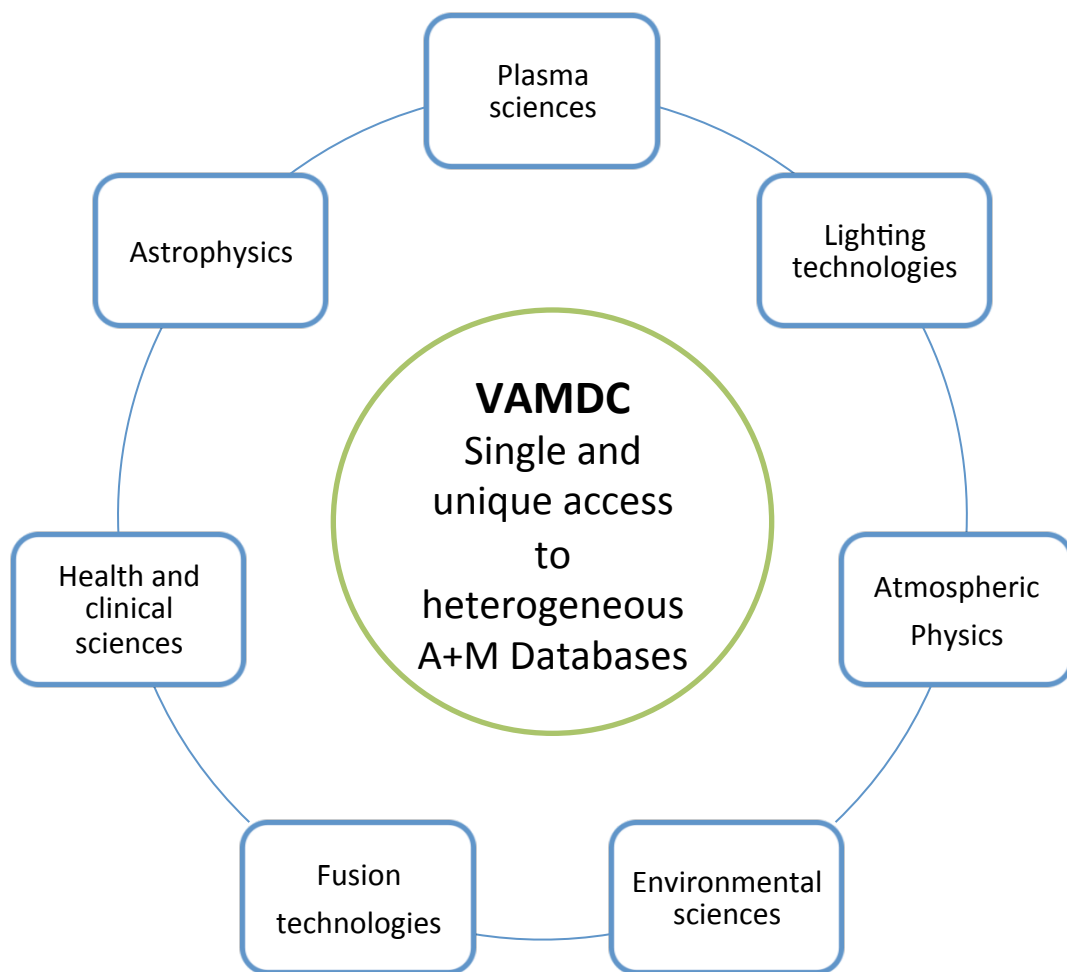


The Virtual Atomic and Molecular Data Centre

C.M. Zwölf, N. Moreau,
M.L. Dubernet & VAMDC consortium



The Virtual Atomic and Molecular Data Centre



➤ Federates ~30 heterogeneous databases
<http://portal.vamdc.org/>

➤ The “V” of VAMDC stands for Virtual in the sense that the e-infrastructure does not contain data. The infrastructure is a wrapping for exposing in a unified way a set of heterogeneous databases.

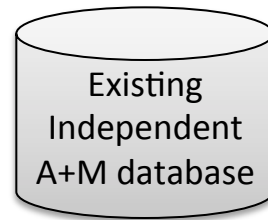
➤ The consortium is politically organized around a Memorandum of understanding (15 international members have signed the MoU, 1 November 2014)

➤ High quality scientific data come from different Physical/Chemical Communities

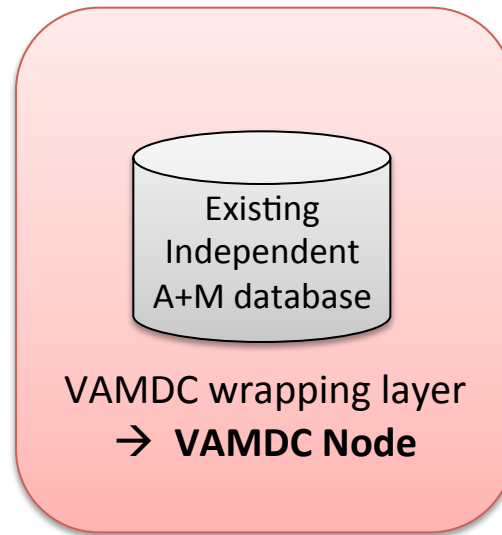
➤ Provides data producers with a large dissemination platform

➤ Remove bottleneck between data-producers and wide body of users

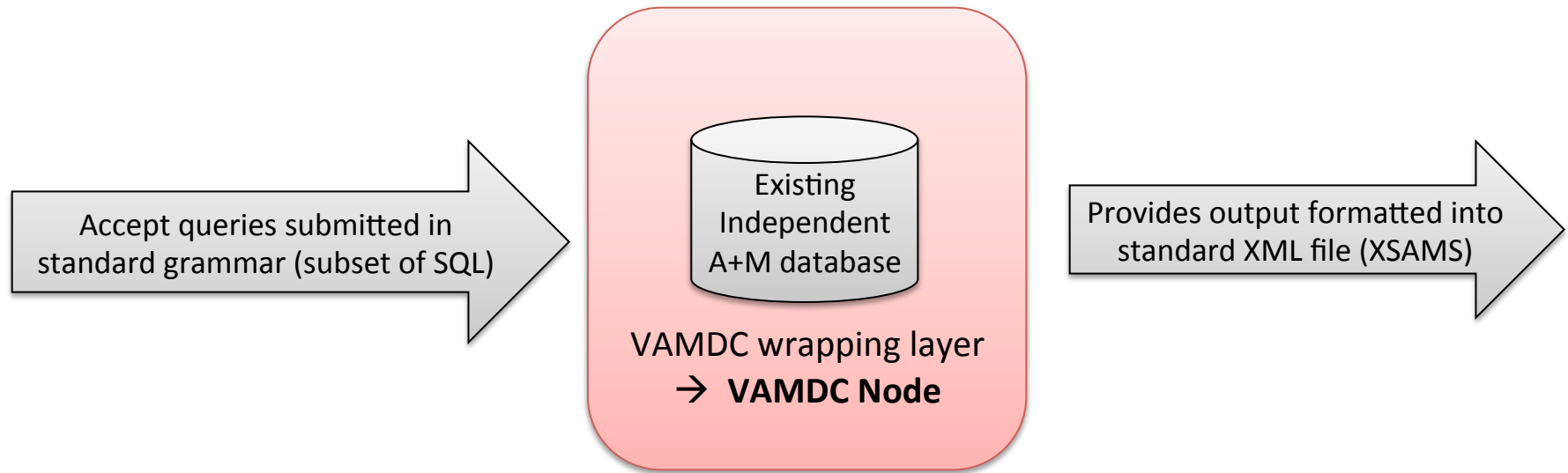
The VAMDC infrastructure technical architecture



