LOFAR2.0 Data Processing & Management Busy Week

R. F. Pizzo
Head Science Data Centre Operations

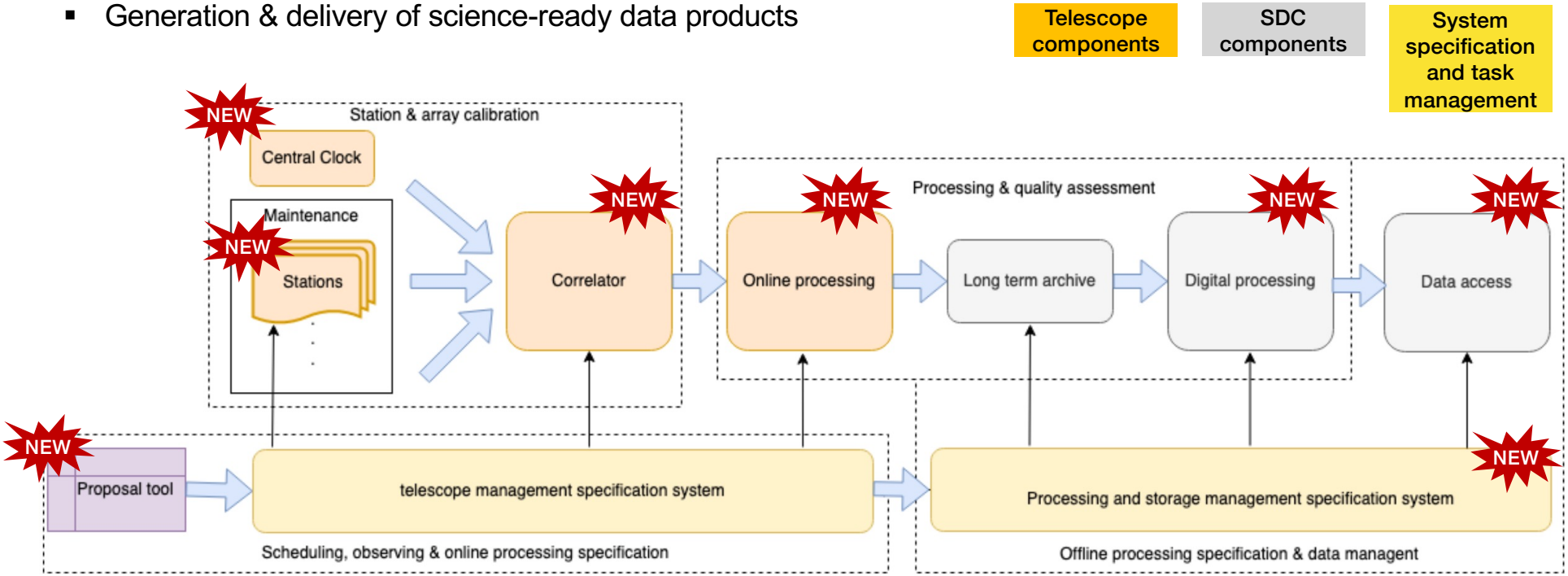
with contributions from BW participants

30 October 2024

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR2.0: New Challenges

- LOFAR2 will:
  - Double the number of active LBA antennas
  - Realize simultaneous HBA and LBA observations
  - use the Megamode – interferometric and TAB data products at the same time
  - Generation & delivery of science-ready data products



