Lessons learned from other missions

Gaia ground segment – practical matters

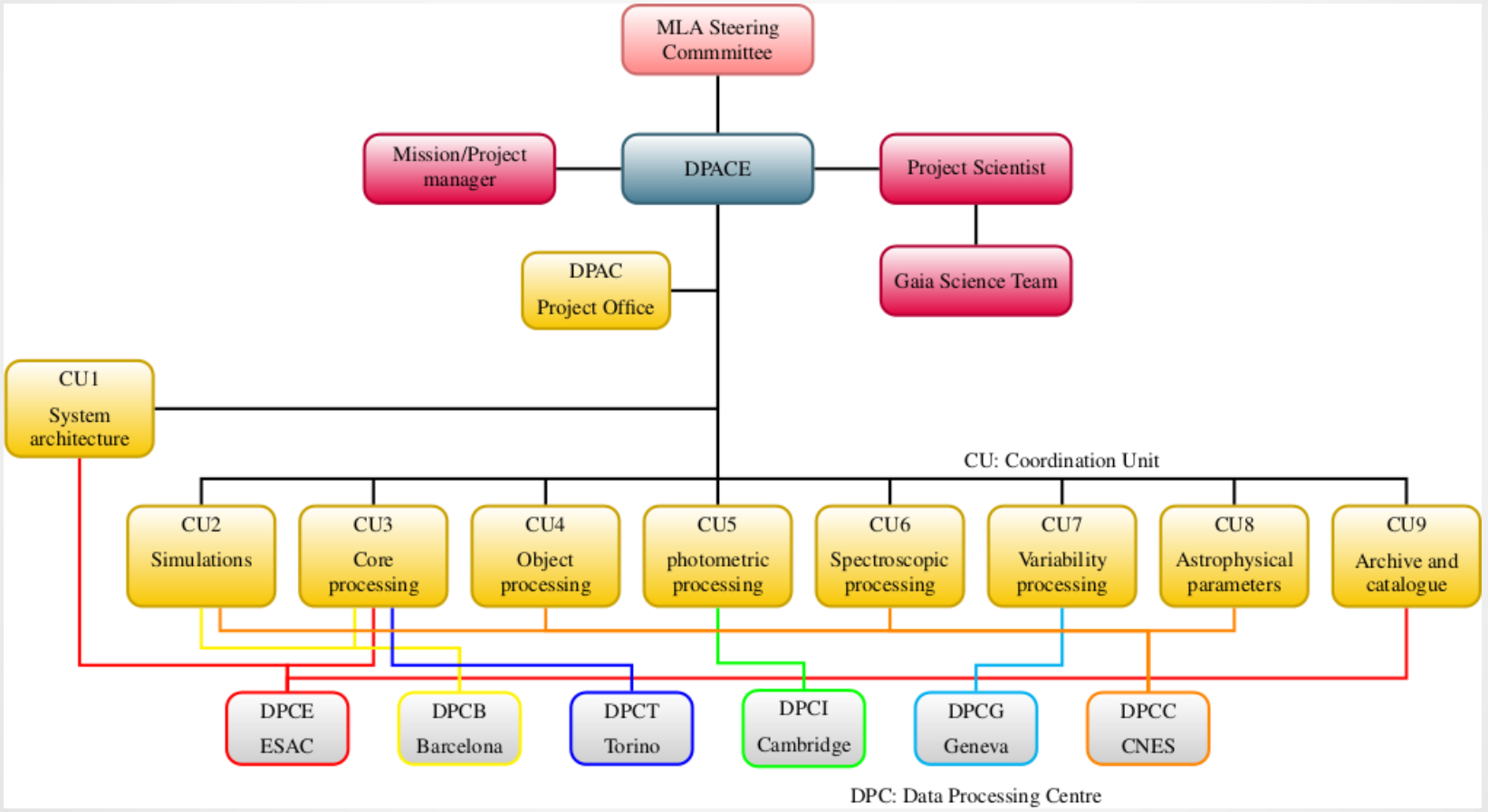
Orlagh Creevey (CU8-Lead) (Gaia-CU8 WP lead since 2014)