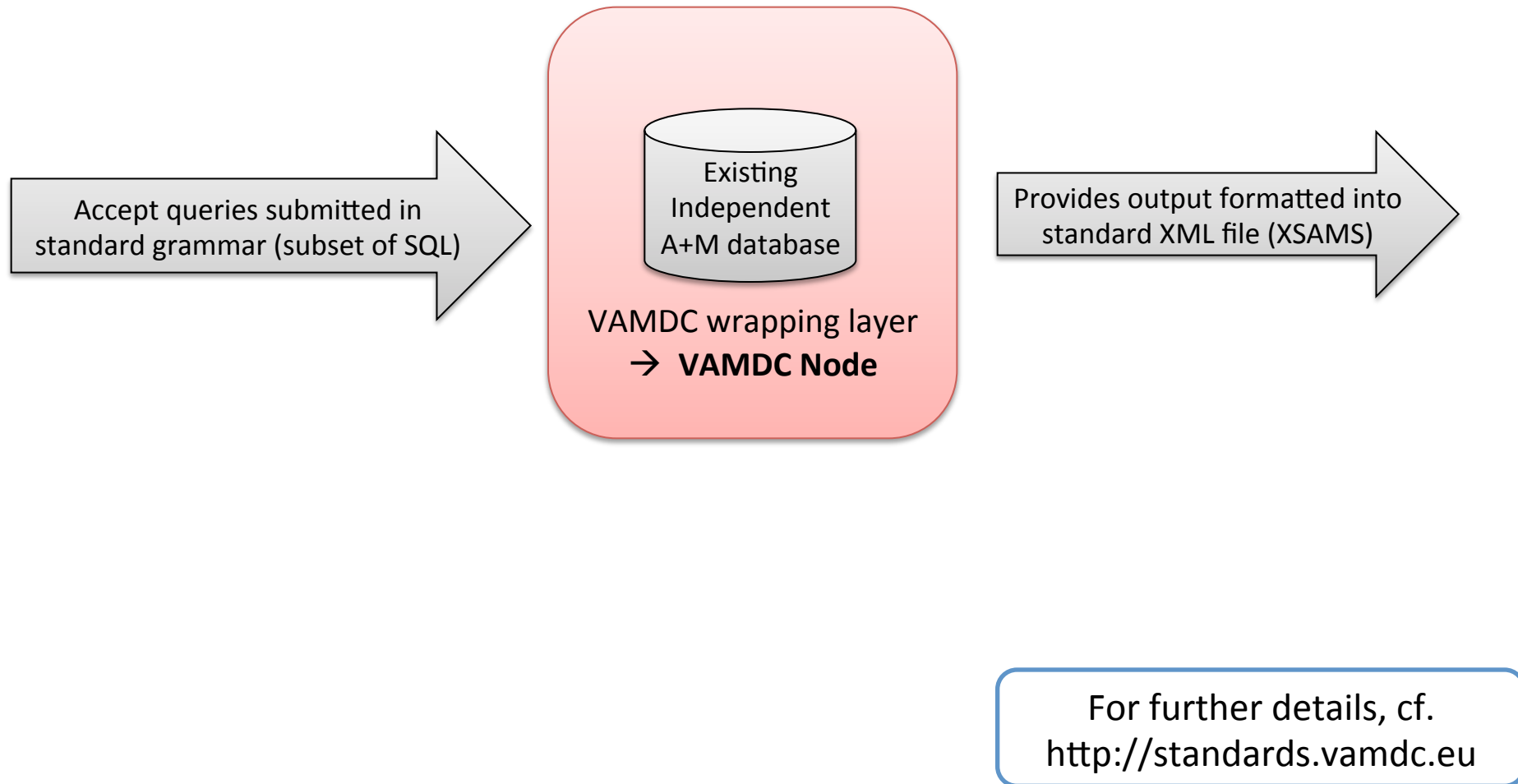
The VAMDC infrastructure technical architecture



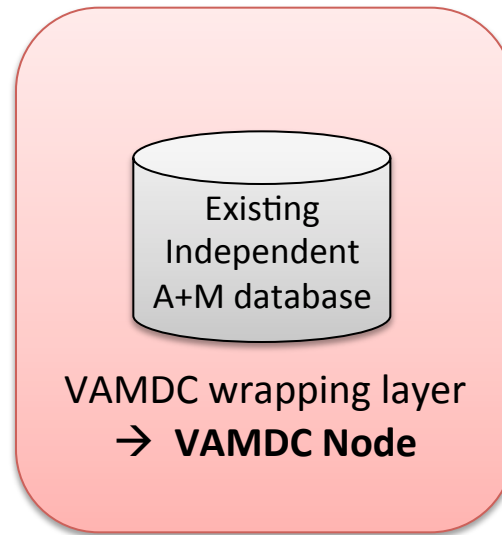
The VAMDC infrastructure technical architecture



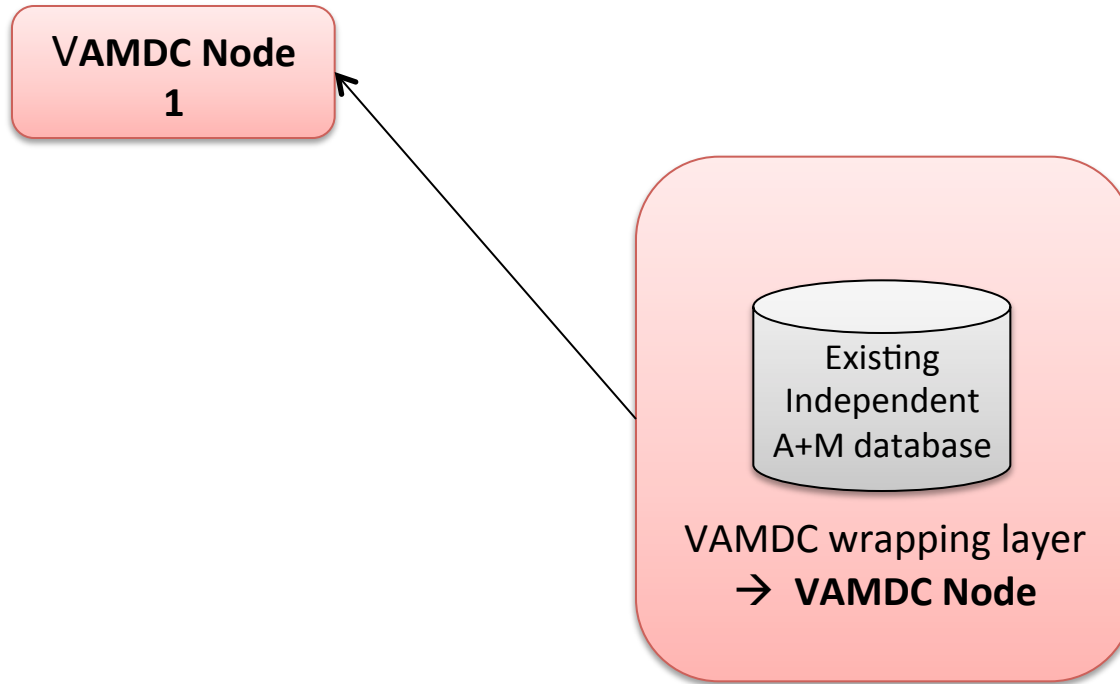
The VAMDC infrastructure technical architecture



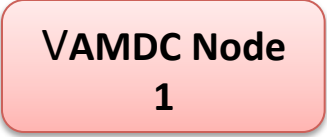
The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



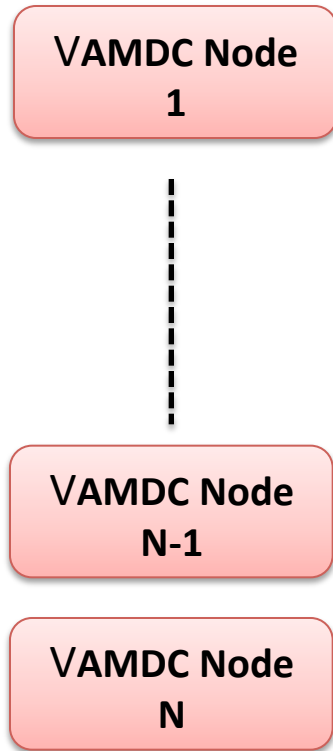
The VAMDC infrastructure technical architecture