The LOFAR2.0 Observing Program

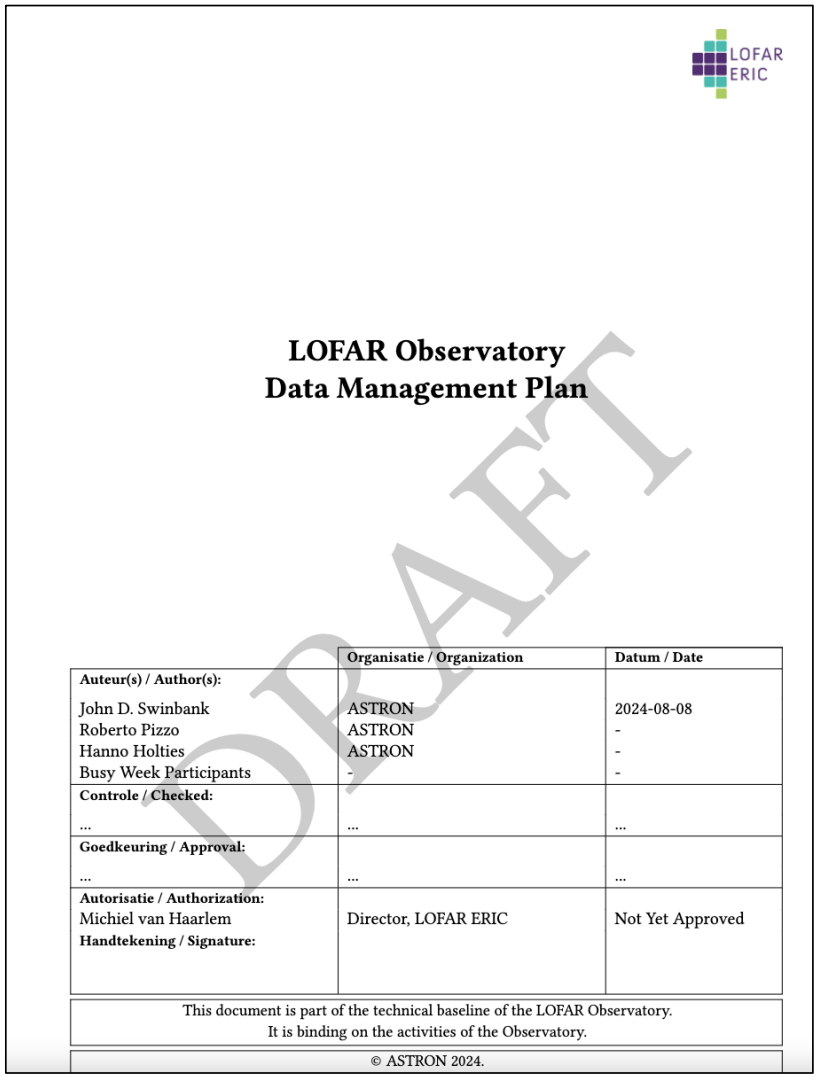
B. Hare	L2LP_001	lightning
L. Morabito & P. Best	L2LP_002	extragalactic
M. Arias	L2LP_003	galactic
A. Rowlinson	L2LP_004	transients and pulsars
S. Bouma, K. Mulrey	L2LP_005	cosmic rays
T. Shimwell	L2LP_006	extragalactic
A. Basu & D. Schwartz	L2LP_007	galactic/extragalactic
F. de Gasperin	L2LP_008	galactic/extragalactic
L. Koopmans	L2LP_009	extragalactic
E. Vardoulaki	L2LP_010	extragalactic
P. Zucca	L2LP_011	space weather
H. Vedantham	L2LP_012	planets
J. Hessels	L2LP_013	transients and pulsars
V. Heesen	L2LP_014	extragalactic
C. Tiburzi	L2LP_015	transients and pulsars

- 15 Large Programs proposals
  - Requesting ~ 48000 observing hours over 5 years
  - Oversubscription ~ 1.5
  - 70% of available time
- Allocations: PC
- Open Skies (30% of available time)
- Archival science

Busy Week Goals

- How the LOFAR2.0 observing program will be realized.
  - Large programmes **and** open skies **and** archival science.
- Establish a common understanding around the LOFAR Observatory:
  - Establish a common set of goals within the community;
  - From the goals, derive a way of working together;
  - What constitute a *success*?
  - Capture this within a written framework presented to the LOFAR ERIC for approval.
- Set expectations