orlagh.creevey@oca.eu

Outline

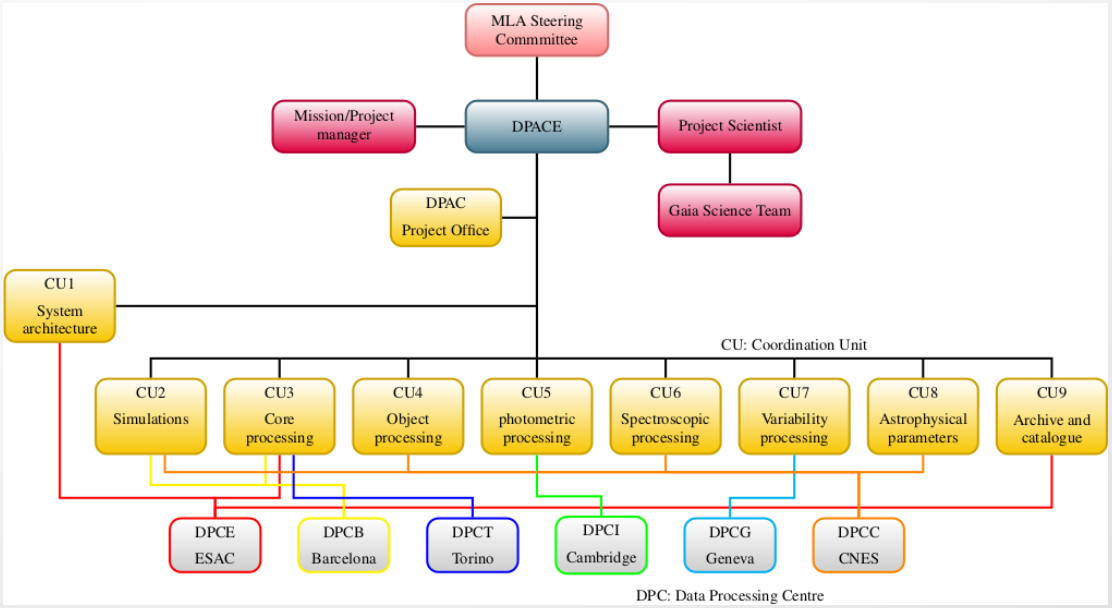
- Gaia DPAC structure
- Coordination Unit Lead (Level 4 data)
- (few words on validation added at the end)

Gaia DPAC Structure



Gaia DPAC Management

- 6 main Coordination units (CUs)
- 1 System architecture / 1 archive / 1 simulations (<fte)
- Data processing centers (DPCs)
- ESA : Project scientist / Project Office / Mission manager
- (MLA Steering committee / Gaia Science Team)
- « DPACE »



Communication Tools within DPAC

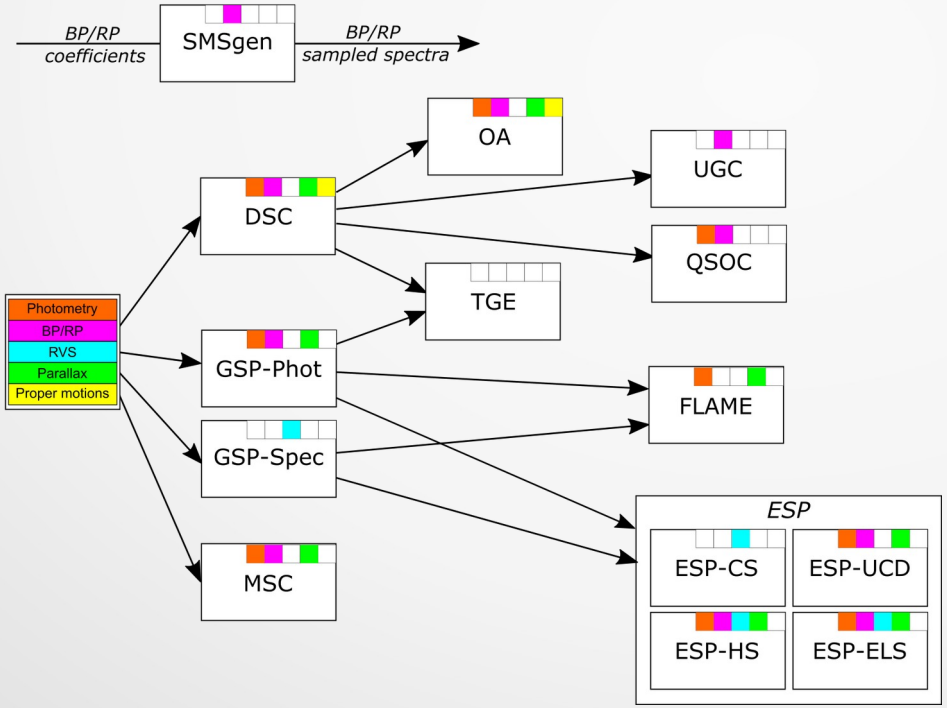
- General tools :
  - JIRA project for « issues » (actions, operations, discussions, decisions)
  - Wiki
  - Technical Documents (« livelink » server)
  - Mailing lists / email
  - Telecons (different levels) webex, zoom, teams (not so popular)
  - In-person/hybrid « plenaries »
  - On-line/in-person meetings on focussed items
- Specific :
  - CNES also use confluence and teams/skype(?) for chat but rest of DPAC do not have access
  - CU8 have used slack for preparation of data release, but we prefer to use that only temporarily (avoid being constantly interrupted)
  - Overleaf for writing
  - Sometimes google (easiness of use)

