

Building the Gaia ground segment revisited

William O'Mullane

Gaia Science Operations Centre
European Space Astronomy Centre
European Space Agency

January 10th 2017, LSST Tucson

what did and did not work ..

Or

the good

the bad

and

the ugly.

The western is quintessentially American, that one was Italian though and it was shot in Spain.

1 Initial Set up

2 Standards and Tools

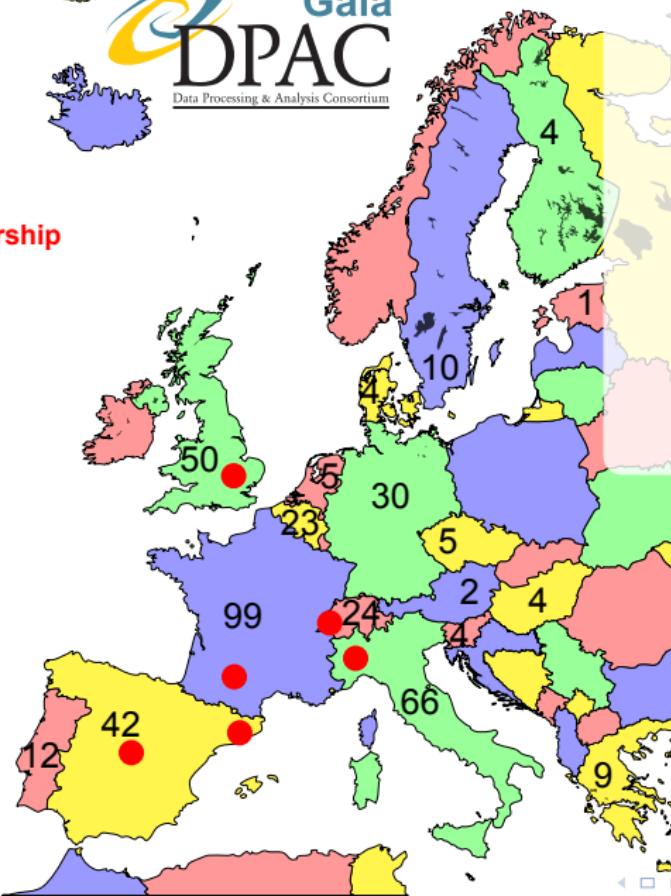
3 Management of Gaia

4 Conclusion



DPAC membership
January 2013
432 total

BR: 5
CA: 1
CL: 1
ESA: 28
IL: 1
US: 2



- Gaia data processing is a Pan-European cooperation
 - ▶ Over 1000 staff years effort since 2006
 - ▶ Processing power spread over 6 centres
 - ▶ Supported through national funding
 - ▶ Additional support from EC Marie Curie and ESF



In the beginning



- initial ideas for ground segment were in place for the study in 2000
ESA-SCI(2000)4
- already clear then we would have distributed processing in multiple centres (though not mentioned)
- Intention was to have autonomy between coordination units
- Interviewed several project leaders for O'Mullane (WOM-003) in 2004/5 — tried to learn from them ...
 - ▶ Included LSST (Kantor) , SDSS (Borowski)
 - ▶ Management came out as the most difficult part of all projects ... so I will leave that until last.
- some things started before DPAC — simulations and GIS studies for example.
- Finally DPAC is large — I am sure you can find someone in DPAC to disagree with anything I say.

In the beginning



- initial ideas for ground segment were in place for the study in 2000
ESA-SCI(2000)4
- already clear then we would have distributed processing in multiple centres (though not mentioned)
- Intention was to have autonomy between coordination units
- Interviewed several project leaders for O'Mullane (WOM-003) in 2004/5 — tried to learn from them ...
 - ▶ Included LSST (Kantor) , SDSS (Borowski)
 - ▶ Management came out as the most difficult part of all projects ... so I will leave that until last.
- some things started before DPAC — simulations and GIS studies for example.
- Finally DPAC is large — I am sure you can find someone in DPAC to disagree with anything I say.