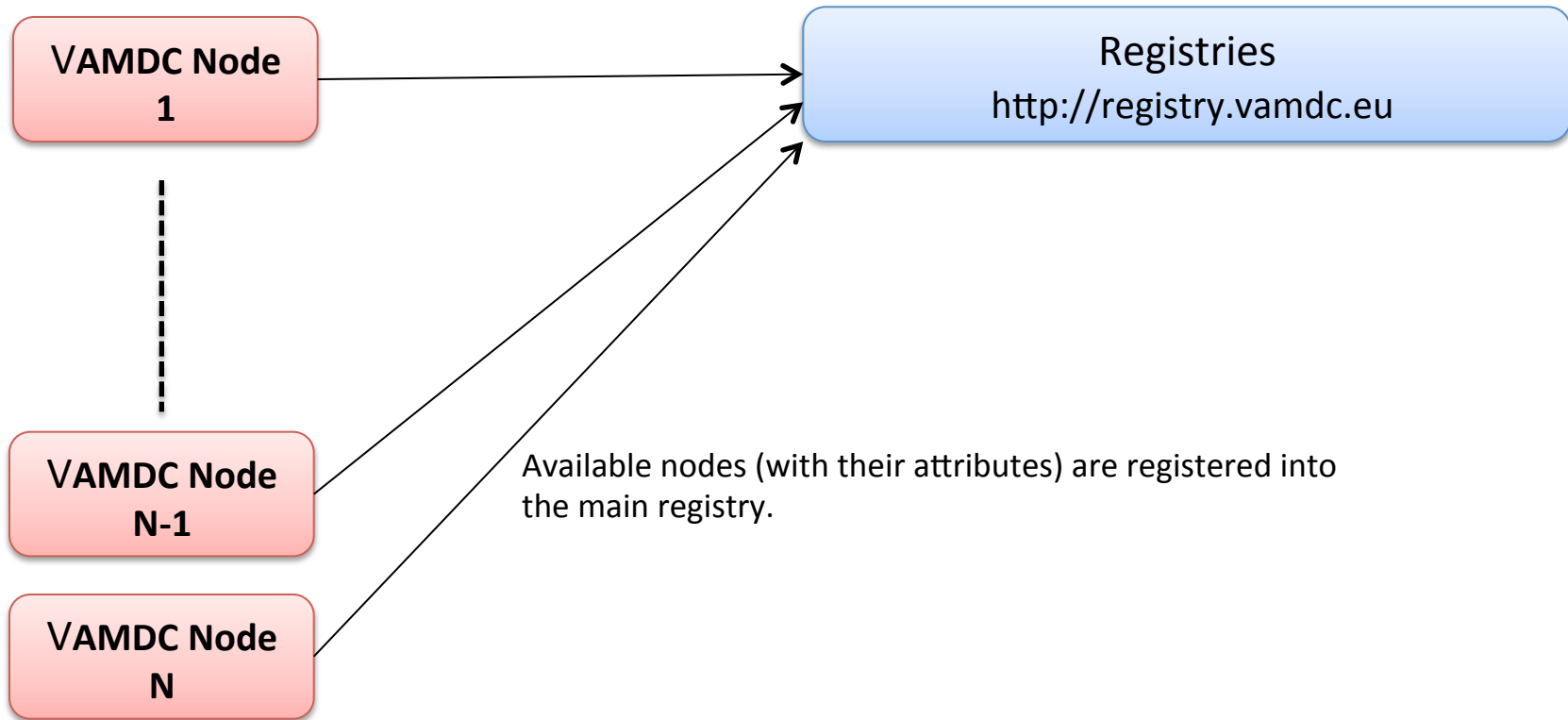
**VAMDC Node
1**

A red rounded rectangular box with a thin black border and a slight drop shadow, containing the text "VAMDC Node 1" in bold black font.

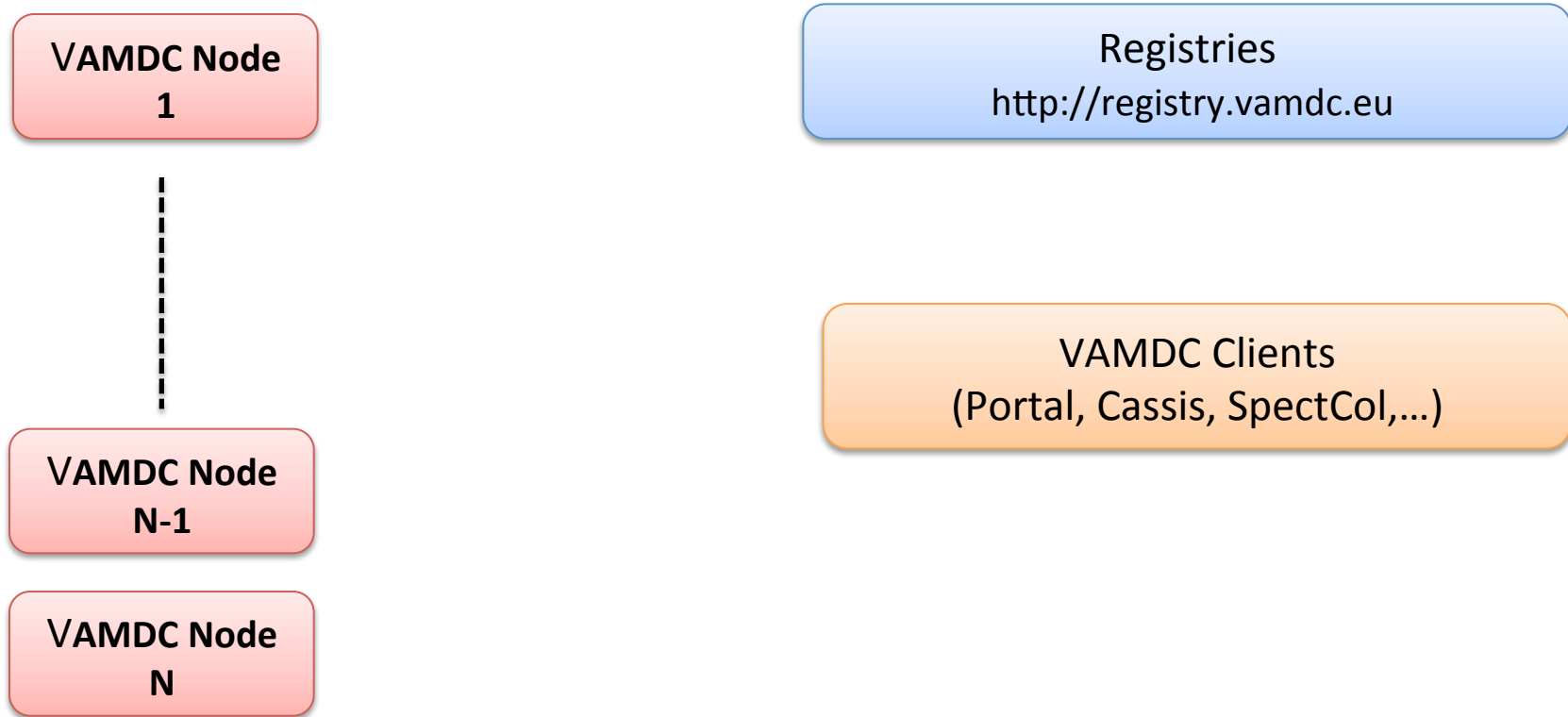
The VAMDC infrastructure technical architecture



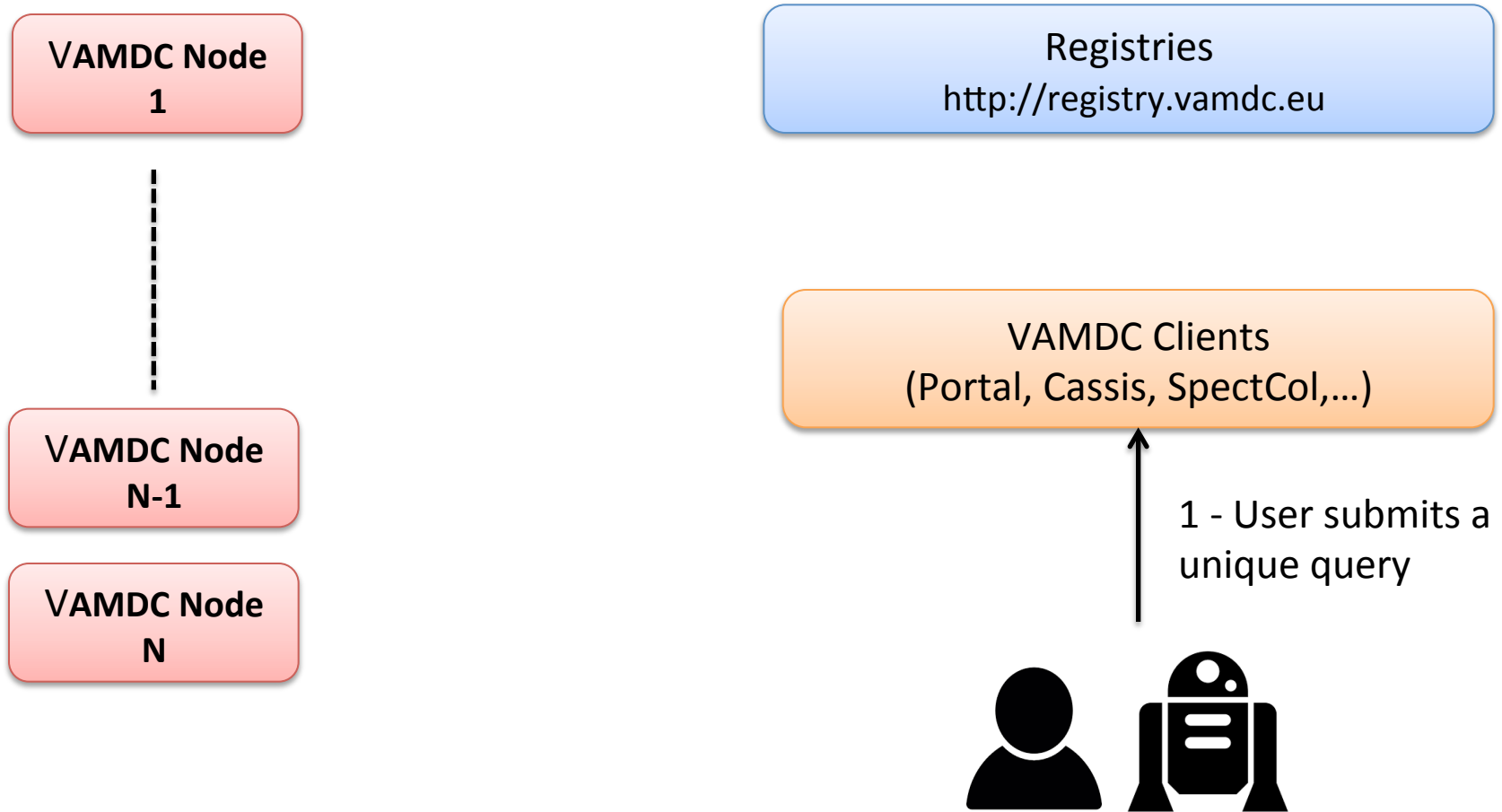
The VAMDC infrastructure technical architecture



