR@DPCT)



4. How?

(Where are we?)

“Lift and Shift” approach within Oracle (already tested)



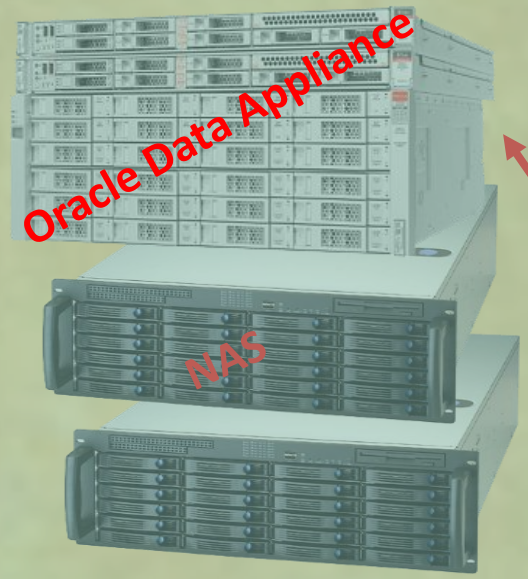
TLS prototype: building a Proof of Concept (POC)

Today

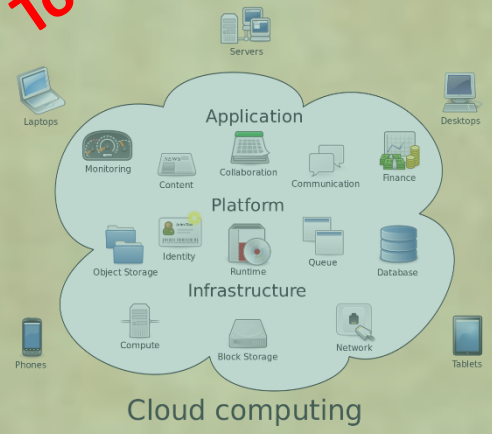
Gaia DPCT (constraints):
H/W – Gaia GBIN Repository
S/W – Gaia Libraries JAVA
DBMS – Oracle 18C
Network – Close

Queen queries & bit reformatting

TLS Ecosystem (POC):
H/W – ODAX4 + Storage + Cloud
S/W – Fortran, C++, Java, Piton..
DBMS – Oracle products galaxy
Network – Open



Tomorrow



Conclusions so far:

- ❑ **TLS prototype feasible within given funding framework and resources, and well worth the efforts!**
- ❑ **Delays mostly due to the increased workload imposed by the Gaia schedule especially after implementation of mission extension on ALTEC and INAF personnel**
- ❑ **Initial trial extractions for Exp. 1 already underway, while Exp. 2 will use known GW antennas discoveries as study cases.**
- ❑ **DB, processing, and storage HW identified**
- ❑ **Recruiting of a Data Scientist underway**