- Work started and will continue after the Busy Week



LOFAR ERIC

LOFAR Observatory
Data Management Plan

DRAFT

	Organisatie / Organization	Datum / Date
Auteur(s) / Author(s):		
John D. Swinbank	ASTRON	2024-08-08
Roberto Pizzo	ASTRON	-
Hanno Holties	ASTRON	-
Busy Week Participants	-	-
Controle / Checked:		
...	...	...
Goedkeuring / Approval:		
...	...	...
Autorisatie / Authorization: Michiel van Haarlem **Handtekening / Signature:**	Director, LOFAR ERIC	Not Yet Approved

This document is part of the technical baseline of the LOFAR Observatory.
It is binding on the activities of the Observatory.

Working Groups

- 21-25 October
- 48 participants (37 in person)
- LOFAR2 LP PI's + people involved in the realization of the observing program
- Defined a set of Working Groups:
  - discuss various fundamental aspects of the realisation of the LOFAR2.0 observing program
  - Responsible for sections of Observatory framework description
  - 2 chairs (ASTRON-internal + external)
- People asked to join the WG most relevant to their interests.

Working Groups

- WG1: LOFAR Observatory structure and processes (chairs Pizzo + Hardcastle)
  - How do the various parts of the Observatory collaborate to produce and provide science-ready data products?
  - Who is responsible for what?
  - What sort of governance or other structures need to be in place?

- WG2: Supported modes (chairs Orru' + Mulrey)
  - What are the various modes in which LOFAR will operate?
  - What is needed for each mode to begin operations?
  - Who is responsible for delivering it?
  - What's the staged timeline for when functionality will become available?

- WG3: Data Products and ICDs (chairs Holties + van Weeren)
  - What data products will each of the observing modes produce?
  - What structure, format, etc, will they be provided in?
  - What metadata will accompany them?
  - How will quality be checked?
  - What information is needed to enable them to be ingested into the LTA?

Working Groups

- WG4 Pipeline Development Processes (chairs Shimwell + Timmermann)
  - How do we work together to develop code?
  - What standards are necessary for that code to be delivered and supported within the context of the Observatory?
  - Who is responsible for meeting those standards?
  - Who makes decisions about when code becomes operational, and when it is retired?
- WG5 Editorial (Swinbank + Arias)
  - Compile all the above into a coherent document.
  - Note that this group is expected to be small — perhaps just two people — and to source input from all of the groups above.

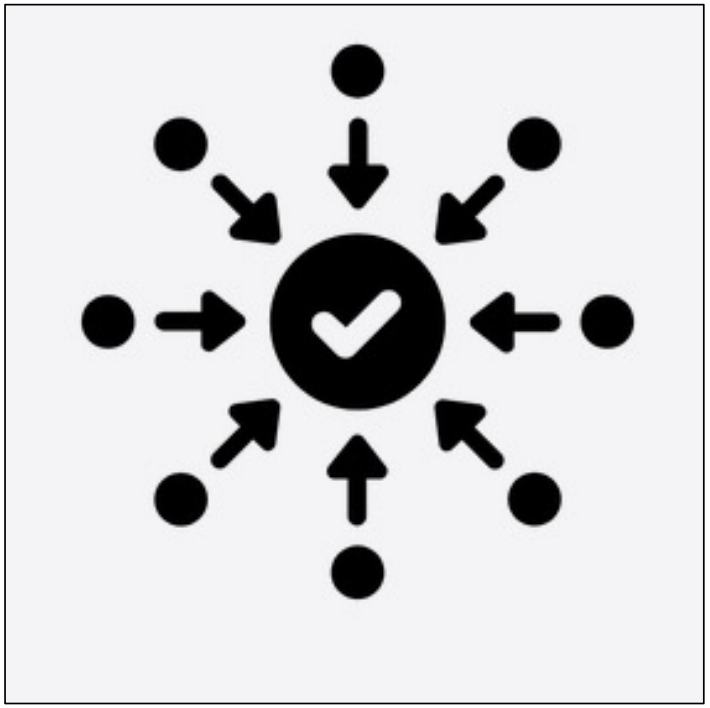
WG1 – LOFAR Observatory Structure and Processes