1 Initial Set up

2 Standards and Tools

3 Management of Gaia

4 Conclusion

Language (spoken) and conventions

- Very important in large groups
- Which way is up on Gaia, which is a row and column on focal plane, which quaternion to use
- dealt with quite early on in BAS-003
 - ▶ still Astrium have a different definition for X , Y , Z on Gaia
 - ▶ at least in the consortium there is only one — could have been much worse
- What is a product, a Work Package — why is 10 months = 1 year
- also dealt with early on in WOM-001
- Then there are Acronyms
<http://gaia.esac.esa.int/gpdb/glossary.txt> and an acronym tool for TeX files (e.g. Appendix 5)
- having a complete engineering guideline early was good JH-001

Language (spoken) and conventions

- Very important in large groups
- Which way is up on Gaia, which is a row and column on focal plane, which quaternion to use
- dealt with quite early on in BAS-003
 - ▶ still Astrium have a different definition for X , Y , Z on Gaia
 - ▶ at least in the consortium there is only one — could have been much worse
- What is a product, a Work Package — why is 10 months = 1 year
- also dealt with early on in WOM-001
- Then there are Acronyms
<http://gaia.esac.esa.int/gpdb/glossary.txt> and an acronym tool for TeX files (e.g. Appendix 5)
- having a complete engineering guideline early was good JH-001

Language (spoken) and conventions

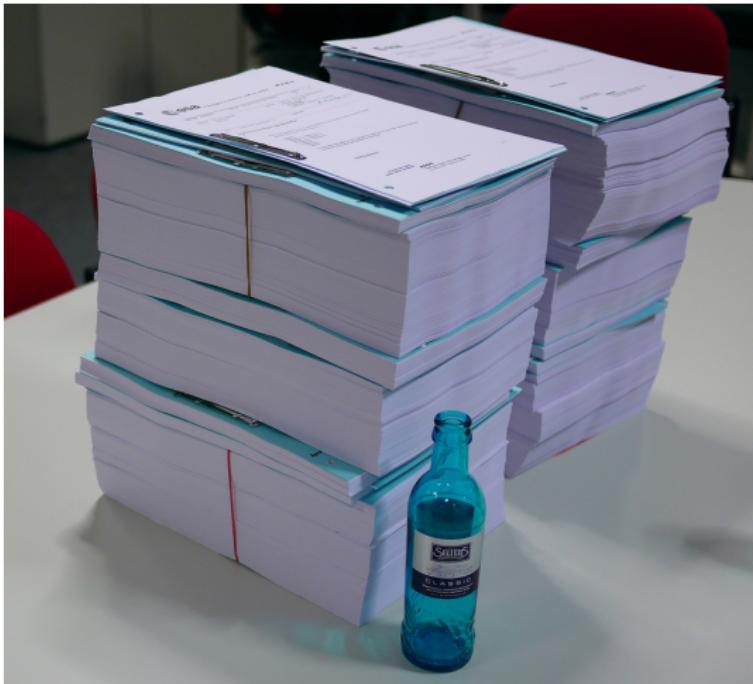
- Very important in large groups
- Which way is up on Gaia, which is a row and column on focal plane, which quaternion to use
- dealt with quite early on in BAS-003
 - ▶ still Astrium have a different definition for X , Y , Z on Gaia
 - ▶ at least in the consortium there is only one — could have been much worse
- What is a product, a Work Package — why is 10 months = 1 year
- also dealt with early on in WOM-001
- Then there are Acronyms
 - <http://gaia.esac.esa.int/gpdb/glossary.txt> and an acronym tool for TeX files (e.g. Appendix 5)
- having a complete engineering guideline early was good JH-001