DPACE communication

- 1 representative from all CUs DPCs, PO, Mission Manager, PScientist
- Monthly telecons (2.5 hrs with 10 min break)
  - Date decided during previous telecon (4 weeks is a good time)
  - Agenda provided always 1 week before
  - Clear indication of topic, speaker, and time allocated
  - Presence from all CU/DPC expected
  - Minutes in the form of a « TN » usually within ~10 days (after feedback)
- 2 meetings of1.5 days per year (now all provide hybrid possibility)
- Discuss many issues and make decisions (usually within a timeframe after the meeting) : schedule, particular issues from CUs or DPCs, status reports, new propositions, archive, delivery dates,...
- Action items/Decisions to be taken → bring to CU level

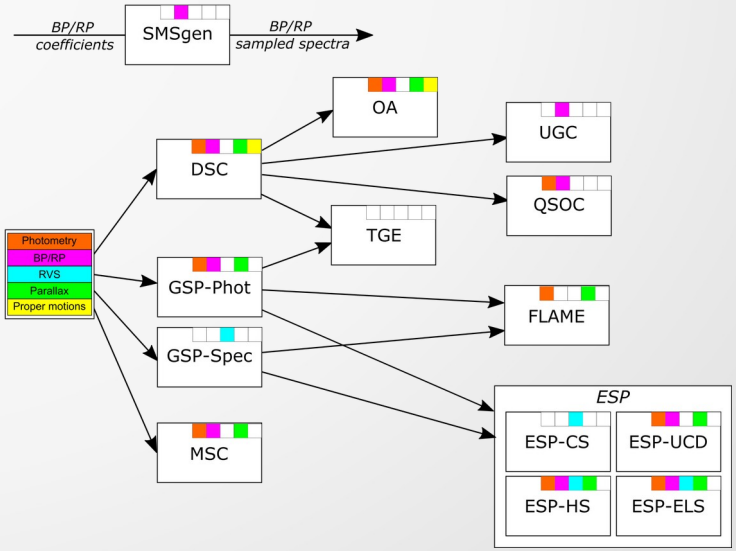
CU-8 Structure (Astrophysical parameters)

- ~75 members (of which ~15 DPCE/C),
  - Total of ~20 FTE
  - around 32 > 0.3 FTE, many 0.1-0.2
- 14 Processing Workpackages (~12.7 FTE)



CU-8 Structure (Astrophysical parameters)

- ~75 members (of which ~15 DPCE/C),
  - Total of ~20 FTE
  - around 32 > 0.3 FTE, many 0.1-0.2
- 14 Processing Workpackages (~12.7 FTE)
- Process upstream data to derive APs :
  - Classification
  - Atmospheric/Evolution params
  - Extra-galactic / 2D dust map
- Archive GDR3:
  - 10 tables (some MDD)
  - ~500 APs



CU-8 Structure (Astrophysical parameters)

- 14 Processing Workpackages (~12.7 FTE)

  +

- CU-Lead (me) + CU-Technical Lead (DPCC /CNES)
- Management team (0.7 FTE)
- Non-processing Workpackages (processing/validation support+) : simulations, validation, communication, outreach, documentation ~3.3 FTE (2.3 = sci. validation)
- DPCC (CNES) support (integration/operations/schedule) ~2.6 FTE
- DPCE (ESA) support (validation database) ~1FTE

