



Data Citation Working Group Mtg @ P9 April 5th 2017, Barcelona

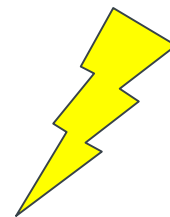
research data sharing without barriers
rd-alliance.org

- 14:00 Introduction, Welcome
- 14:10 Short description of the WG recommendations
- 14:25 Report on Standardisation & Adoption
- 14:30 Brief reports on use cases
 - Virtual and Atomic Molecular Data Center (C.M. Zwolf)
 - EURO ARGO (TBA)
 - BCO-DMO (Slides from C. Chandler, presented by Asmi)
 - Pilot project in Japan (Yasuhiro Muyarama)
 - Others
- 15:10 Collaboration with other WGs and IGs
- 15:20 Other issues, future of the WG

Welcome!
to the adoption meeting
of the
WGDC

- 14:00 Introduction, Welcome
- 14:10 Short description of the WG recommendations
- 14:25 Report on Standardisation & Adoption
- 14:30 Brief reports on use cases
 - Virtual and Atomic Molecular Data Center (C.M. Zwolf)
 - EURO ARGO (TBA)
 - BCO-DMO (Slides from C. Chandler, presented by Asmi)
 - Pilot project in Japan (Yasuhiro Muyarama)
 - Others
- 15:10 Collaboration with other WGs and IGs
- 15:20 Other issues, future of the WG

- Usually, datasets have to be static
 - Fixed set of data, no changes:
no corrections to errors, no new data being added
- But: (research) data is **dynamic**
 - Adding new data, correcting errors, enhancing data quality, ...
 - Changes sometimes highly dynamic, at irregular intervals
- Current approaches
 - Identifying entire data stream, without any versioning
 - Using “accessed at” date
 - “Artificial” versioning by identifying batches of data (e.g. annual), aggregating changes into releases (time-delayed!)
- Would like to identify precisely the **data as it existed at a specific point in time**



Granularity of Subsets

- What about the **granularity** of data to be identified?
 - Enormous amounts of CSV data
 - Researchers use specific subsets of data
 - Need to identify precisely the subset used
- Current approaches
 - Storing a copy of subset as used in study -> scalability
 - Citing entire dataset, providing textual description of subset -> imprecise (ambiguity)
 - Storing list of record identifiers in subset -> scalability, not for arbitrary subsets (e.g. when not entire record selected)
- Would like to be able to identify precisely the **subset of (dynamic) data used** in a process



- Research Data Alliance
- WG on **Data Citation: Making Dynamic Data Citeable**
- March 2014 – September 2015
 - Concentrating on the problems of **large, dynamic (changing) datasets**
- Final version presented Sep 2015 at P7 in Paris, France
- Endorsed September 2016 at P8 in Denver, CO



<https://www.rd-alliance.org/groups/data-citation-wg.html>

- We have
 - Data & some means of access („query“)

- We have
 - Data & some means of access („query“)
- Make data: time-stamped and versioned
- Prepare some way of storing the queries

- We have
 - Data & some means of access („query“)
- Make data: time-stamped and versioned
- Prepare some way of storing the queries
- Data Citation:
 - Store query with timestamp
 - **Assign the PID to the timestamped query**
(which, dynamically, leads to the data)
- Access:
 - Re-execute query on versioned data according to timestamp
- Dynamic Data Citation:
 - Dynamic data & dynamic citation of data

- Researcher uses workbench to identify subset of data
- Upon executing selection („download“) user gets
 - Data (package, access API, ...)
 - PID (e.g. DOI) (Query is time-stamped and stored)
 - Hash value computed over the data for local storage
 - Recommended citation text (e.g. BibTeX)
- PID resolves to landing page
 - Provides detailed metadata, link to parent data set, subset,...
 - Option to retrieve original data OR current version OR changes
- Upon activating PID associated with a data citation
 - Query is re-executed against time-stamped and versioned DB
 - Results as above are returned
- Query store aggregates data usage

Data Citation – Deployment

12

- **Note: query string provides excellent provenance information on the data set!**
- subset of data per gets
 - Data (package, access API, ...)
 - PID (e.g. DOI) (Query is time-stamped and stored)
 - Hash value computed over the data for local storage
 - Recommended citation text (e.g. BibTeX)
- PID resolves to landing page
 - Provides detailed metadata, link to parent data set, subset,...
 - Option to retrieve original data OR current version OR changes
- Upon activating PID associated with a data citation
 - Query is re-executed against time-stamped and versioned DB
 - Results as above are returned
- Query store aggregates data usage