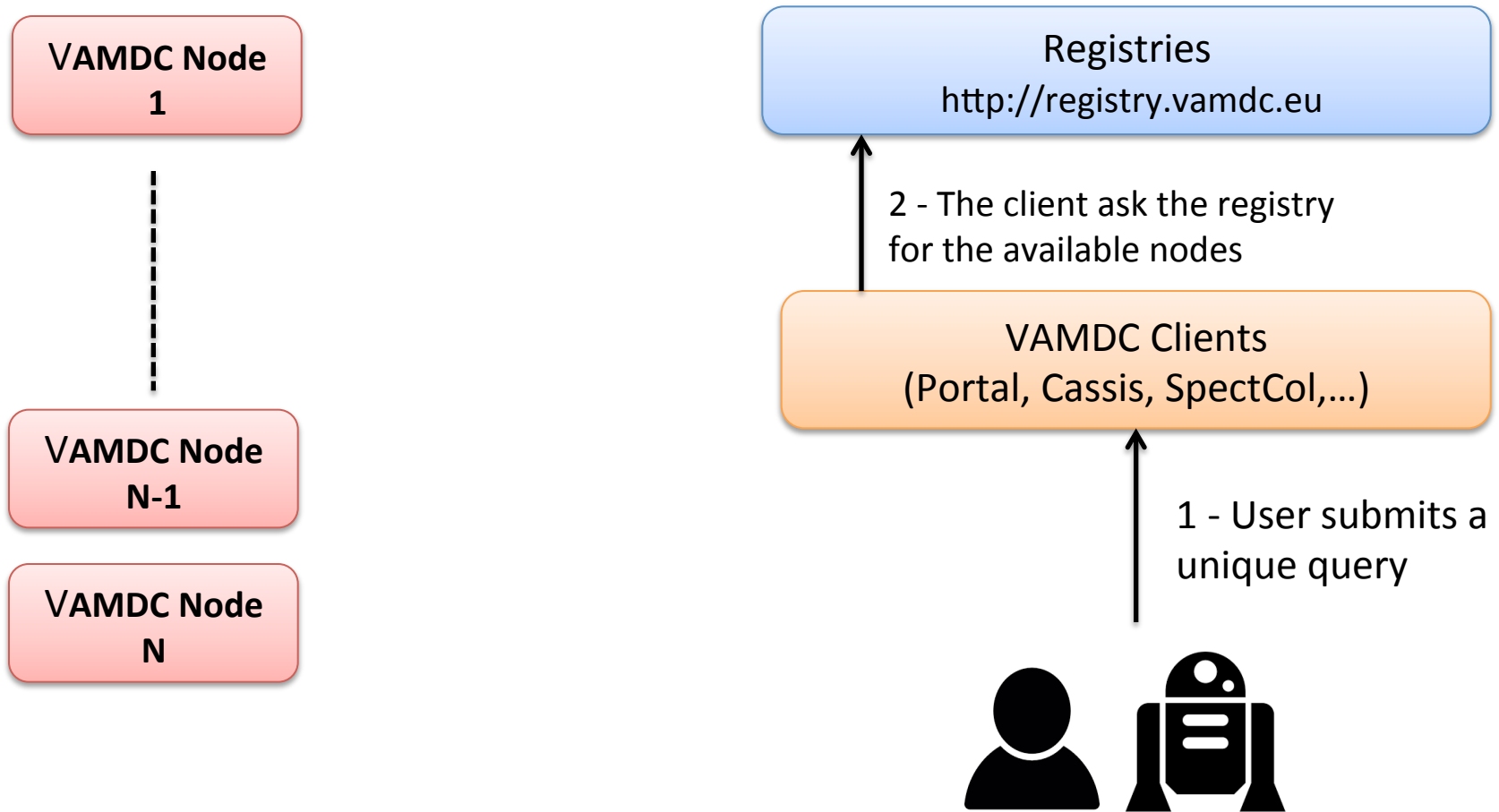
The VAMDC infrastructure technical architecture



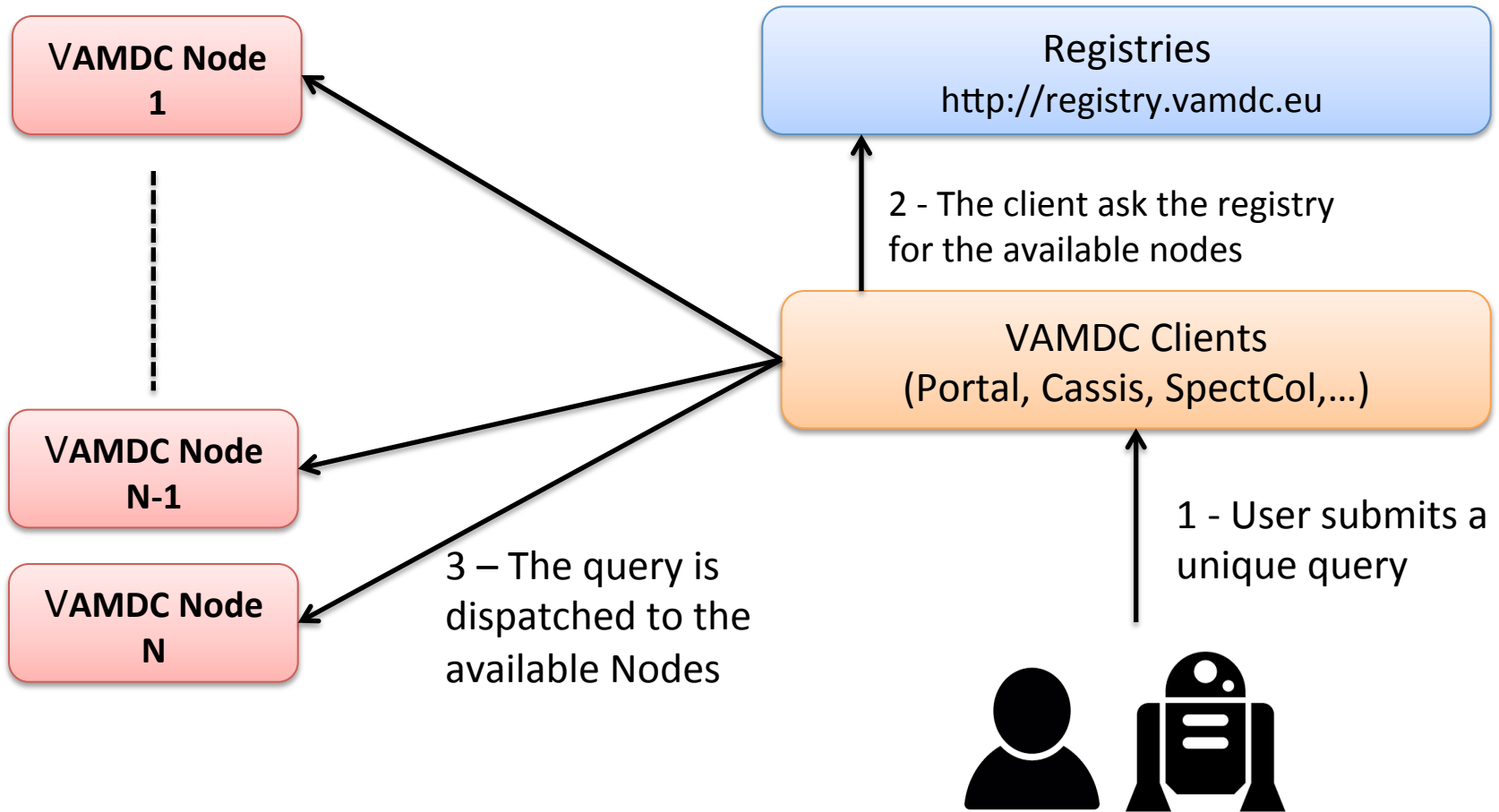
The VAMDC infrastructure technical architecture