Language (spoken) and conventions

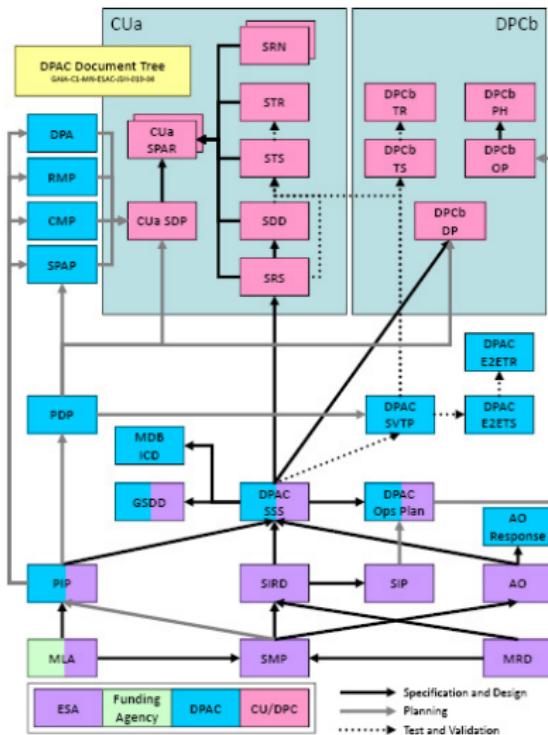
- Very important in large groups
- Which way is up on Gaia, which is a row and column on focal plane, which quaternion to use
- dealt with quite early on in BAS-003
 - ▶ still Astrium have a different definition for X , Y , Z on Gaia
 - ▶ at least in the consortium there is only one — could have been much worse
- What is a product, a Work Package — why is 10 months = 1 year
- also dealt with early on in WOM-001
- Then there are Acronyms
<http://gaia.esac.esa.int/gpdb/glossary.txt> and an acronym tool for TeX files (e.g. Appendix 5)
- having a complete engineering guideline early was good JH-001

- Avoid different values of constants in peoples code ...
- The Gaia Parameter Database was set up early on for this (de Bruijne et al., 2005)
 - ▶ all constants in one place; web searchable configuration controlled (Only updated by Jos De Bruijne)
 - ▶ published as constants for Java (can also do C, Fortran...) so you may refer to a particular version
- then the actual data model — what exactly is an AstroElementary?
 - ▶ entire data model defined in multi-user dictionary tool; includes Units on each field.
 - ★ good for astronomers — computer people find it harder to handle
 - ▶ from it we generate data instance classes and schemas for storage.
 - ▶ **ONLY data model not processing** — all objects are dumb (had discussion with KT and Mario on this)
- These are logical extensions of having agreed conventions... .

Flight Operations Procedures in MOC



The FOP is followed by the spacecraft operators - the paper copy is just in case the computers fail - could be useful! **But we should avoid *write only* documents.**



European Cooperation for Space Standardization

- ECSS tailored as in figure
 - ▶ LaTeX Templates/examples provided (by SOC)
 - ▶ Documents are iterated — All of this is done for all DPAC products.
 - ▶ It is very good to have a standard set of documents augmented by technical notes and streamlined ECSS
 - ▶ Some still found it too heavy — other reports requested beyond the standard ones.
- DPAC had sufficient QA people (~ 1/CU) from the start to help with this

- No standard - no Overview of DM docs I could find some web pages
 - ▶ LDM-493 .. covers some aspects .. but still no Doctree - templates ?
 - ▶ not nice to have tex/markdown/word others ? Markdown ok for implementation docs but for technotes for science notes?
 - ▶ did not find a document management plan for DM ..
- Mention made of LPM-051 .. but is DM following this ? Or just not declaring controlled documents .. prefer term ISSUED rather than controlled
 - ▶ docushare is the LSST repo but DM docs seem old (did find some from KT) - should all issued docs be there (I think so ..)
 - ▶ would expect to find all Technotes from DM in docushare .. like in a technotes folder ... SQR-006 gives an API which I may be able to use - should not be this hard.
- would like to find all draft docs in one DM repo ..
- everything is *referenceable* - but I do not find a single bibtex database or such for issued docs .. (Gaia livelink metadata is exported as a bibtex every night)

- LDTN-030 - ok for code documentation - understand its evolving-
 - ▶ overlaps LDM-493 somewhat - Why so many documentation projects - should we have ONE for DM ?
 - ▶ Provides document shards - no structure - how do you get a an old version (users will not get from git) .. LTD perhaps - but not well explained.
 - ▶ All products should be listed and for each product the set of docs .. the Audience seems too broad to cover with one type of document.
 - ▶ This "In our case, the situation is complicated by the fact that Data Management software is being built by multiple teams across many coupled repositories. Taken together, the software repositories of the Data Management System are typically called the Stack, but not all parts of the Stack are used together. There are server-side database and display components, as well as pipelines algorithms bound together with middle ware." **does not sound like good system engineering is in place ..**
 - ▶ Unclear in a massive web how one can check consistency to any given level for docs .. **Potential problem for reviewers**