Data Citation – Deployment

13

- Note: query string provides excellent provenance information on the data set!
- This is an important advantage over traditional approaches relying on, e.g. storing a list of identifiers/DB dump!!!
 - Data (package)
 - PID (e.g. DOI)
 - Hash value
 - Recommended citation text (e.g. BibTeX)
- PID resolves to landing page
 - Provides detailed metadata, link to parent data set, subset,...
 - Option to retrieve original data OR current version OR changes
- Upon activating PID associated with a data citation
 - Query is re-executed against time-stamped and versioned DB
 - Results as above are returned
- Query store aggregates data usage

Data Citation – Deployment

14

- **Note: query string provides excellent provenance information on the data set!**

- Data (package)
- PID (e.g. DOI)
- Hash value
- Recommended citation text (e.g. PID text)

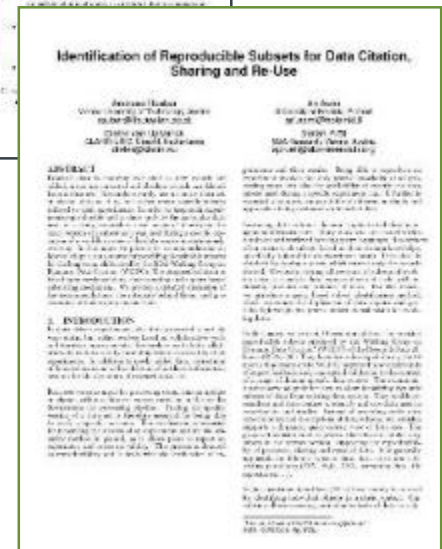
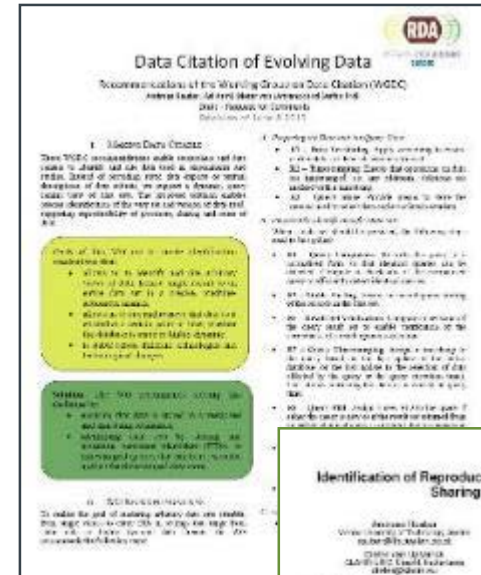
This is an important advantage over traditional approaches relying on, e.g. storing a list of identifiers/DB dump!!!

- PID resolves
 - Provides details
 - Option to retrieve
- Upon activating PID associated with a data citation
 - Query is re-executed against time-stamped and versioned DB
 - Results as above are returned
- Query store aggregates data usage

Identify which parts of the data are used. If data changes, identify which queries (studies) are affected



- **14 Recommendations** grouped into 4 phases:
 - Preparing data and query store
 - Persistently identifying specific data sets
 - Resolving PIDs
 - Upon modifications to the data infrastructure
- **2-page flyer**
<https://rd-alliance.org/recommendations-working-group-data-citation-revision-oct-20-2015.html>
- **More detailed report: IEEE TCDL 2016**
http://www.ieee-tcdl.org/Bulletin/v12n1/papers/IEEE-TCDL-DC-2016_paper_1.pdf



- 14:00 Introduction, Welcome
- 14:10 Short description of the WG recommendations
- 14:25 Report on Standardisation & Adoption
- 14:30 Brief reports on use cases
 - Virtual and Atomic Molecular Data Center (C.M. Zwolf)
 - EURO ARGO (TBA)
 - BCO-DMO (Slides from C. Chandler, presented by Asmi)
 - Pilot project in Japan (Yasuhiro Muyarama)
 - Others
- 15:10 Collaboration with other WGs and IGs
- 15:20 Other issues, future of the WG

- RDA applied for WGDC recommendations to become **ICT Technical Specification**
- Application to be considered by MSP (European Multi Stakeholder Platform) on 7 March
- Hillary Hanahoe presents to MSP on 23 March
- Evaluation group to investigate the Market Acceptance (basically the adoption & implementation) in April
- Evaluation group will make a recommendation to the MSP in June and the MSP will vote
- If favourable, the recommendation will be officially recognised as ICT Tech Specifications