- Observatory: the sum of all the entities contributing to the generation of science-ready data products

- Data lifecycle:
  - All proposals should follow a process in which final data products are returned to the archive -> accessible LTA needed
  - Provenance of data products and linkage back to papers
  - List of supported observing modes and pipelines maintained

- For LOFAR2, the Observatory should move towards centralization of all services (bulk: timescale 2-3 years). In the meantime, a *federal* model is recommended
  - We need a definition of a process to become a federated partner (rights and responsibilities, including support)
  - Specification and commissioning of a system tracking processing across distributed systems

- Model for providing infrastructure:
  - Preference to a bulk compute/storage allocation

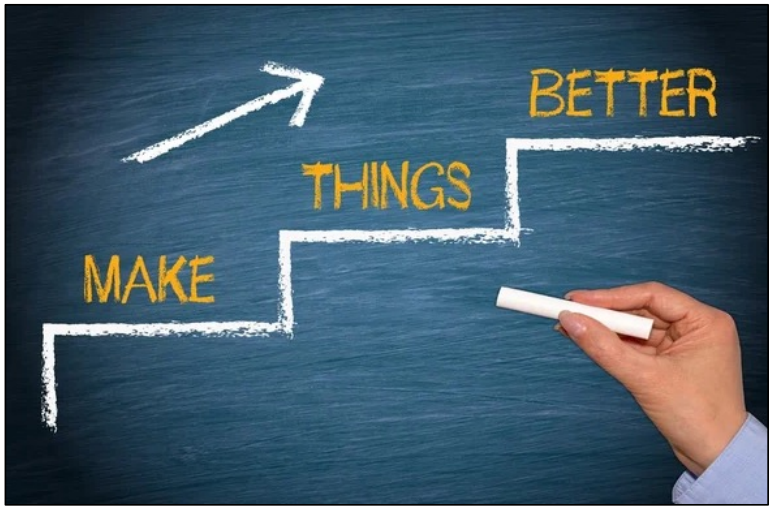


WG1 – LOFAR Observatory Structure and Processes

- User-shared support model should continue
  - Benefits should be made clear
  - On-boarding should be setup (manuals, trainings, certificates)
  - Expand/update data quality/management policies

- Open skies users
  - Point of contact within Observatory
  - Should be encouraged to make use of commissioned observing modes and production-ready pipelines
  - Support is a limited resource

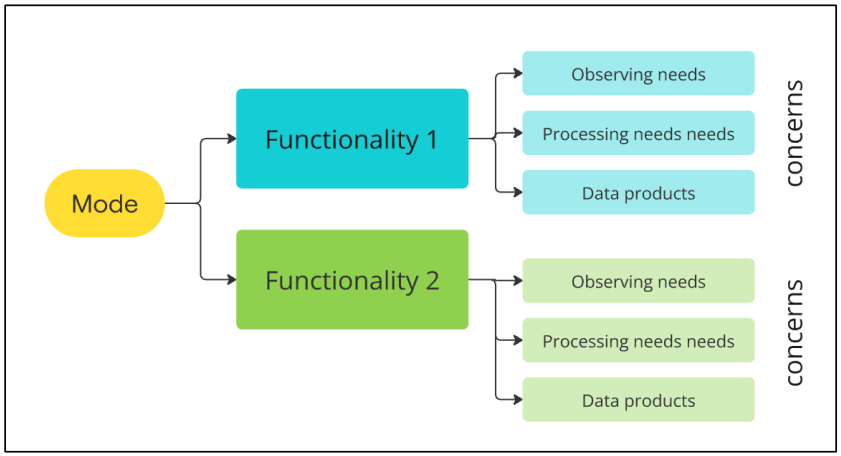
- Archive proposal functionality
  - People should be able to propose processing of archival data to enable new science