Single Sign on Gaia Portal

- <http://www.rssd.esa.int/index.php?project=MYPORTEL&page=index> hosted at ESTEC; set up eons ago...
- Names, emails and affiliations (phone numbers, photo, address) of all Gaia people
- Single login (LDAP) for
 - ▶ Livelink — for all published documents
 - ▶ Wiki — for wiki things (meeting setup etc.); always draft nearly always out of date
 - ▶ Mantis — for all issues
- Same LDAP for SVN, MDB dictionary etc
- Single sign on is perhaps not great but having one LDAP for authentication of everything is fabulous!
- Having information in SVN, Livelink and possibly on a wiki is not great but we do not have a solution
- Having single agreed set of collaboration tools from the outset excellent.

Development tools

- All DPAC code and docs in Subversion, hosted at ESAC
 - ▶ Access control according to Group membership in the LDAP
- Mantis for centralised issue tracking (includes risks and actions)
 - ▶ ALL DPAC issues in one system hosted at ESTEC
 - ▶ Jira would probably be better
- Having one language is good (O'Mullane et al., 2011) agreed 2006 (JH-001) — only one verification part is NOT in Java.
 - ▶ Can have a library of standard routines GaiaTools (Relativity, Field Angle Calculator, Ephemeris handling...)
 - ★ The set of routines were not defined hence GaiaTools is a bit of hodgepodge mess...
 - ★ Counter argument for common tools is (unnecessary) interdependence...
 - ▶ all libs in Nexus
 - ▶ builds with Ant; Maven might be the thing to use now
 - ▶ automated builds with Hudson/Jenkins (previously cruisecontrol)
 - ▶ virtual machines make some reasons for Java invalid

Development tools



- All DPAC code and docs in Subversion, hosted at ESAC
 - ▶ Access control according to Group membership in the LDAP
- Mantis for centralised issue tracking (includes risks and actions)
 - ▶ ALL DPAC issues in one system hosted at ESTEC
 - ▶ Jira would probably be better
- Having one language is good (O'Mullane et al., 2011) agreed 2006 (JH-001) — only one verification part is NOT in Java.
 - ▶ Can have a library of standard routines GaiaTools (Relativity, Field Angle Calculator, Ephemeris handling...)
 - ★ The set of routines were not defined hence GaiaTools is a bit of hodgepodge mess...
 - ★ Counter argument for common tools is (unnecessary) interdependence...
 - ▶ all libs in Nexus
 - ▶ builds with Ant; Maven might be the thing to use now
 - ▶ automated builds with Hudson/Jenkins (previously cruisecontrol)
 - ▶ virtual machines make some reasons for Java invalid

Outline



- 1 Initial Set up
- 2 Standards and Tools
- 3 Management of Gaia
- 4 Conclusion

- Internal communication:
 - ▶ Some say could have been better.
 - ▶ Few focused working groups and working meetings.
 - ▶ As for any project cost of entry for new people is very high — no obvious solution.
- ESA policy initially to reduce contact between DPAC and Astrium (who construct Gaia) **not good**.
- External communication:
 - ▶ Perhaps could have had a better DPAC website.
 - ▶ ESA PR also not great (ok as they point out they have a tiny fraction of NASA budget).
 - ▶ Publication policy was dealt with very late.

Requirements Management

- Some requirements at an appropriate level are very useful.
 - ▶ Many could have been better formulated.
 - ▶ Separation of performance, software, and science requirements should have been clearer.
- All requirements and test reports then ingested in the Information Management Tool (Comoretto et al., 2012):
 - ▶ Automated testing provided much requirements verification.
 - ▶ Operations Rehearsals used to validate many requirements.
 - ▶ Could probably have put more effort in this earlier.
- Cumbersome ESA Reviews are an unavoidable part of all this:
 - ▶ Just do it.
 - ▶ Had to convince all collaborators to also support reviews.

Management and science I



All large projects, and especially science projects, have management issues.