- **Series of Webinars: Adoption reports**
<https://www.rd-alliance.org/group/data-citation-wg/webconference/webconference-data-citation-wg.html>
- **Adoption of the RDA Data Citation of Evolving Data Recommendation to Electronic Health Records**
Leslie McIntosh, PHD, MPH, Director Center for Biomedical Informatics, Washington University in St.Luis
Tue, Jan 17 2017
- **Implementation of Dynamic Data Citation at the Vermont Monitoring Cooperative** Presenter: James Duncan, VMC, University of Vermont, Burlington, VT
Mon, Feb 13 2017
- **Implementing the RDA Data Citation Recommendations in the Distributed Infrastructure of the Virtual and Atomic Molecular Data Center (VAMDC)**
Presenter: Carlo Maria Zwölf, VAMDC, Observatoire de Paris, FR
Fri, Mar 31 2017

- **Series of Webinars**

<https://www.rd-alliance.org/group/data-citation-wg/webconference/webconference-data-citation-wg.html>

- All webinars available for off-line viewing
- More webinars to come
 - Long-tail data / CSV (Stefan Pröll)
 - ???

- 14:00 Introduction, Welcome
- 14:10 Short description of the WG recommendations
- 14:25 Report on Standardisation & Adoption
- 14:30 Brief reports on use cases
 - Virtual and Atomic Molecular Data Center (C.M. Zwolf)
 - EURO ARGO (TBA)
 - BCO-DMO (Slides from C. Chandler, presented by Asmi)
 - Pilot project in Japan (Yasuhiro Muyarama)
 - Others
- 15:10 Collaboration with other WGs and IGs
- 15:20 Other issues, future of the WG



Implementing WGDC Recommendations at VAMDC

**C.M. Zwölf, N. Moreau,
VAMDC Consortium
carlo-maria.zwolf@obspm.fr**

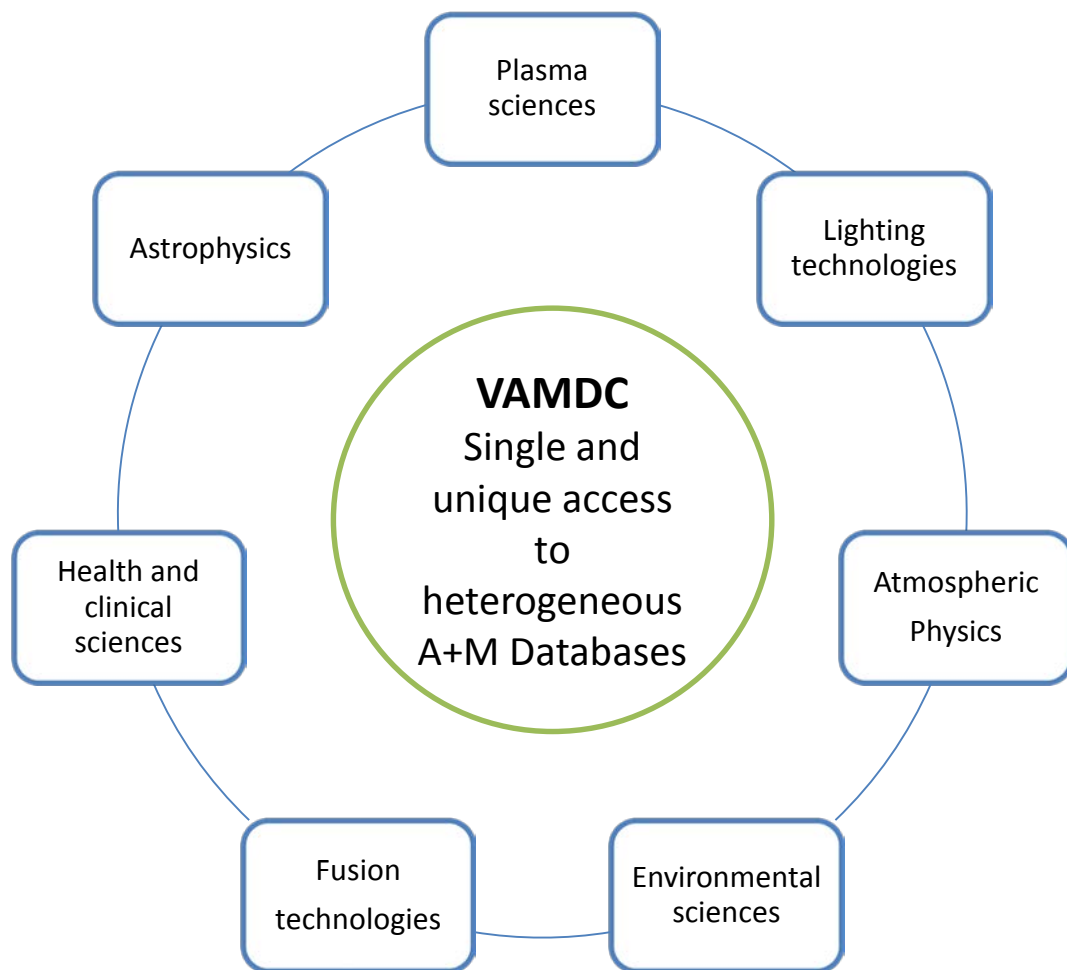
research data sharing without barriers
rd-alliance.org