WG2: Supported Modes

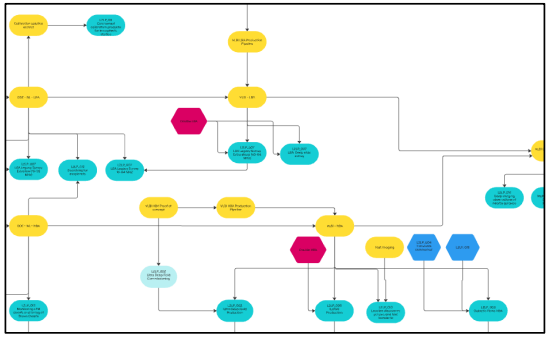
1) Inventory of different LOFAR2 observation modes, functionalities, and concerns

- Imaging
- Beamformed
- Transient buffer
- AARTFAAC
- Station Statistics
- EuroFlash



LINK [here]

2) Mapping of interconnectivity between modes



WG2: Supported Modes

3) Detailed spreadsheet of each L2LP: observing needs, status, and concerns

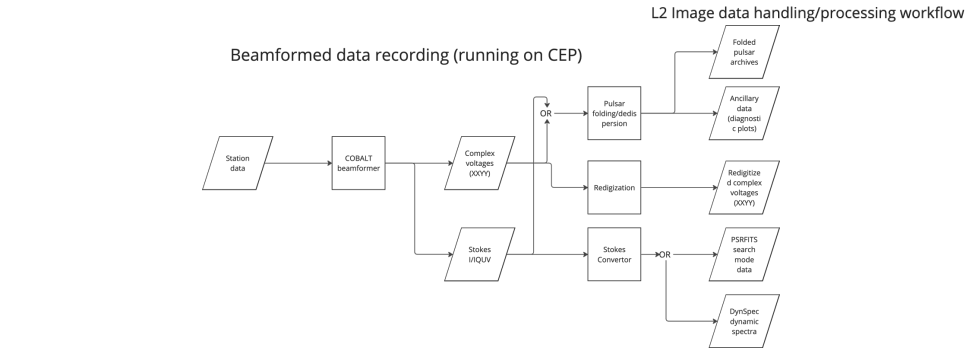
Observing mode	Processing needs	Data products	Resolution	Observational needs	Frequency	Retention
Imaging	Preprocessing	AARTFAAC visibilities	Single station	multi-beam (multi SAP)	LBA	until fully processed
Beam formed	LINC	images	ST	single-beam (single SAP)	HBA	until survey/program complete
AARTFAAC	DDcal NL station	cubes	CS	shared bandwidth	LBA+HBA	valuable to keep for post-program re-processing'
TBB	Pulsar Search Pipeline	catalogues	NL	daily calibrator		legacy/permanent
Commensal analysis	Dynamic spectrum	time-series	IS	dedicated calibrator		1-3 months
Imaging + Beam formed	TraP	autocorrelation	12CS	widefield survey		rolling buffer
Station XST	Euroflash	dynamic spectra		deep integration		intermediate (18 months)
Station Raw UDP	DDcal NL+IS station	XST station		target of opportunity		
	Cosmic Ray Pipeline	folded pulsar data		rapid response mode		
	Lightning Pipeline	antenna time series TBB		requested cadence		
	Subtraction imaging	light curves		double LBA		
	AARTFAAC imager	solution tables		double HBA		
	XST imager	Incoherent Stokes		nenufar widefield		
	WF Rmsynthesis pipeline	Coherent Stokes		nenufar full		
	PULP2 - dspsr folding	Complex Voltages		single antenna per tile		
	PULP2 - XXYY 8-bit redigit	raw data		piggy-back observing		
	PULP2 - stokes - 8-bit - ps	pre-processed visibilities		dedicated observing		
	PULP2 - dynspec	calibrated visibilities		commensal to observing		