CU-L & CU-T

- CU-Lead has the best overview and is the main articulator between the board, the other CUs and the team
  - The CU-L should have the team's interest and goals in mind and not their own agenda, and must be willing to dedicate time to (non-hierarchical) managing
- CU-Technical lead is the DPCC-dedicated person and also has the top-level view of CU + DPC schedule (+possibly other DPCs)
- Very helpful if they have a GOOD relationship and listen to/trust each other
- Monthly telecon with PO to discuss schedule/interfaces
- 1-2 hour weekly telecons is mandatory
  - Fix the hour each week, but are flexible to move if needed
  - Take the time to discuss all of the CU-level and the DPC-level issues along with the constraints from 'above'
  - Focus on (1) schedule (development/operational), (2) development and testing, (3) any other issues e.g. specific WP, (4) help to define the DPCC priorities for CU8
  - Detailed Validation/Operation schedule managed between us. But we also include non-DPCC issues (e.g. delivery of test data, validation from archive, simulations, …). One unique information figure with all information for WP manager

CU-8 Management Team

- Impossible to manage properly if one is alone
  - If management goes wrong, the team most likely will too
  - 7 heads is way better than one. Different expertise / ideas/ views
  - Many things to do (manage WP, manage interfaces, archive, documentation,….)
- Mng Team: CU-L + CU-T + 5 other members
- Each member is responsible for communication with another CU
- 1.5 hour -telecon every 5 weeks
  - Agenda fixed ahead of time
  - Report from different Cus/DPACE
  - Discussion with aim to make proposals to CU8 (to avoid a 70-person discussion going nowhere) e.g. how to present the data in the archive
  - Actions on team members to 'spread the load'
  - Minutes written within a week

CU-8 WPs

- JIRA/Wiki/Emails (mailing lists) in moderation/Slack/Overleaf (if writing)
  - Identify which method is best for what (and when)
- Regular telecons with WP managers (between 15-20 people)
  - weekly/bi-weekly/tri-weekly… depending on needs (e.g. during Operations)
  - Fixed a regular time/day of the week each year
  - (google) Shared calendar
  - Opportunity to ensure everyone is on the same page
- Focussed telecons with a subset of the managers (as often as needed, e.g. 1 per 2 months)
- Introduced a « CU8 Newsletter » ~ 1-2 months, gives other members a chance to know what is going on because they don't participate in the frequent meetings
- Scientific Validation half-day meetings
- ~2-3 CU8 plenaries per year
  - Before 2 in-person
  - Now 1 in-person + ~2-4 half-days online
  - Physical contact important because of long project
- Much Communication happens without me too !

Tips to make it work for the WP manager

- Clear delivery objectives and dates (schedule) with long-term view of things, and regular reminders
- Don't assume everyone remembers what to do. Make (short) documents available in a clear structured manner, or add a section to the wiki
- Regular meetings at different levels (one-on-one, group, …)
- Online tools : JIRA project management, svn for code delivery, templates for documentation, shared (google) calendar, online telecons for discussions, **shared documents / presentations / templates**
- Regular reminders and **updates** (but not too much)
- Common software tools
- Realistic timeline – Don't underestimate the time needed for any task in particular scientific validation

