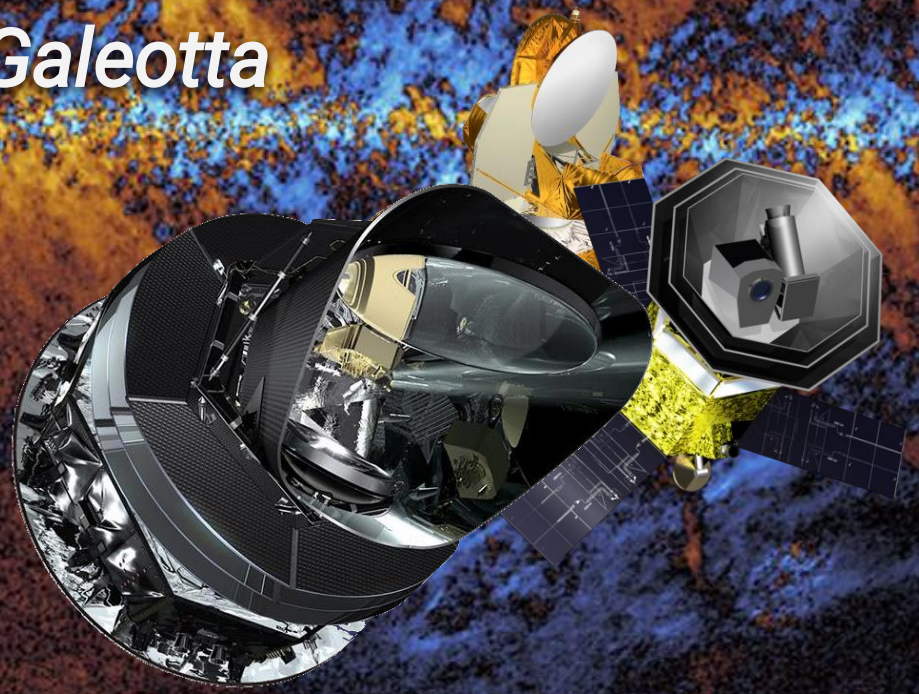
WP 2
Data Selection and Flagging

Samuele Galeotta



BeyondPlanck Final Review, December 10, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776282

Summary

- WP2 serves as starting point to select which data, at the timelines level, will be used in our analysis.

- It will, based on pre-defined criteria, flag the data that should be excluded like maneuvers period, gain changes in the data acquisition electronics that cause saturation, abrupt changes in voltage outputs caused by gain fluctuation.
  - Note: maneuvers where flagged but not excluded in BP analysis

- The main objective of this WP is to give to the subsequent WPs consolidated data.

WP2 Deliverables

- Prototype version of data flagging module at Month 3
  Delivered on May 31st, 2018
- Final version of data flagging and selection module at Month 6
  Delivered on September 14th, 2018
- Added a compression module for optimization

WP2 Software Components

WP 2 consists of 3 software components:

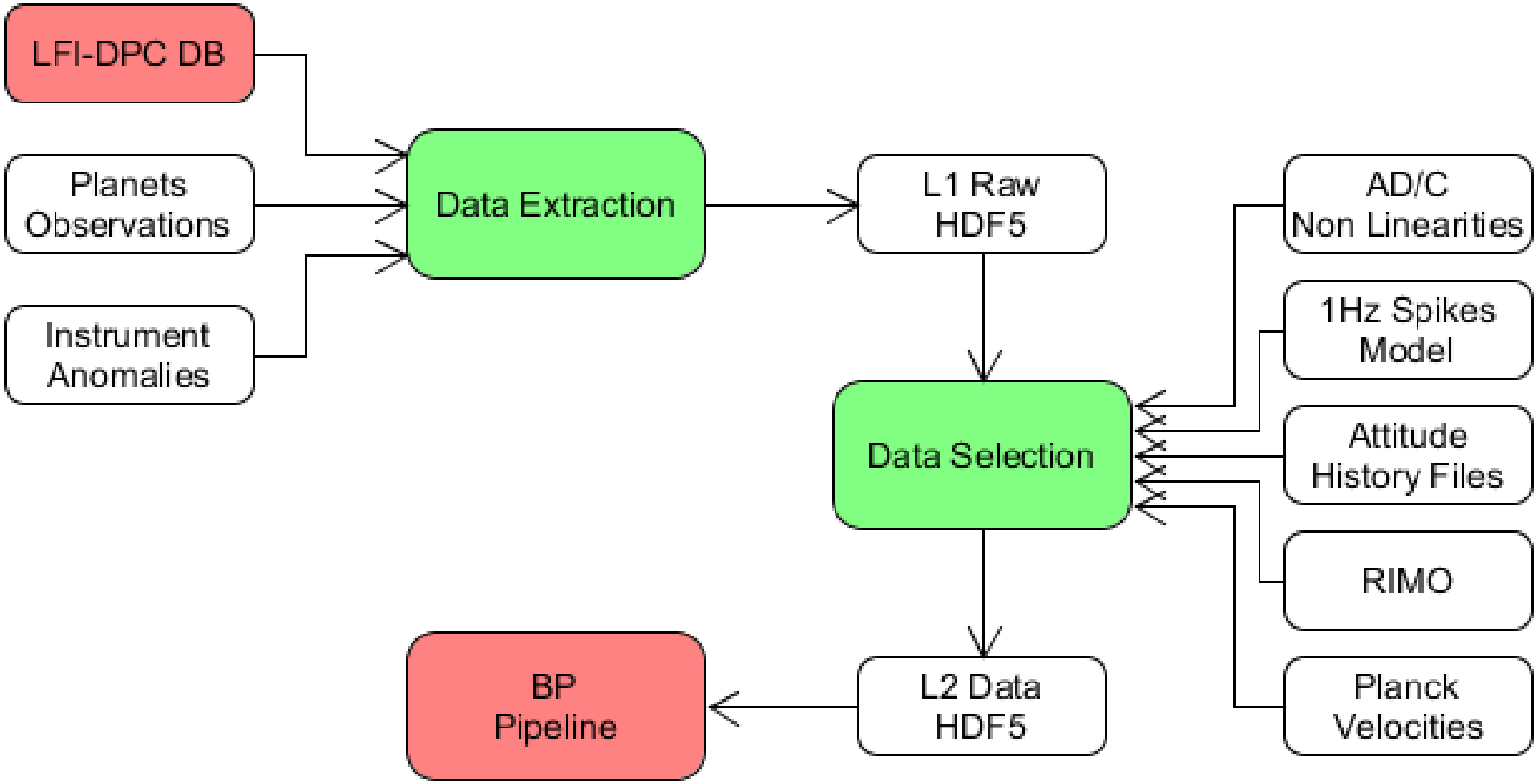
- Data Extraction: retrieve data from Planck LFI DB
- Data Selection: systematics correction, flagging, differentiation and detector pointings
- Data Compression: compression of input data using Huffman coding

Output data from Data Compression are the input data for the Gibbs sampler and are part of the deliverable.