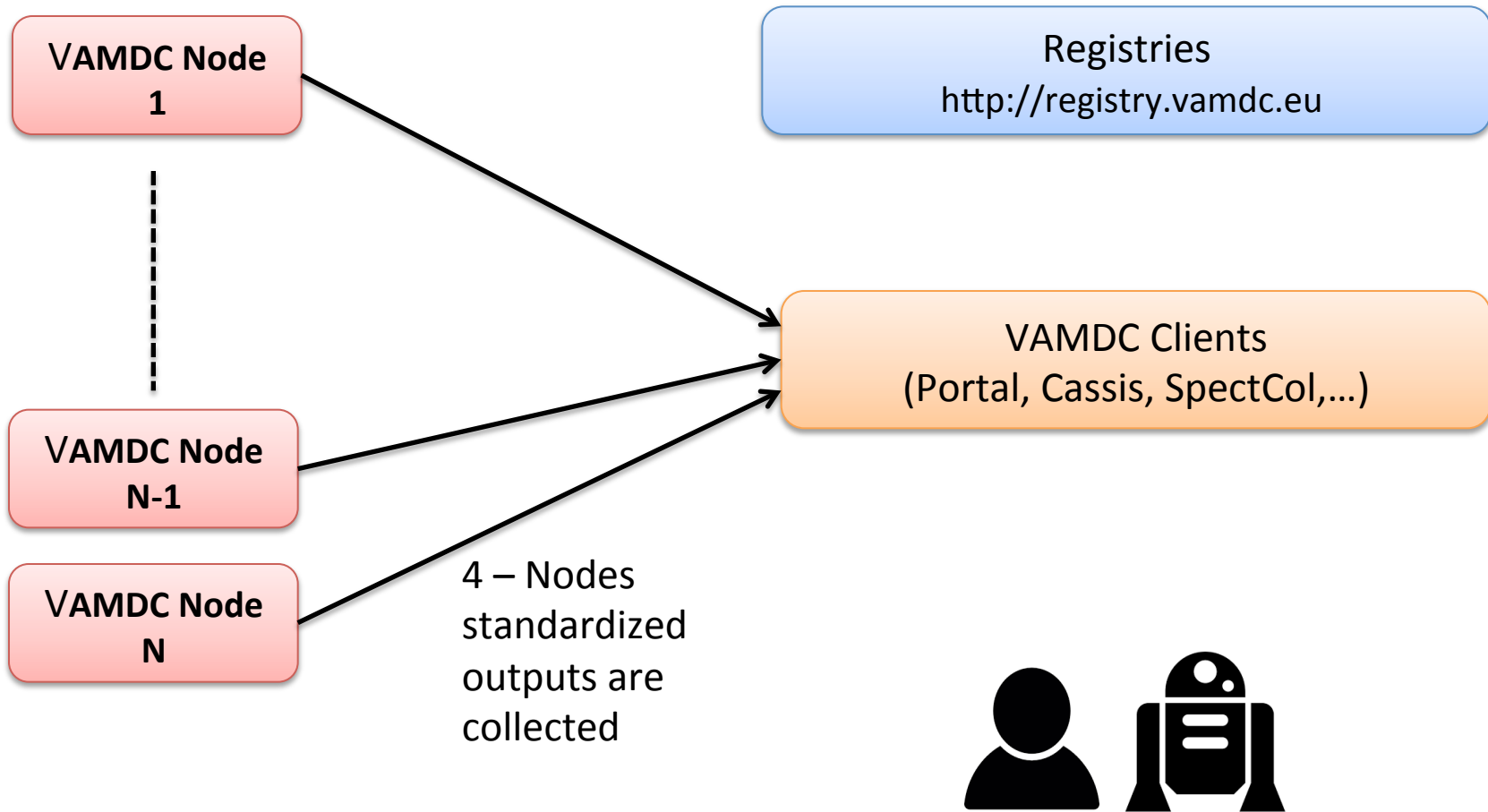
The VAMDC infrastructure technical architecture



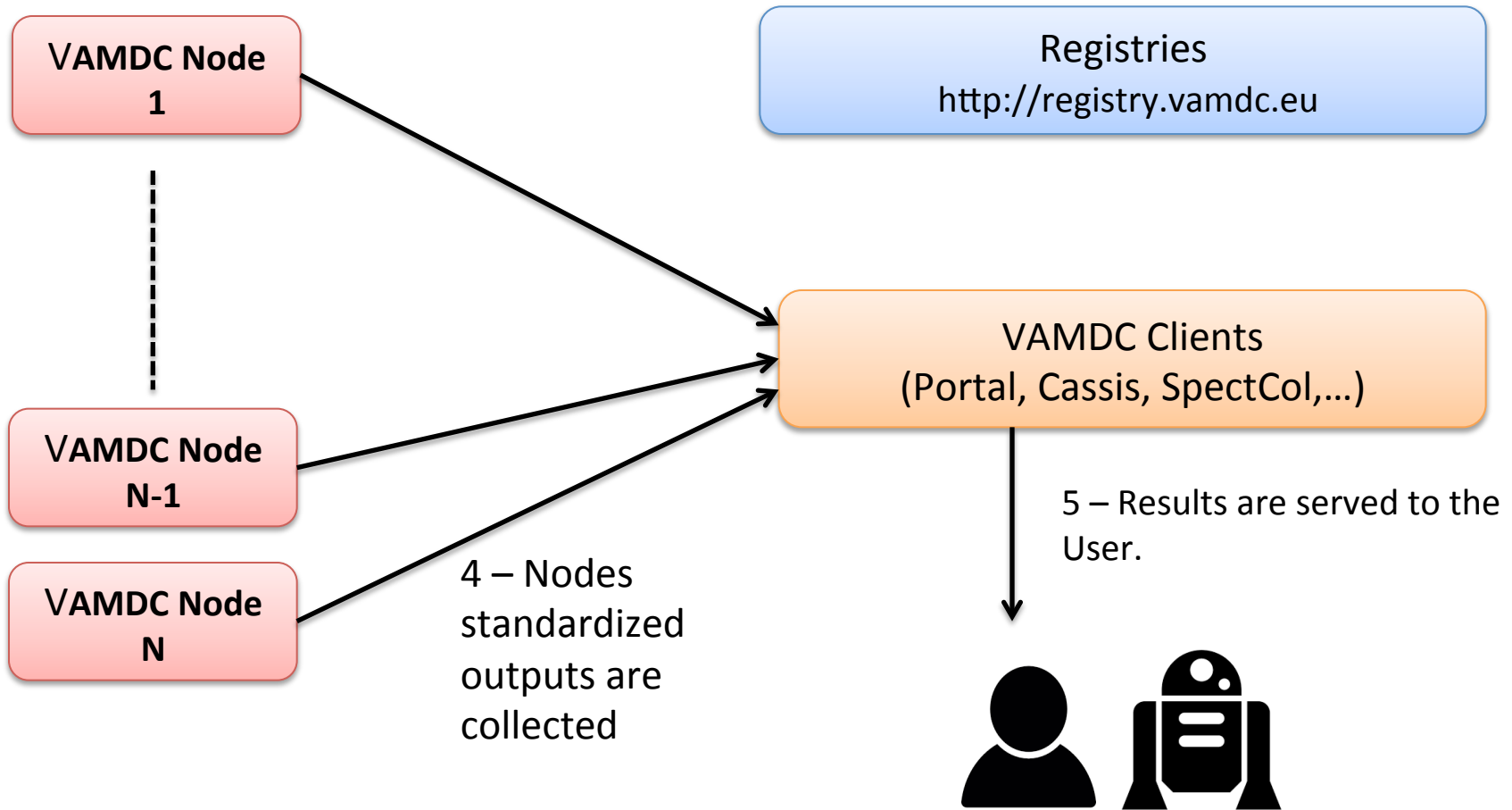
The VAMDC infrastructure technical architecture



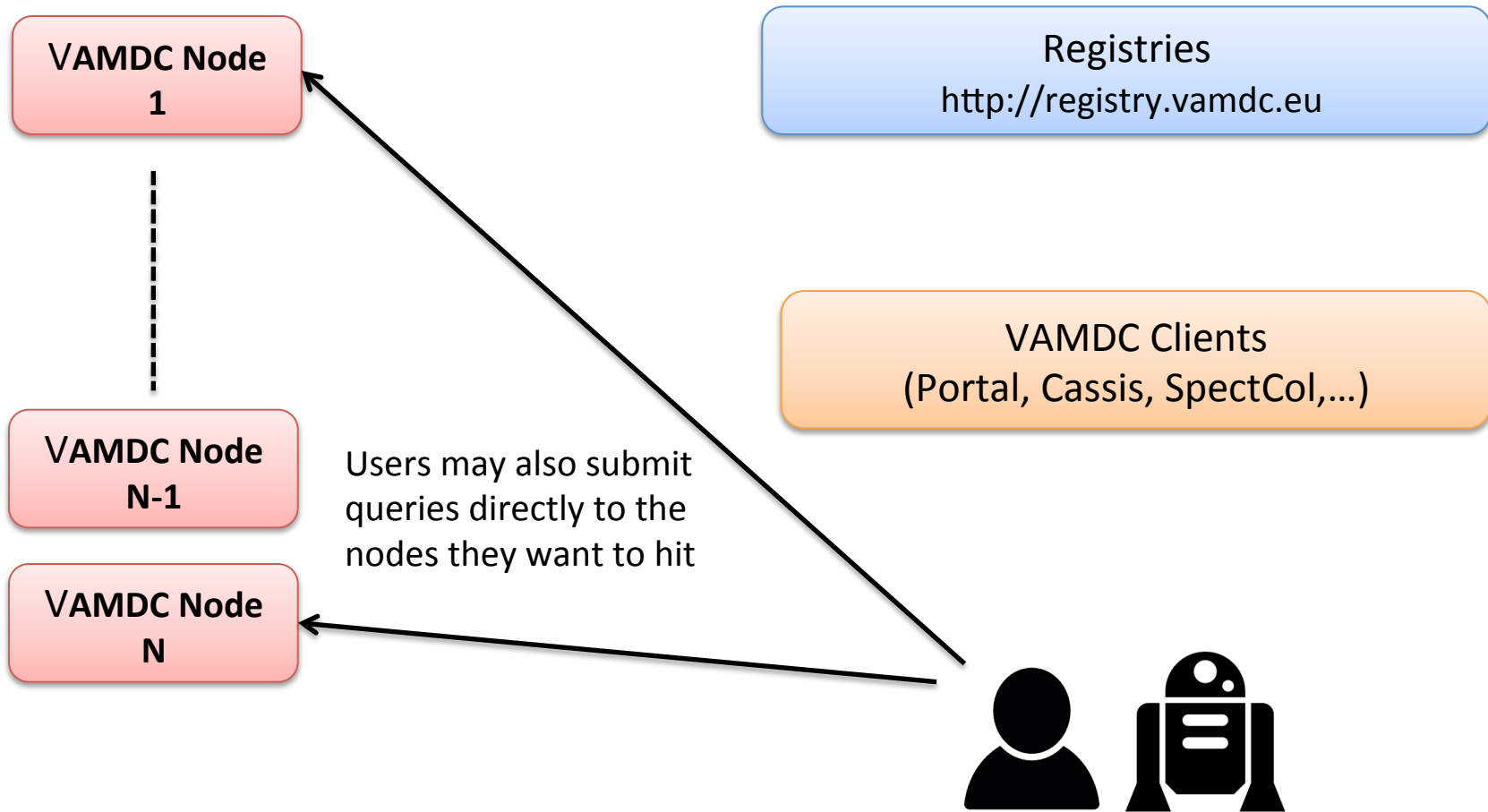
The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