Tips for WP managers to make it work for the full team

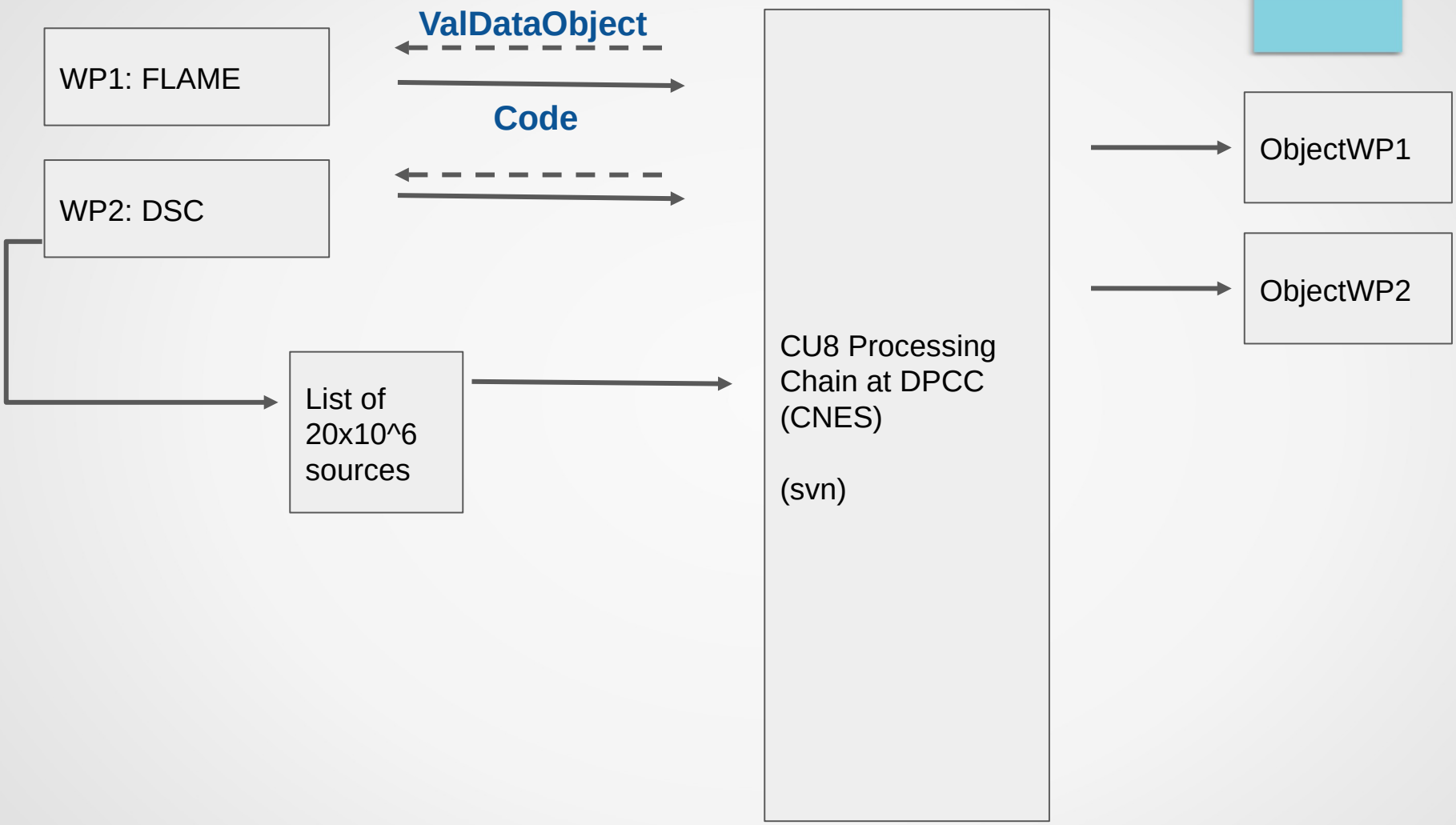
- Participate (and listen/ask) in regular telecons. If not available read the minutes and make sure you have no action items. Be aware of other things that are going on.
- Respect the deadlines that have been set, usually in coordination with the full team (given the top-level constraints)
- Test your codes before submitting to svn (unit tests, null pointers…) this saves signficant time at the DPCC level.
- Don't underestimate time needed for scientific and technical validation. Be prepared, use templates / python notebooks ...
- Go to main 'page' for information (wiki in Gaia) before sending emails to 3 people. Usually it is there, and if it is not, then it should be put there
- Report on ongoing activities