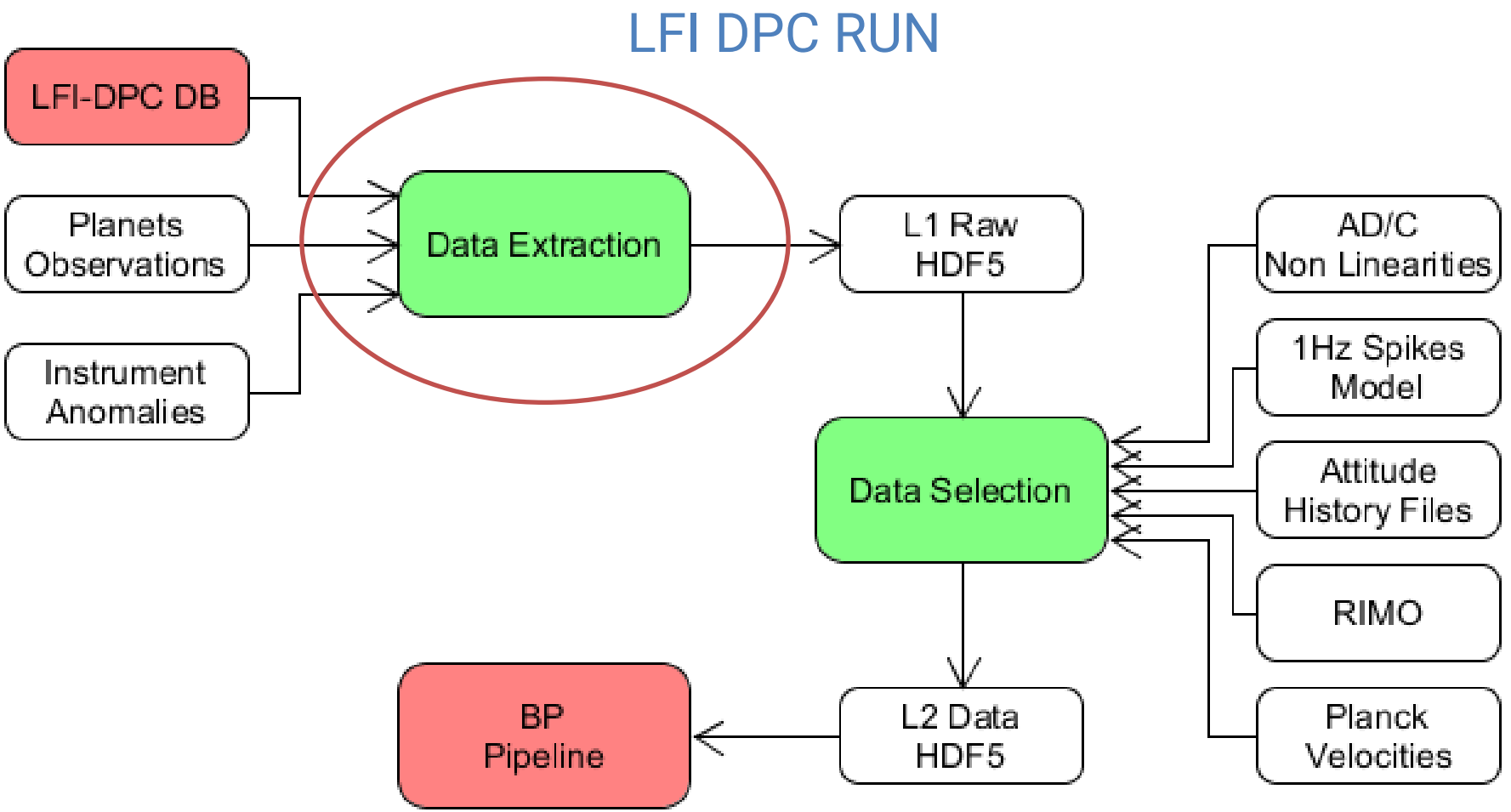
WP2 Criteria Definitions

- Definition of interfaces between software components
  - HDF5 format for easy management of big sizes and portability
  - Interfaces build to reduce data duplication
- Definition and distribution of operations to reduce repetition of the same computations
- Operations defined by instrument history in Data Extraction
  - Planet Flagging
  - Anomalies
- Operations defined by algorithm chosen in Data Selection
  - Systematic corrections
  - Detector pointing computation
  - Data differentiation