- In 2006 we had a big management training week for the DPAC management — though sceptical to start most found it good:
 - ▶ Despite this management support was still lacking.
- Science project management is a little different, still books like (Handy, 1993) are quite useful.
- Cyclic (Agile type) development seems well suited to science:
 - ▶ We chose six month cycles — probably too long.
 - ▶ Some prototypes started very early (O'Mullane & Lindegren, 1999).
 - ▶ We have great simulations — they started in 1998 before Gaia was accepted:
 - ★ Still they always seemed to be a little behind what people wanted — we have no solution for that, could not start earlier.
 - ★ Simulator fell out of the ECSS rigour — testing etc... .

Management and science II

- ▶ We did a lot of testing; some tests were probably not appropriate in hindsight.
- ▶ Despite aiming for test driven development — NOT ENOUGH effort in testing and many systems only recently got continuous integration.
- Scientific institutes are not good at managing things like software projects (hard anyway):
 - ▶ Interface control between software was insufficient — data model was not enough.
 - ▶ Perhaps ESA should have taken control of all critical software.
 - ▶ ESA is stepping back from this type of role in future missions and was not totally happy about the level of involvement of ESAC in DPAC.
- Industrial contracts for scientific software are very difficult:
 - ▶ Did an experiment with this on Gaia very early on.
 - ▶ XMM have their own woeful tale to tell.
 - ▶ Science consortia need managers and engineers already in the early phases

Management and science III

- ★ Noted lack of *engineering* in many areas — **lack of software engineers in initial phases, many hires were astronomers.**
- DPAC too collaborative?
 - ▶ DPAC was broadly inclusive — CU leaders on paper had the chance to include or not groups and WPs but in fact no one was left out.
 - ▶ This has lead to some inefficiencies — perhaps we could be smaller and more focused.
 - ▶ We possibly should have jettisoned some work packages, groups, and individuals early on.
 - ▶ In a proper phased approach some CUs could probably have started 2 years later with minimum presence at kick off
- Too flexible and too inflexible:
 - ▶ There are many configuration control boards and other groups to manage change.
 - ▶ No one wants any change to anything — *until the moment they want a change then it should be immediate!*
 - ▶ We have found no solution to this.

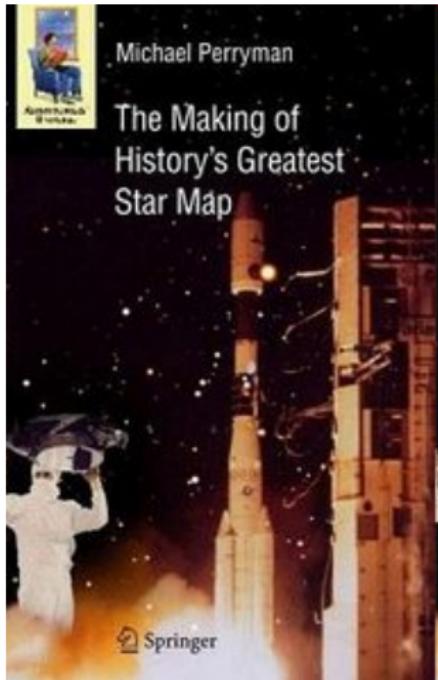
1 Initial Set up

2 Standards and Tools

3 Management of Gaia

4 Conclusion

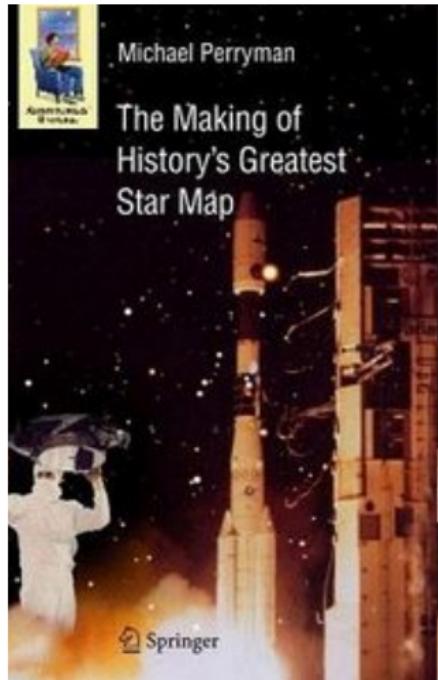
Conclusion



- many of us dreamed of a Gaia Processing Institute - only a dream for us.
- Much mentioned here is contentious — others do not see it exactly as I do.
- For me DPAC started off well largely due to strong initial leadership especially from Perryman
- All agree we are fortunate to have some excellent people in DPAC
- thanks to DPACE for their input to this presentation and to DPAC for all their work over the past years.