Relevance of VAMDC for planetary community

The VAMDC infrastructure is cross-disciplinary. The databases containing data for planetary community are:

Database	Application	VAMDC partner
HITRAN	Planets, exo-planets	Harvard
SMPO	Earth, exo-planets	Reims + IAO Toms
MeCaSDa	Planets, exo-P, Brown Darfs	Dijon
CDSD	Planets, Brown Dwarfs	IAO Toms
W@SIS	Planets	IAO Toms
KIDA	Planets	Bordeaux
UdfA	Planets	Belfast
GhoSST	Planets	Grenoble
LASP	Planets	Catania Observatory
PAH	Planets	Cagliari Observatory
ECaSDa	Planets	Reims

Relevance of VAMDC for planetary community

The VAMDC infrastructure is cross-disciplinary. The databases containing data for planetary community are:

Database	Application	VAMDC partner
HITRAN	Planets, exo-planets	Harvard
SMPO	Earth, exo-planets	Reims + IAO Tomsk
MeCaSDa	Planets, exo-P, Brown Darfs	Dijon
CDSD	Planets, Brown Dwarfs	IAO Tomsk
W@SIS	Planets	IAO Tomsk
KIDA	Planets	Bordeaux
UdfA	Planets	Belfast
GhoSST	Planets	Grenoble
LASP	Planets	Catania Observatory
PAH	Planets	Cagliari Observatory
ECaSDa	Planets	Reims

E.G. Data for planetary atmospheres

Biomolecular reactions é
3body-assisted reactions

- Titan model from Hébrard et al. 2009
- Origianl data from Baulch et al. 2005 and Atkinson et al. 2006
- Hot Jupiters Data

Strengths of the VAMDC infrastructure

All data from VAMDC comes with detailed context information:

- References,
- Origins & Methods (theoretical, observed,...),
- Uncertainty.

We focus the end of the presentation on two added values of the VAMDC infrastructure

- Data Interoperability (example: Hitran output format from any VAMDC extracted data).
- Data Citation & Reproducibility of Research

XSAMS-TO-HITRAN

Needs :

- Molecular spectroscopy community : no way to read XSAMS and to compare data between databases
- Molecular data users : radiative transfer code, atmospheric models, ... use HITRAN file format
- VAMDC Portal has some conversion tools but no HITRAN format

XSAMS Processor:

- Java web service using XSAMS library provided by VAMDC
- Produce HITRAN2004 (160 characters) file from a valid XSAMS file
- Will be integrated in the VAMDC Portal (early 2018)
- Code is freely available on Github