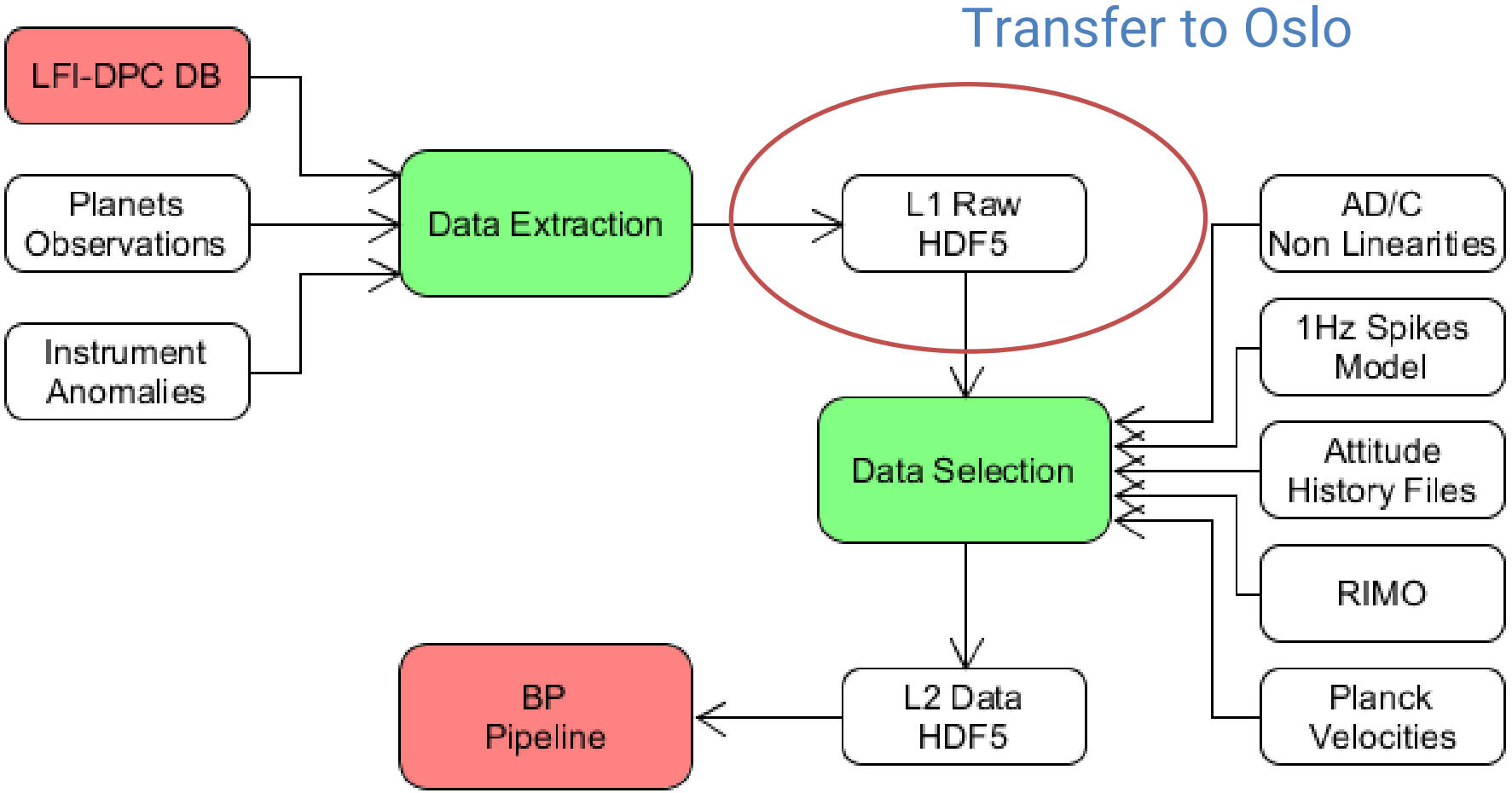
WP2 Design


LFI-DPC DB
Planets Observations
Instrument Anomalies
Data Extraction
L1 Raw HDF5
AD/C Non Linearities
1Hz Spikes Model
Attitude History Files
RIMO
Planck Velocities
Data Selection
L2 Data HDF5
BP Pipeline