Perryman (2010)

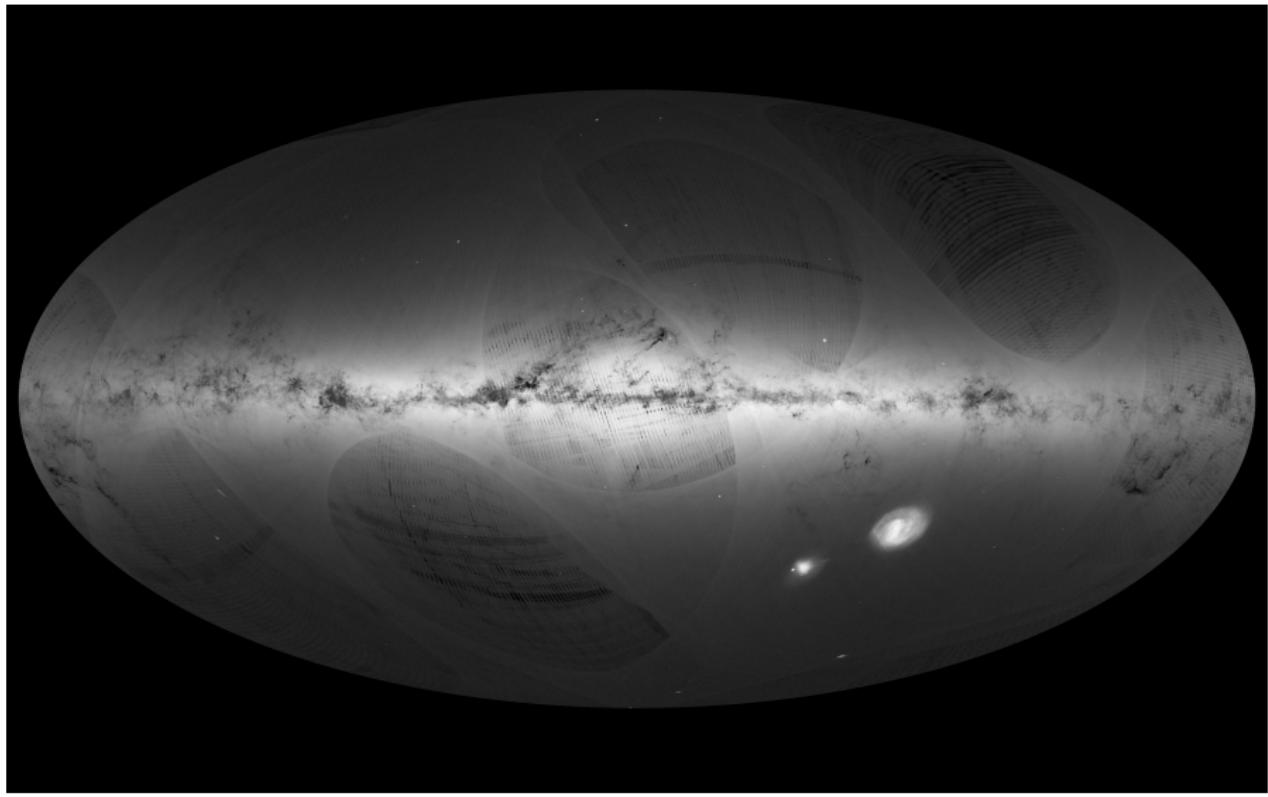
Conclusion



- many of us dreamed of a Gaia Processing Institute - only a dream for us.
- Much mentioned here is contentious — others do not see it exactly as I do.
- For me DPAC started off well largely due to strong initial leadership especially from Perryman
- All agree we are fortunate to have some excellent people in DPAC
- thanks to DPACE for their input to this presentation and to DPAC for all their work over the past years.

Perryman (2010)

It worked out pretty well !