HITRAN output example

Ozone (O₃) from JPL database

31	0.026341	5.502e-29	1.382e-13.0000.0000	495.51740.000.000000	0 0 0	0 0 0 31	5 27	32	4 28	600000	0 0 0 0 0 0	63.0	65.0
31	0.036284	9.341e-29	3.461e-13.0000.0000	522.55760.000.000000	0 0 0	0 0 0 33	4 30	32	5 27	600000	0 0 0 0 0 0	67.0	65.0
31	0.086529	5.595e-29	4.727e-12.0000.0000	1061.68640.000.000000	0 0 0	0 0 0 46	7 39	47	6 42	600000	0 0 0 0 0 0	93.0	95.0
32	0.108216	1.519e-27	1.235e-11.0000.0000	47.95820.000.000000	0 0 0	0 0 0 9	2 8	10	1 9	600000	0 0 0 0 0 0	19.0	21.0
33	0.127852	3.495e-28	3.594e-12.0000.0000	474.35710.000.000000	0 0 0	0 0 0 33	2 32	32	3 29	600000	0 0 0 0 0 0	67.0	65.0
33	0.155233	4.677e-27	2.326e-11.0000.0000	169.53180.000.000000	0 0 0	0 0 0 19	2 18	18	3 15	600000	0 0 0 0 0 0	39.0	37.0
31	0.167725	2.505e-28	4.101e-11.0000.0000	751.29220.000.000000	0 1 0	0 1 0 10	1 9	9	2 8	600000	0 0 0 0 0 0	21.0	19.0
32	0.169145	2.764e-27	3.294e-11.0000.0000	194.98000.000.000000	0 0 0	0 0 0 20	3 17	21	2 20	600000	0 0 0 0 0 0	41.0	43.0
31	0.197571	6.450e-29	5.523e-11.0000.0000	1401.61460.000.000000	0 0 0	0 0 0 53	8 46	54	7 47	600000	0 0 0 0 0 0	107.0	109.0
33	0.211917	4.276e-28	7.208e-11.0000.0000	804.44600.000.000000	0 1 0	0 1 0 16	2 14	15	3 13	600000	0 0 0 0 0 0	33.0	31.0
35	0.212814	2.164e-27	7.710e-11.0000.0000	127.47330.000.000000	0 1 0	0 1 0 15	3 13 17.5	16	2 14 18.5	600000	0 0 0 0 0 0	36.0	38.0
35	0.212819	1.576e-27	7.775e-11.0000.0000	127.47330.000.000000	0 1 0	0 1 0 15	3 13 12.5	16	2 14 13.5	600000	0 0 0 0 0 0	26.0	28.0
35	0.212830	2.032e-27	7.666e-11.0000.0000	127.47330.000.000000	0 1 0	0 1 0 15	3 13 16.5	16	2 14 17.5	600000	0 0 0 0 0 0	34.0	36.0
35	0.212831	1.680e-27	7.693e-11.0000.0000	127.47330.000.000000	0 1 0	0 1 0 15	3 13 13.5	16	2 14 14.5	600000	0 0 0 0 0 0	28.0	30.0
35	0.212837	1.790e-27	7.652e-11.0000.0000	127.47330.000.000000	0 1 0	0 1 0 15	3 13 14.5	16	2 14 15.5	600000	0 0 0 0 0 0	30.0	32.0
35	0.212837	1.907e-27	7.644e-11.0000.0000	127.47320.000.000000	0 1 0	0 1 0 15	3 13 15.5	16	2 14 16.5	600000	0 0 0 0 0 0	32.0	34.0
32	0.215346	6.660e-28	7.782e-11.0000.0000	762.61830.000.000000	0 0 0	0 0 0 40	6 35	41	5 36	600000	0 0 0 0 0 0	81.0	83.0
32	0.215554	5.559e-27	1.567e-10.0000.0000	7.71680.000.000000	0 0 0	0 0 0 4	0 4	3	1 3	600000	0 0 0 0 0 0	9.0	7.0
31	0.217284	6.206e-29	9.055e-11.0000.0000	1153.40990.000.000000	0 0 1	0 0 1 10	1 9	9	2 8	600000	0 0 0 0 0 0	21.0	19.0
33	0.232396	2.246e-27	9.199e-11.0000.0000	661.40210.000.000000	0 0 0	0 0 0 37	5 33	36	6 30	600000	0 0 0 0 0 0	75.0	73.0
32	0.239595	7.196e-28	1.044e-10.0000.0000	795.47590.000.000000	0 0 0	0 0 0 42	5 38	41	6 35	600000	0 0 0 0 0 0	85.0	83.0
32	0.271715	3.452e-28	1.955e-10.0000.0000	740.85410.000.000000	0 1 0	0 1 0 9	2 8	10	1 9	600000	0 0 0 0 0 0	19.0	21.0
33	0.273148	1.169e-27	1.534e-10.0000.0000	896.52380.000.000000	0 0 0	0 0 0 42	7 35	43	6 38	500000	0 0 0 0 0 0	85.0	87.0
35	0.278478	2.671e-27	1.451e-10.0000.0000	170.55510.000.000000	0 0 0	0 0 0 18	3 15 18.5	19	2 18 19.5	600000	0 0 0 0 0 0	38.0	40.0
35	0.278479	2.818e-27	1.454e-10.0000.0000	170.55510.000.000000	0 0 0	0 0 0 18	3 15 19.5	19	2 18 20.5	600000	0 0 0 0 0 0	40.0	42.0
35	0.278481	2.533e-27	1.452e-10.0000.0000	170.55510.000.000000	0 0 0	0 0 0 18	3 15 17.5	19	2 18 18.5	600000	0 0 0 0 0 0	36.0	38.0
35	0.278485	2.971e-27	1.460e-10.0000.0000	170.55520.000.000000	0 0 0	0 0 0 18	3 15 20.5	19	2 18 21.5	600000	0 0 0 0 0 0	42.0	44.0
35	0.278487	2.402e-27	1.458e-10.0000.0000	170.55510.000.000000	0 0 0	0 0 0 18	3 15 16.5	19	2 18 17.5	600000	0 0 0 0 0 0	34.0	36.0
35	0.278495	2.278e-27	1.469e-10.0000.0000	170.55520.000.000000	0 0 0	0 0 0 18	3 15 15.5	19	2 18 16.5	600000	0 0 0 0 0 0	32.0	34.0
34	0.289037	1.503e-27	2.987e-10.0000.0000	7.86120.000.000000	0 0 0	0 0 0 4	0 4 4.5	3	1 3 3.5	600000	0 0 0 0 0 0	10.0	8.0
34	0.289063	1.908e-27	3.158e-10.0000.0000	7.86110.000.000000	0 0 0	0 0 0 4	0 4 5.5	3	1 3 4.5	600000	0 0 0 0 0 0	12.0	10.0
34	0.289065	1.055e-27	2.619e-10.0000.0000	7.86110.000.000000	0 0 0	0 0 0 4	0 4 3.5	3	1 3 2.5	600000	0 0 0 0 0 0	8.0	6.0
34	0.289099	2.682e-27	3.807e-10.0000.0000	7.86110.000.000000	0 0 0	0 0 0 4	0 4 6.5	3	1 3 5.5	600000	0 0 0 0 0 0	14.0	12.0
34	0.296729	2.115e-27	2.127e-10.0000.0000	138.79160.000.000000	0 0 0	0 0 0 16	3 14 18.5	17	2 15 19.5	600000	0 0 0 0 0 0	38.0	40.0
34	0.296731	1.570e-27	2.143e-10.0000.0000	138.79160.000.000000	0 0 0	0 0 0 16	3 14 13.5	17	2 15 14.5	600000	0 0 0 0 0 0	28.0	30.0
34	0.296756	1.667e-27	2.123e-10.0000.0000	138.79160.000.000000	0 0 0	0 0 0 16	3 14 14.5	17	2 15 15.5	600000	0 0 0 0 0 0	30.0	32.0
34	0.296759	1.994e-27	2.117e-10.0000.0000	138.79150.000.000000	0 0 0	0 0 0 16	3 14 17.5	17	2 15 18.5	600000	0 0 0 0 0 0	36.0	38.0
34	0.296770	1.770e-27	2.113e-10.0000.0000	138.79150.000.000000	0 0 0	0 0 0 16	3 14 15.5	17	2 15 16.5	600000	0 0 0 0 0 0	32.0	34.0
34	0.296771	1.879e-27	2.112e-10.0000.0000	138.79150.000.000000	0 0 0	0 0 0 16	3 14 16.5	17	2 15 17.5	600000	0 0 0 0 0 0	34.0	36.0
31	0.302783	7.561e-28	4.258e-10.0000.0000	708.98330.000.000000	0 1 0	0 1 0 4	0 4	3	1 3	600000	0 0 0 0 0 0	9.0	7.0
31	0.306889	1.681e-26	1.795e-10.0000.0000	205.32850.000.000000	0 0 0	0 0 0 21	2 20	20	3 17	600000	0 0 0 0 0 0	43.0	41.0
32	0.307737	6.934e-27	2.206e-10.0000.0000	328.41500.000.000000	0 0 0	0 0 0 26	4 22	27	3 25	600000	0 0 0 0 0 0	53.0	55.0
32	0.321635	9.324e-27	2.089e-10.0000.0000	211.96640.000.000000	0 0 0	0 0 0 22	2 21	21	3 18	600000	0 0 0 0 0 0	45.0	43.0
32	0.327670	3.990e-27	2.719e-10.0000.0000	523.00660.000.000000	0 0 0	0 0 0 33	5 28	34	4 31	600000	0 0 0 0 0 0	67.0	69.0
33	0.333919	5.224e-27	2.805e-10.0000.0000	630.83330.000.000000	0 0 0	0 0 0 35	6 30	36	5 31	600000	0 0 0 0 0 0	71.0	73.0
31	0.341093	3.072e-26	3.549e-10.0000.0000	50.30210.000.000000	0 0 0	0 0 0 10	1 9	9	2 8	600000	0 0 0 0 0 0	21.0	19.0
31	0.342647	6.602e-29	1.400e-10.0000.0000	1230.59110.000.000000	0 0 1	0 0 1 16	2 14	17	1 17	600000	0 0 0 0 0 0	33.0	35.0
31	0.350853	1.511e-28	6.785e-10.0000.0000	1111.16810.000.000000	0 0 1	0 0 1 4	0 4	3	1 3	600000	0 0 0 0 0 0	9.0	7.0
31	0.354025	2.252e-27	8.151e-11.0000.0000	534.74910.000.000000	0 0 0	0 0 0 35	2 34	34	3 31	600000	0 0 0 0 0 0	71.0	69.0
31	0.357105	2.125e-28	3.520e-10.0000.0000	1182.73040.000.000000	1 0 0	1 0 0 17	2 15	16	3 14	600000	0 0 0 0 0 0	35.0	33.0
31	0.359772	7.847e-29	8.490e-11.0000.0000	1233.86120.000.000000	0 1 0	0 1 0 34	3 31	35	2 34	600000	0 0 0 0 0 0	69.0	71.0
31	0.369337	3.338e-26	7.953e-10.0000.0000	8.02170.000.000000	0 0 0	0 0 0 4	0 4	3	1 3	600000	0 0 0 0 0 0	9.0	7.0
32	0.387909	4.503e-28	3.514e-10.0000.0000	922.26710.000.000000	0 1 0	0 1 0 23	2 22	22	3 19	600000	0 0 0 0 0 0	47.0	45.0
31	0.388316	6.076e-28	4.448e-10.0000.0000	1507.66000.000.000000	0 0 1	0 0 1 31	5 27	32	4 28	600000	0 0 0 0 0 0	63.0	65.0

Future services

- Including **GEISA** into the VAMDC infrastructure
- **GEISA format** XSAMS processor
- This approach can be generalized to any format

Data citation and reproducibility issues

Data citation and reproducibility are two facets of the same problem

- Provide somebody with all the elements and information about the production of a given result.



Data citation & Reproducibility

≠ Repeatability



Virtual machines and containers (e.g. Docker) may be distributed for sharing processes:

- These tools work as black-boxes
- Result may be repeated, but
- With no information this is useless for our purposes.

Data citation and reproducibility issues

Data citation and reproducibility are two facets of the same problem

- Provide somebody with all the elements and information about the production of a given result.



Data citation & Reproducibility

≠ Repeatability

This work is made in collaboration with the Research Data Alliance,
since 2014

Data Citation WG



Group details

Status: [Recognised & Endorsed](#)

Chair(s): [Andreas Rauber](#), [Ari Asmi](#), [Dieter van Uytvanck](#)

Case Statement: [Download](#)

The RDA Working Group on Data Citation (WG-DC) aims to bring together a group of experts to discuss the issues, requirements, advantages and shortcomings of existing approaches for efficiently citing subsets of data. The WG-DC focuses on a narrow field where we can contribute significantly and provide prototypes and reference implementations.



We applied the RDA
recommendation by
implementing over the
VAMDC infrastructure a
Query Store