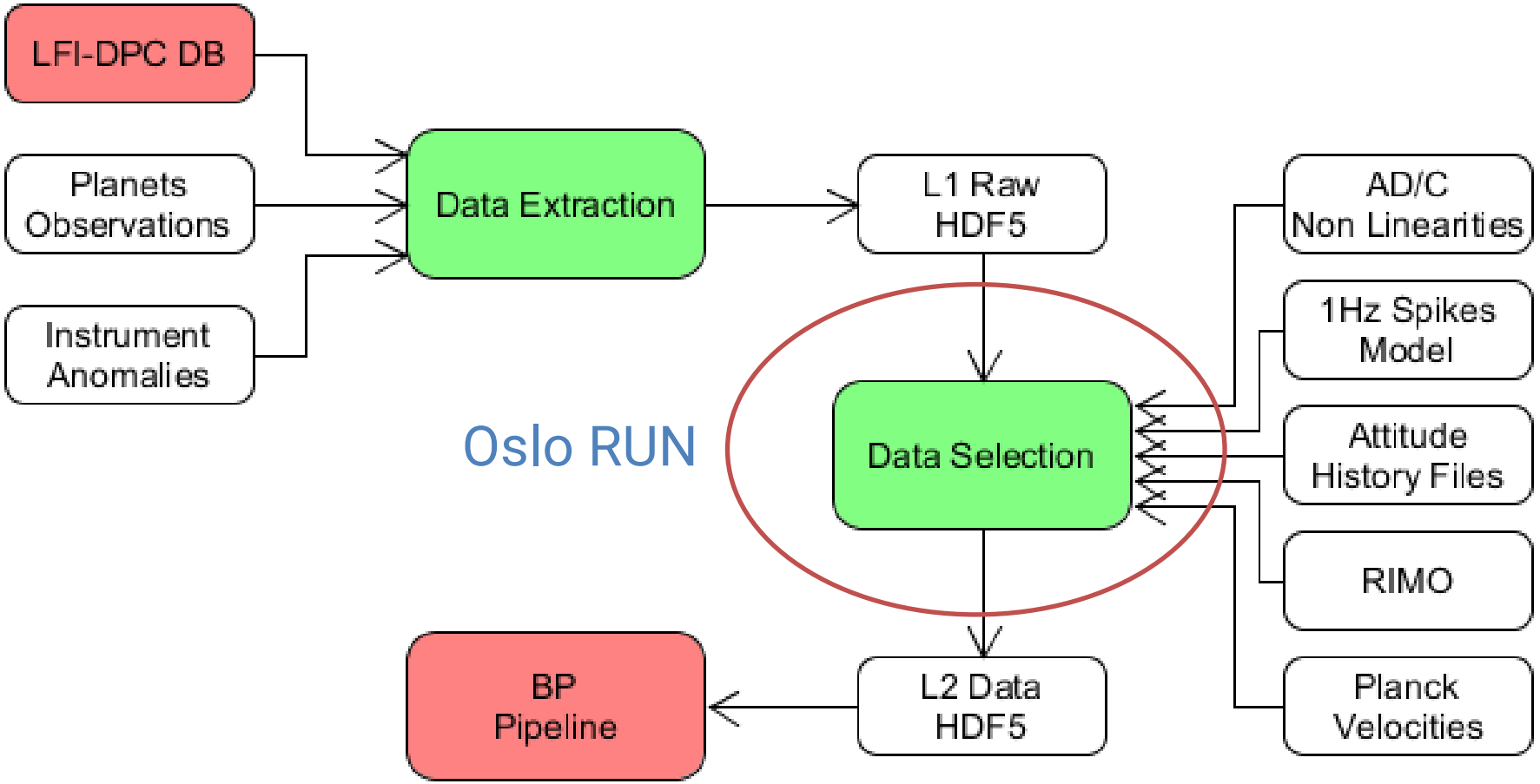
WP2 Design


LFI DPC RUN
LFI-DPC DB
Planets Observations
Instrument Anomalies
Data Extraction
L1 Raw HDF5
AD/C Non Linearities
1Hz Spikes Model
Attitude History Files
RIMO
Planck Velocities
Data Selection
L2 Data HDF5
BP Pipeline