Outline



5 Acronyms

6 References

Acronyms I

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
API	Application Programming Interface
CU	Coordination Unit (in DPAC)
DM	Data Management (LSST)
DPAC	Data Processing and Analysis Consortium
DPACE	Data Processing and Analysis Consortium Executive
ECSS	European Cooperation for Space Standardisation
ESA	European Space Agency
ESAC	European Space Astronomy Centre (ViSpa)
ESF	European Science Foundation
ESTEC	European Space research and TECnology Centre (ESA)
FOP	Flight Operation Procedure (Plan)
GIS	(Astrometric) Global Iterative Solution
IOA	Institute of Astronomy (Cambridge; also denoted IOA)
LDAP	Lightweight Directory Access Protocol
LSST	Large Synoptic Survey Telescope
LTD	LSST the Docs
LaTeX	(Leslie) Lamport TeX (document markup language and document preparation system)
MDB	Main Database
MOC	Mission Operations Centre
NASA	National Aeronautics and Space Administration (USA)
PR	Progress Report
QA	Quality Assurance

Acronyms II

SDSS	Sloan Digital Sky Survey
SOC	Science Operations Centre
SVN	SubVersioN
TOC	Table of Contents
USA	United States of America
WP	Work Package
XMM	X-ray Multi-mirror Mission (ESA; officially known as XMM-Newton)

Outline



5 Acronyms

6 References

[BAS-003], Bastian, U., 2007, *Reference systems, conventions and notations for Gaia*,
GAIA-CA-SP-ARI-BAS-003,
URL <http://www.rssd.esa.int/cs/livelink/open/358698>

Comoretto, G., Gallegos, J., Els, S., et al., 2012, In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8449 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ADS Link

de Bruijne, J.H.J., Lammers, U., Perryman, M.A.C., 2005, In: C. Turon, K. S. O'Flaherty, & M. A. C. Perryman (ed.) *The Three-Dimensional Universe with Gaia*, vol. 576 of ESA Special Publication, 67–+, ADS Link

[RD-010], Drimmel, R., Els, S., O'Mullane, W., et al., 2014, *DPAC Project Development Plan*,
GAIA-CD-PL-INAF-RD-010,
URL <http://www.rssd.esa.int/cs/livelink/open/2786669>

[ESA-SCI(2000)4], GAIA Science Advisory Group, 2000, *GAIA. Composition, Formation and Evolution of the Galaxy [The GAIA Study Report (ESA-SCI(2000)4)]*,
ESA-SCI(2000)4,
URL <http://www.rssd.esa.int/cs/livelink/open/359232>

Handy, C., 1993, *Understanding organizations*, Penguin Books, London, England New York, N.Y., USA

[JH-001], Hernandez, J., 2014, *Main Database Interface Control Document*,
GAIA-C1-SP-ESAC-JH-001,
URL <http://www.rssd.esa.int/cs/livelink/open/2786145>

[VI-001], Innocente, V., 2007, *External review report for Gaia SOC/DPAC SRR*,
VI-001,
URL <http://www.rssd.esa.int/cs/livelink/open/2803649>

[VI-002], Innocente, V., 2009, *External review report for Gaia SOC/DPAC DR*,
VI-002,
URL <http://www.rssd.esa.int/cs/livelink/open/2905956>

References II

- [WOM-003], O'Mullane, W., 2005, *Large scientific data systems - analysis of some existing projects and their applicability to Gaia*,
GAIA-C1-TN-ESAC-WOM-003,
URL <http://www.rssd.esa.int/cs/livelink/open/497678>
- [WOM-019], O'Mullane, W., 2007, *CU1 Progress Report #1*,
GAIA-C1-PR-ESAC-WOM-019,
URL <http://www.rssd.esa.int/cs/livelink/open/2803205>
- [WOM-001], O'Mullane, W., Lammers, U., 2007, *Work breakdown structures for DPAC*,
GAIA-C1-TN-ESAC-WOM-001,
URL <http://www.rssd.esa.int/cs/livelink/open/497865>
- O'Mullane, W., Lindegren, L., 1999, Baltic Astronomy, 8, 57, ADS Link
- O'Mullane, W., Luri, X., Parsons, P., et al., 2011, ArXiv e-prints, ADS Link
- Perryman, A., 2010, *The Making of History's Greatest Star Map*, Astronomers' universe, Springer,
URL <http://books.google.es/books?id=P-5pZ8GNuPIC>