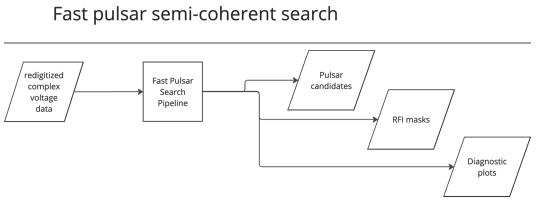
WG2: Summary of Recommendations:

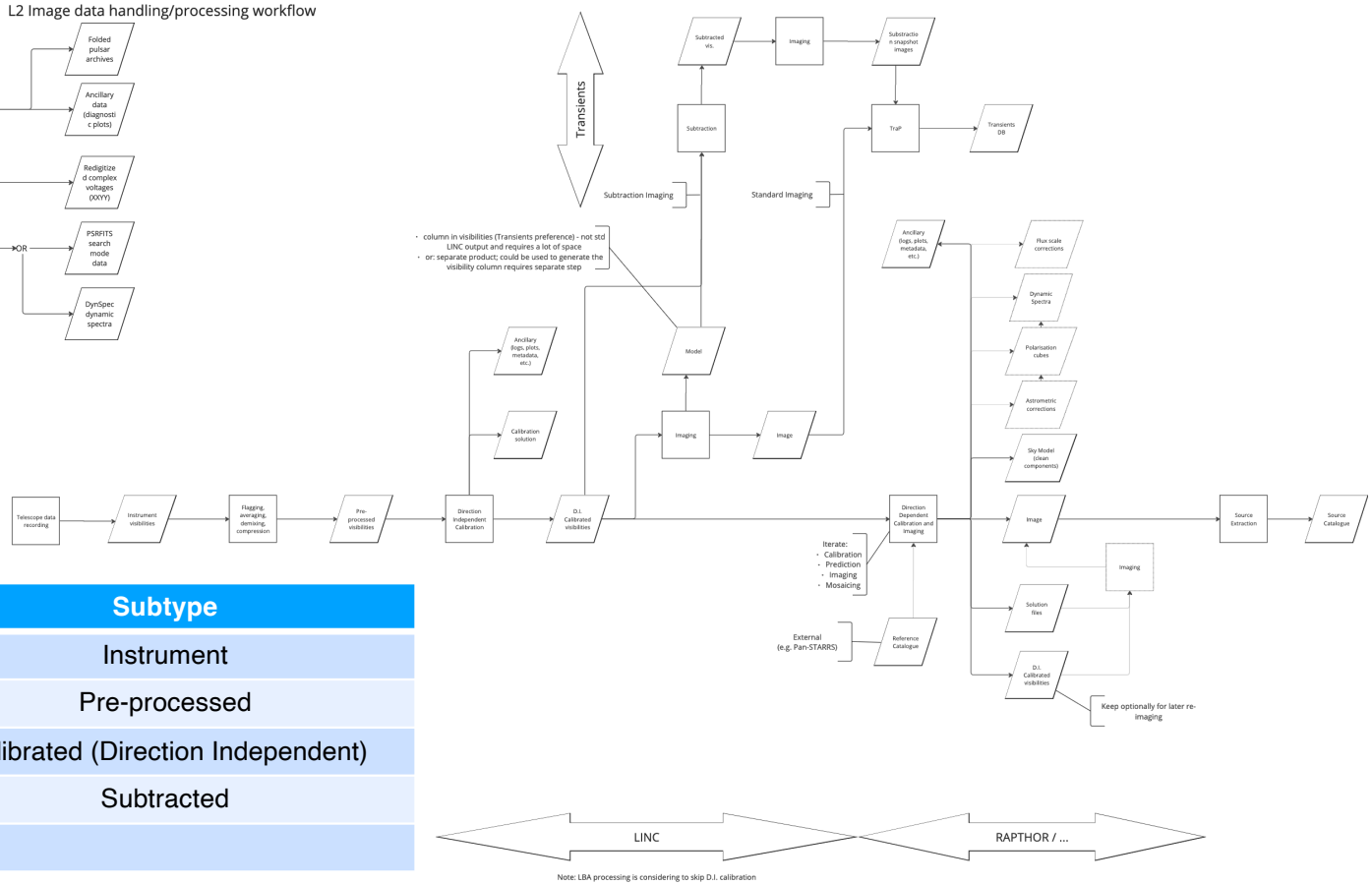
- Missing functionalities from L0 Dupplo requirements should be flagged
- Progress of all observing modes and concerns and status of all modes should be monitored
- Modes should be implemented to offer end-to-end process for L2LP
- status of the timeline and functionalities is communicated
- Resources and competences should be resourced from the Observatory
- The commissioning lead has responsibility for ensuring L0 is reached for each mode.
- Concerns have been raised for certain functionalities that require further investigation. These will be addressed by the commissioning team.