WP2 Design

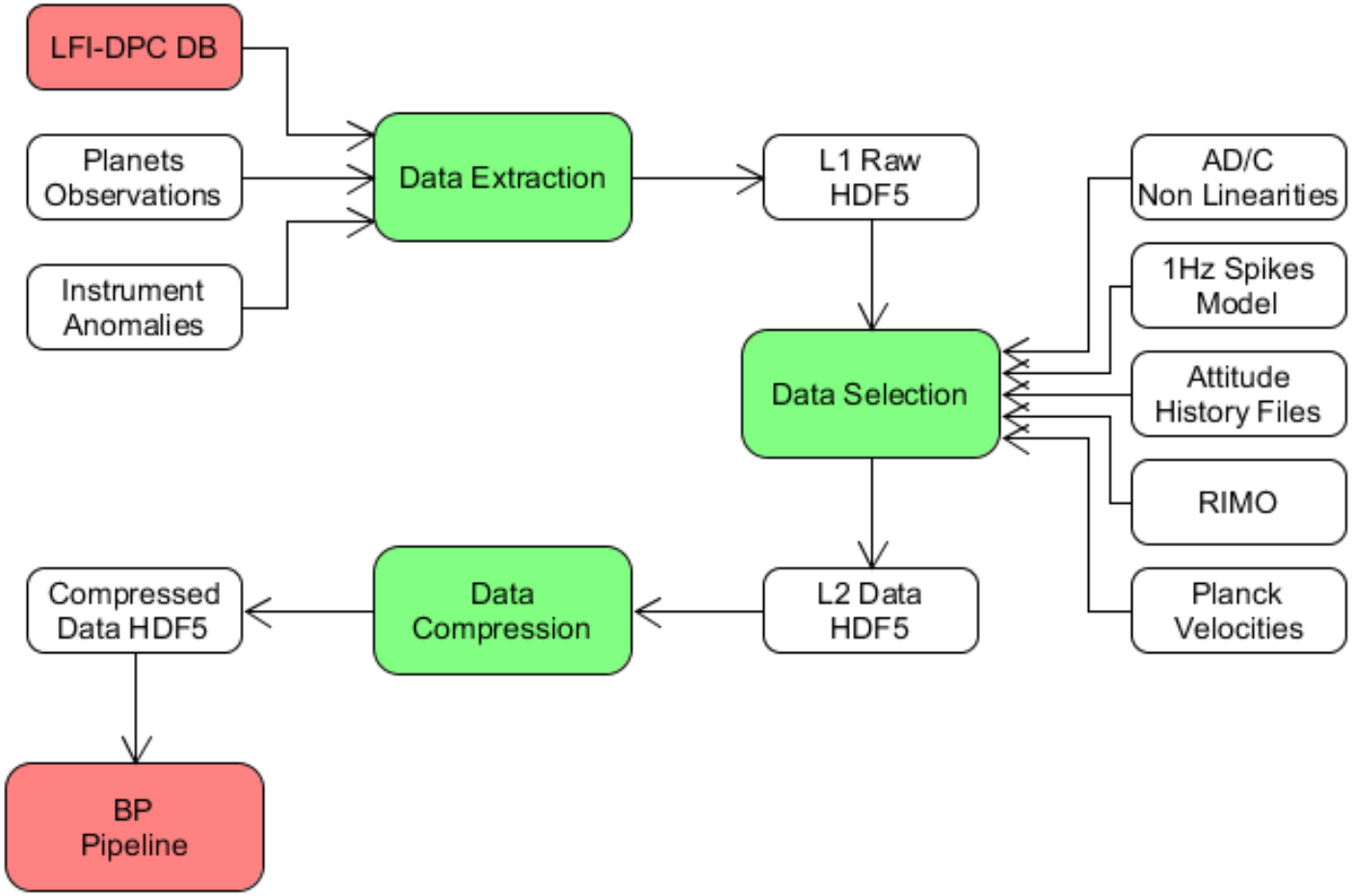

Transfer to Oslo
LFI-DPC DB
Planets Observations
Instrument Anomalies
Data Extraction
L1 Raw HDF5
AD/C Non Linearities
1Hz Spikes Model
Attitude History Files
RIMO
Planck Velocities
Data Selection
L2 Data HDF5
BP Pipeline

WP2 Design


LFI-DPC DB
Planets Observations
Instrument Anomalies
Data Extraction
L1 Raw HDF5
AD/C Non Linearities
1Hz Spikes Model
Attitude History Files
RIMO
Planck Velocities
Oslo RUN
Data Selection
L2 Data HDF5
BP Pipeline

European Commission

Final design with the addition of data compression



Beyond PLANCK

WP2 Data Extraction

- The LFI DPC pipeline was dependent from a proprietary DB (ORACLE).
- Extraction from the LFI DPC DB and instrument operations
  - Read data and fill gaps in data
  - Flag planets
  - Flag instrument anomalies
  - Save raw timelines in HDF5 format

WP2 Data Selection

- Implemented new interfaces and converted to C++11
  - Read timelines
  - Applies ADC non linearities and 1Hz spikes corrections
  - Compute detector pointing
  - Compute differentiated data
  - Save differentiated data and detector pointings in HDF5 format

- Tested and compared with the original data

WP2 Data Compression

Item	Explanation	30 GHz		44 GHz		70 GHz		Total Ratio
		Raw (MB)	Compressed (MB)	Raw (MB)	Compressed (MB)	Raw (MB)	Compressed (MB)	