Overall advice for WP Manager

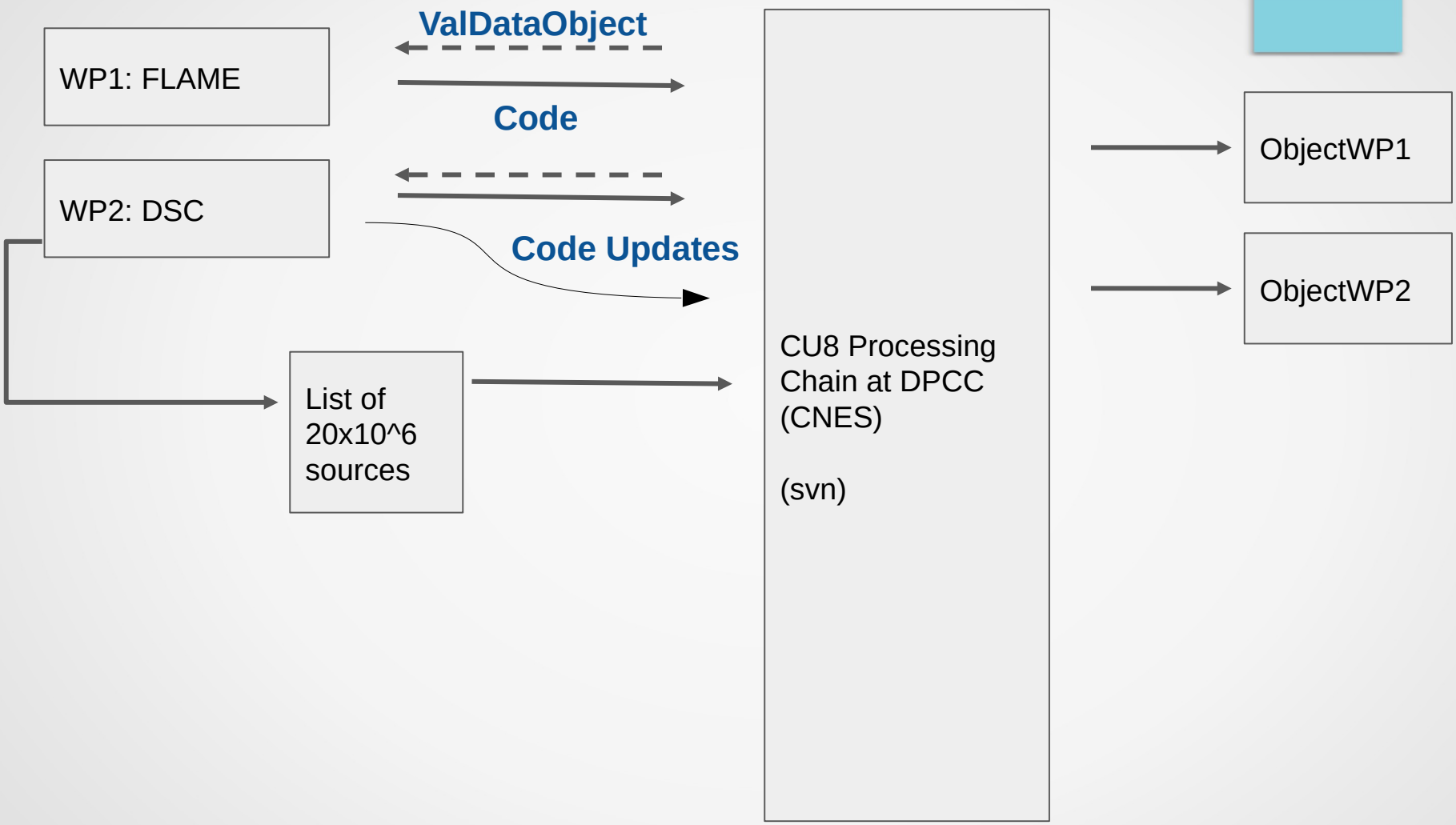
- Producing code means several things of equal importance:
  - Software production (incl Junit tests)
  - Documentation (could be with JIRA or Tns or A&A)
  - Technical & Scientific Validation
- Do not underestimate validation : prepare, practice, practice, …. (data, FTE, tools,...)
- All Data needs to be accessible (non-disclosure agreements)
- Scientific Coordinator needed
- Regular communication DPC - WPs is primordial (weekly/monthly)
- Flexibility DPC is very important
- Processing chains to be quite flexible -- allow modifications within reason (e.g. DataModel update every 6months, Code patches every month, Code integration every year)
- Need time to learn from new data

(a few words on validation)