WG3 – Building the LOFAR2.0 data inventory

Created initial data handling diagrams







Data Product Type	Subtype
Visibilies	Instrument
	Pre-processed
	Calibrated (Direction Independent)
	Subtracted
…	

WG3 – Recommendations & Follow-ups

- Require for each project an agreed Data Management Plan
  - DMP to clarify the relevant/specific data handling flow
    - Data Products needed and delivered back
    - How and where these are processed and stored
  - Start a working group to set requirements and deliver an example/template
- For all supported data products there must be an applicable specification, including:
  - Include applicable formats & standards
  - Include specifying metadata for/about (and how the metadata is to be provided)
    - Provenance
    - Characterisation
    - Quality
    - Discovery
  - Start working groups to write the specifications
- Deliver an archive with
  - Standardized search methods for selected parameters (incl. location, type of data, project, etc.)
  - Effective views (incl. public availability, retirement, data collection(s) data belongs to, etc.)

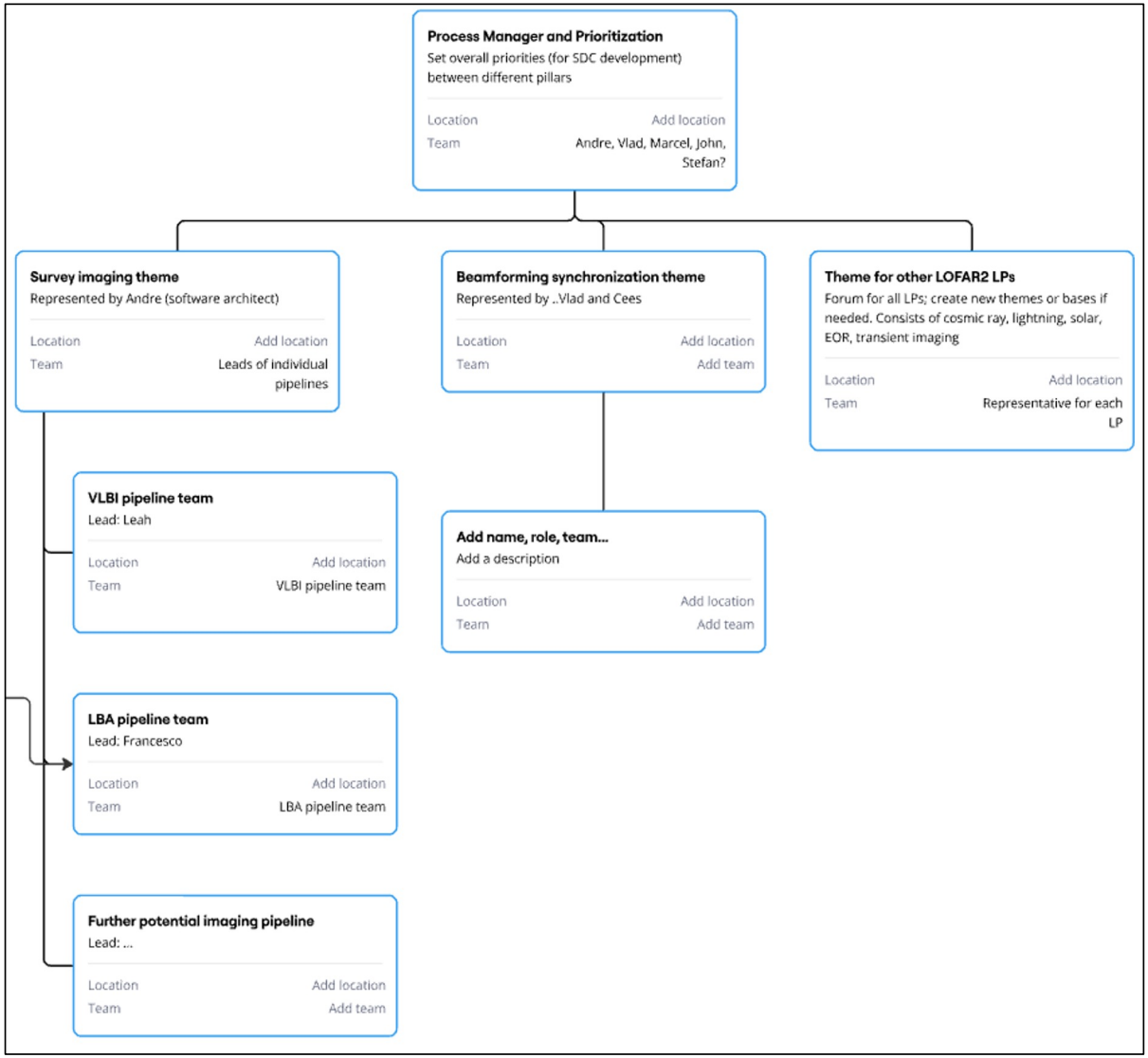
WG4: Pipeline Development Processes

Key recommendations:

- Pipeline workflow development is expected to continue to occur within pipeline teams such as the VLBI or LBA teams for imaging. Whereas component development is largely conducted by development teams (e.g. RAPTHOR, SCHAAP etc).
- We recommend formalising the regular technical meetings that are currently held between pipeline leads, the LOFAR software architect and development team representatives. These meetings will define prioritisation of development of software components. Care must be taken to ensure all pipeline teams are fairly represented and provided with development opportunities.
- The LOFAR software architect and development team representatives are responsible for, within resource limitations, ensuring high priority component development is completed. The pipeline team leads are responsible for representing and communicating with their pipeline team.