TOD	Dtype reduction	90.3	45.2	193.9	97.5	656.2	328.1	0.5
Pixel number	Healpix, Differencing + Huffman Coding	180.5	9.8	387.8	17.5	1312.4	69.7	0.05
Psi	Discretization, Differencing + Huffman Coding	90.3	5.2	193.9	9.6	656.2	24.8	0.04
Flag	Differencing + Huffman Coding	45.1	2.7	96.9	5.9	328.1	10.3	0.03
Time	Runtime Extrapolation	135.4	0.0003	290.8	0.0003	984.3	0.0003	6.3e-7
Scalars	Included from RIMO	0	0.004	0	0.007	0	0.013	∞
Huffman Indexes		0	1.37	0	2.1	0	2.6	∞

- Needed to optimize I/O and RAM usage
- Pixelization of detector pointing to convert into integers
- Differencing reduce complexity
- Huffman Coding compression applied
  - Lossless compression technique
  - Binary representation
  - High frequency numbers shorter representation

WP2 Person Month Effort

Name	EU Fund	In Kind
INAF-OATS	13	0.9
TOTAL	**13**	**0.9**
BUDGET	**13**	

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776282

- "*BeyondPlanck*"
  - COMPET-4 program
  - PI: Hans Kristian Eriksen
  - Grant no.: 776282
  - Period: Mar 2018 to Nov 2020

Collaborating projects:

- "*bits2cosmology*"
  - ERC Consolidator Grant
  - PI: Hans Kristian Eriksen
  - Grant no: 772 253
  - Period: April 2018 to March 2023
- "*Cosmoglobe*"
  - ERC Consolidator Grant
  - PI: Ingunn Wehus
  - Grant no: 819 478
  - Period: June 2019 to May 2024