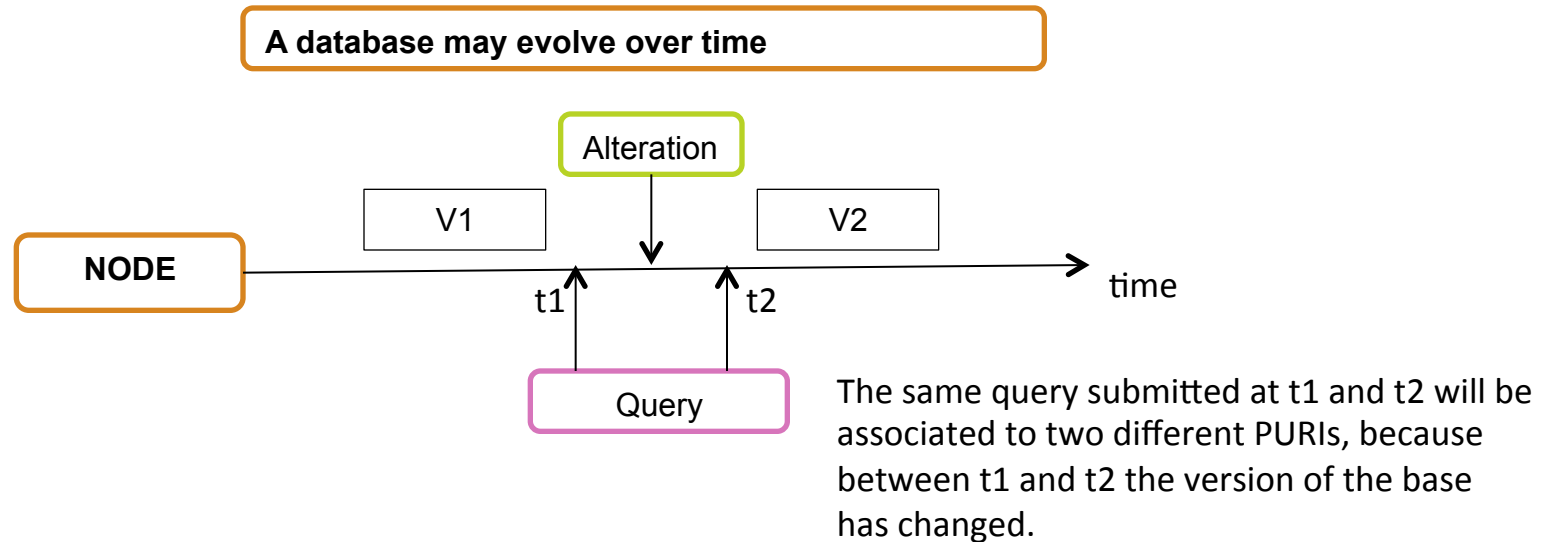
The VAMDC Query Store

- Query Store principles:
 - Each unique query will be associated with a Persistent Unique Resolvable Identifier (PURI)
 - The landing page associated with the PURI resolution gives state information about the Query + bibliographic information

The VAMDC Query Store

- Query Store principles:
 - Each unique query will be associated with a Persistent Unique Resolvable Identifier (PURI)
 - The landing page associated with the PURI resolution gives state information about the Query + bibliographic information
- It was challenging to implement the Query Store on VAMDC (Distributed infrastructure with no central management system)
- We were sub-contractors of RDA-EU3 project
- Architecture + source code + documentation on <https://github.com/VAMDC/QueryStore>
- Deployed in functional validation on : <http://cite.vamdc.eu/>
 - Service both human & machine oriented.
- Video short demonstration at <https://www.youtube.com/watch?v=kDDWFpi22cU>
 - Full video tutorial at <https://www.youtube.com/watch?v=OLe-qcqCcCw&t=5s>

Applications of the Query Store: reproducibility



- Two different PURIs for the same query is a warning to user: something has changed. Result may not be reproducible.
- What has really changed?
 - The alteration may concern He and the query is on H. The query result is unchanged even if the Node version (and PURIs) are different.
 - The approach described in ***New model for datasets citation and extraction reproducibility in VAMDC***, <http://dx.doi.org/10.1016/j.jms.2016.04.009> (Arxiv version: <https://arxiv.org/abs/1606.00405>) is a mechanisms for knowing exactly what changed between v1 and v2.

Applications of the Query Store: working with editors

- The Query Store is not only a technical R&D challenge.
 - It raises question on the collaboration models between data producers & editors.
 - Editors are interested by the Query Store for the Paper-data linking from their website and for building bibliometrics information
 - What is the collaboration model?
 - Who pay for data/metadata storage and for Query Store hosting?
 - Ongoing discussion with Elsevier.
 - Editors may connect with the machine oriented Query Store Api

Applications of the Query Store: working with editors

- The Query Store is not only a technical R&D challenge.
 - It raises question on the collaboration models between data producers & editors.
 - Editors are interested by the Query Store for the Paper-data linking from their website and for building bibliometrics information
 - What is the collaboration model?
 - Who pay for data/metadata storage and for Query Store hosting?
 - Ongoing discussion with Elsevier.
 - Editors may connect with the machine oriented Query Store Api

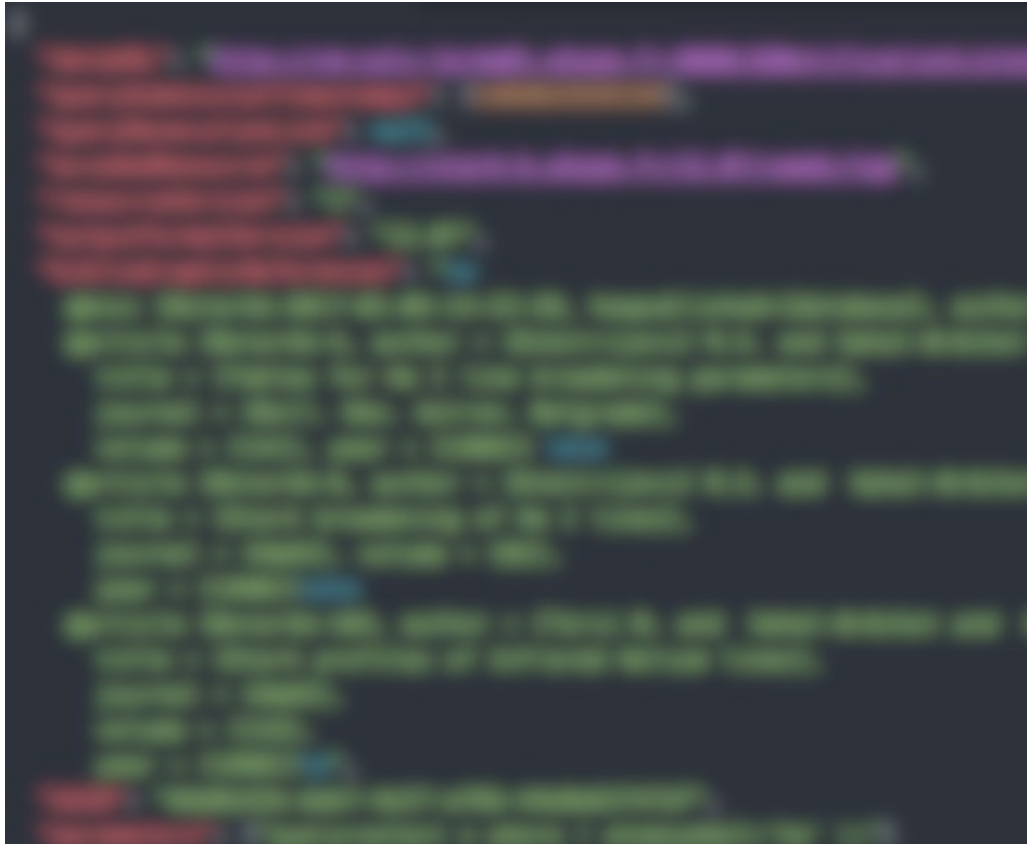
Example of machine oriented service output

```
{
  "dataURL": "http://vm-calc-lerma01.obspm.fr:8080/QSNotificationListener/data/66d6b32b-dab7-4e27-af6b-44e8e62f4fdf.xsams",
  "querySubmissionTimestamps": [1483622535134],
  "queryRexecutionLink": null,
  "accededResource": "http://stark-b.obspm.fr/12.07/vamdc/tap",
  "resourceVersion": "1",
  "outputFormatVersion": "12.07",
  "biblioGraphicReferences": "\n
    @misc {Bstarkb-2017-01-05-14-22-28, howpublished={database}, author = {N.N.}, year = {2017}, url={http://stark-b.obspm.f
    @article {Bstarkb-4, author = {Dimitrijević M.S. and Sahal-Bréchet S.},
      title = {Tables for He I line broadening parameters},
      journal = {Bull. Obs. Astron. Belgrade},
      volume = {141}, year = {1989}} \n\n
    @article {Bstarkb-6, author = {Dimitrijević M.S. and Sahal-Bréchet S.},
      title = {Stark broadening of He I lines},
      journal = {A&AS}, volume = {82},
      year = {1990}}\n\n
    @article {Bstarkb-103, author = {Terzi N. and Sahal-Bréchet and S. Ben Lakhdar Z.},
      title = {Stark profiles of Infrared Helium lines},
      journal = {A&AS},
      volume = {133},
      year = {1998}}\n",
  "UUID": "66d6b32b-dab7-4e27-af6b-44e8e62f4fdf",
  "parameters": ["query=select * where ( atomsymbol='he' ):"]
```

Applications of the Query Store: working with editors

- The Query Store is not only a technical R&D challenge.
 - It raises question on the collaboration models between data producers & editors.
 - Editors are interested by the Query Store for the Paper-data linking from their website and for building bibliometrics information
 - What is the collaboration model?
 - Who pay for data/metadata storage and for Query Store hosting?
 - Ongoing discussion with Elsevier.
 - Editors may connect with the machine oriented Query Store Api

Example of machine oriented service output



Authors may cite in article only the UUIDs

- Delegation of credits applies: editors may use this service for give credits to all the authors of paper appearing into the bibtex.
- This mechanisms is transparent and easy for authors (only a string to cite vs. dozen of papers).

Strengths of the Query Store:

- The QS usage is transparent for users (complexity is hidden).
- Live monitoring of all the queries and users of the VAMDC e-infrastructure
 - Data providers may measure their impact and have detailed statistics of usage.
- It will be easy for authors to cite data coming from VAMDC. Credit to producers will be automatic.