- Pipelines should be transferred from an experimental phase to a production phase through a proof-of-concept phase in which pipeline developers collaborate with software developers to reach the requirements for production. Increasing the links between development teams and pipeline teams hopefully helps transfer pipelines from experimental phase through to proof of concept phase and finally into production level code.

Structure of technical meetings:

- Pipelines are divided into development themes
  (e.g. survey imaging, beamformed)
- Each pipeline is represented by technical expert(s), who is/are the point-of-contact for the community.
- Each theme regularly meets with their relevant pipeline team leads, LOFAR software architect and development team leads to discuss development.
- Global prioritisation of software components is discussed internally and communicated back to the representatives.



Pipelines offered as an observing mode

Key recommendations:

- Pipelines can be either centrally (Observatory) or federally (externally) supported and both types can be offered as a observing mode.
- Centrally supported pipelines have a larger set of standards and are maintained and run by the Observatory.
- Federally supported pipelines are maintained and run by a dedicated support team or individual (this may include legacy pipelines or pipelines developed by pipeline teams or large programmes that are moving towards centralisation). There must be a guarantee of support (e.g. maintenance and running) for a federal pipeline to be offered as an observing mode to large programmes or open skies.
- Aim is to progressively centralise pipelines that may start off as federally supported and resources must be found to aid this transition.
- WG4 group **did not** consider who is responsible for the infrastructure that federally supported and centrally supported pipelines will run on.

Thanks – questions?