Making your code work : Data Flow



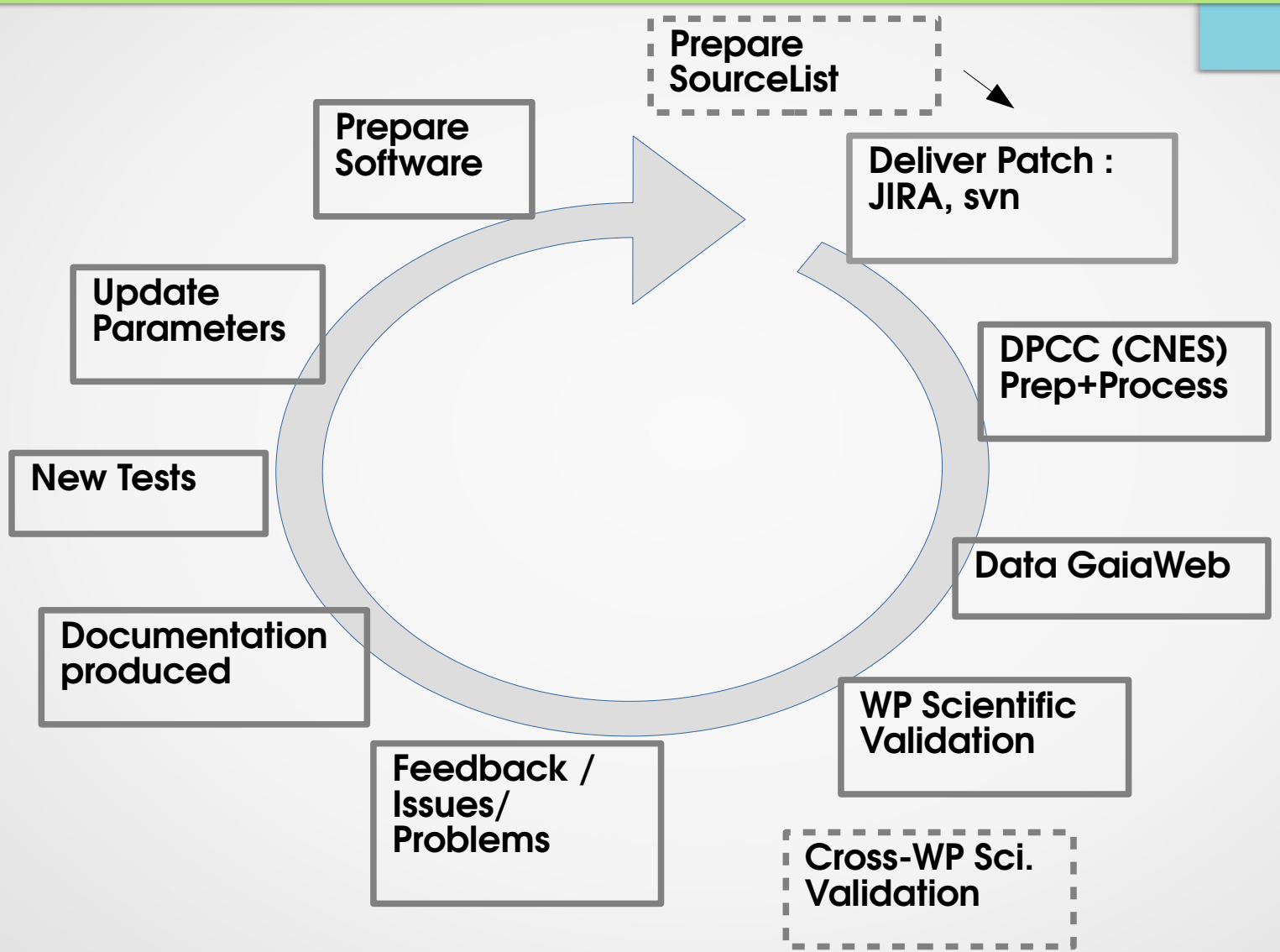
Making your code work : Data Flow



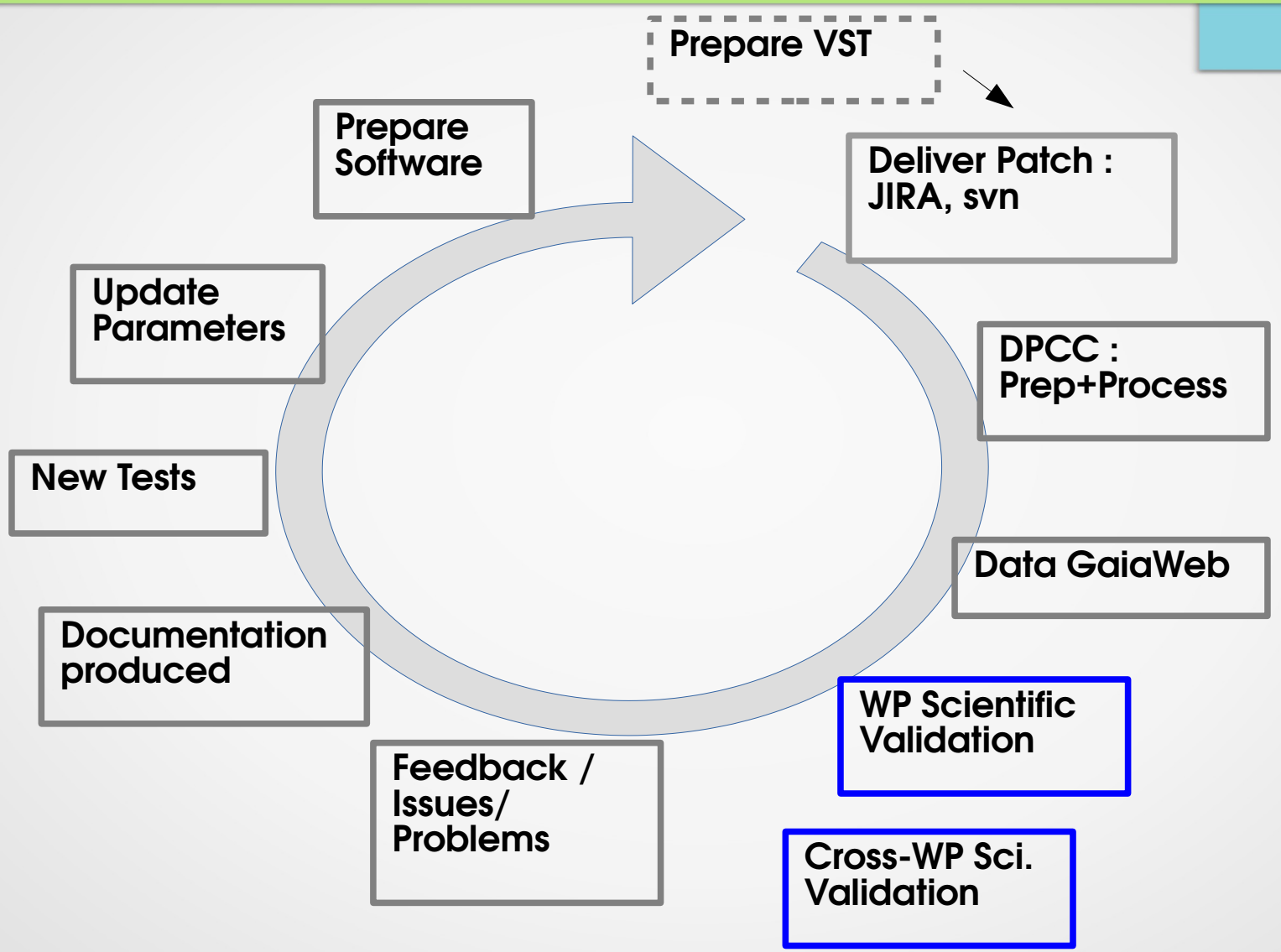
Making your code work : Data Flow

- Requires flexibility with Data Processing Center (not fixed codes)
- Requires access to lots of data to check
  - Does my code do what it should ?
  - Do I get similar results to literature for some sources ?
  - Do the global distributions and statistics make sense ?
  - Are the regions of the parameter space that don't work well ? Why ?
  - Do I get the same results as my offline (local) code ?
- Requires ability to run « validation » tests on large numbers of sources regularly (not too much, not too little)
- Requires ease to interface with scientific team and the DPC team
- Requires flexibility (not too much) on data model structure e.g. every 3 months we make global updates to our Main Data Models

Development once the code is running...



Development once the code is running...



Take away points :what have we learned

- Clarity in communication (everyone should know what to do if… or know how to find that information...)
- Regular communication DPCC-WPs is primordial
- Flexibility DPCC / Code is a must
- Stability of personnel is important
- Practice and practice and test and test (validation, operational processing….)
- Respect of deadlines
- Data needs to be available for quick turn-around time
- More precise data means more « systematic issues » to understand … need time to learn from new data
- Scientific (financial) support is always needed