Implementing the RDA data citation recommendations in the distributed Infrastructure of the Virtual Atomic and Molecular Data Centre

C.M. Zwölf, N. Moreau and 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.

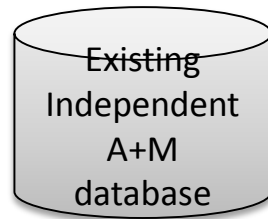
➤ The consortium is politically organized around a Memorandum of understanding (15 international members have signed the MoU, 1 November 2014)

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

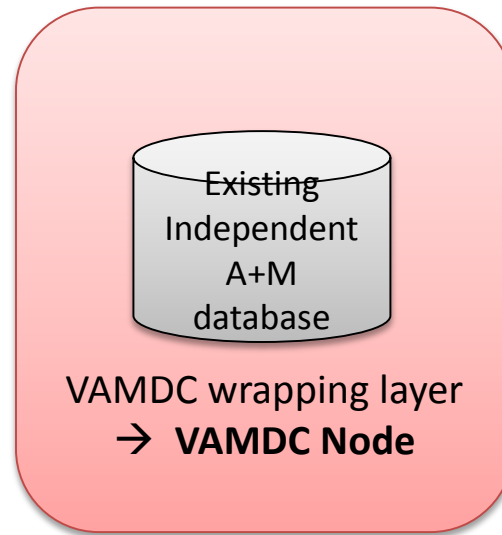
➤ Provides data producers with a large dissemination platform

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

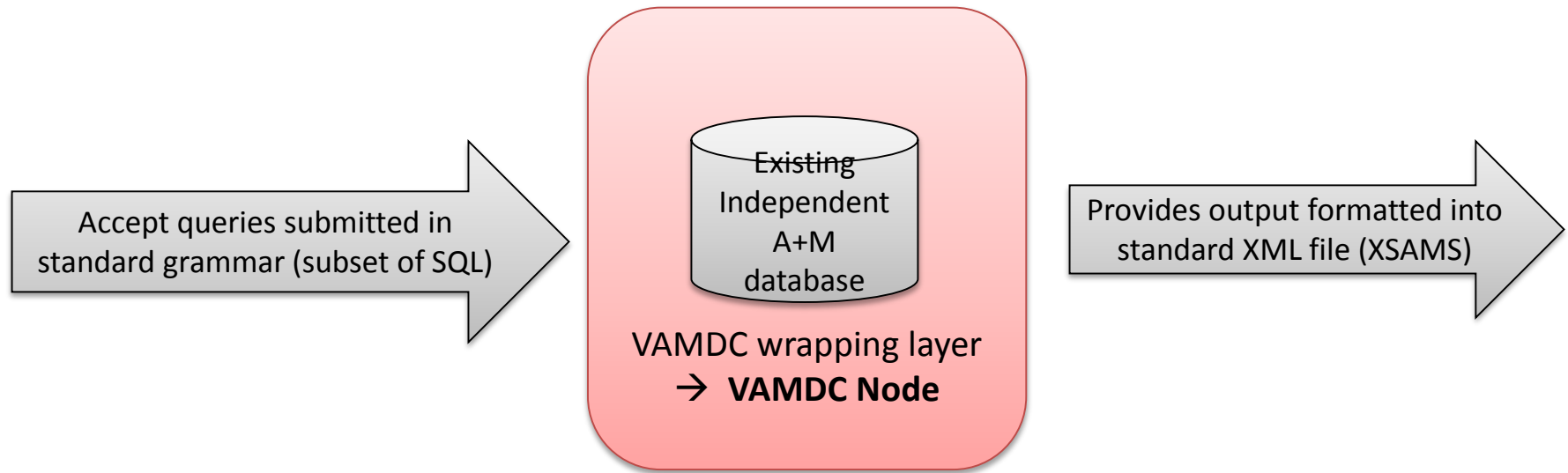
The VAMDC infrastructure technical architecture



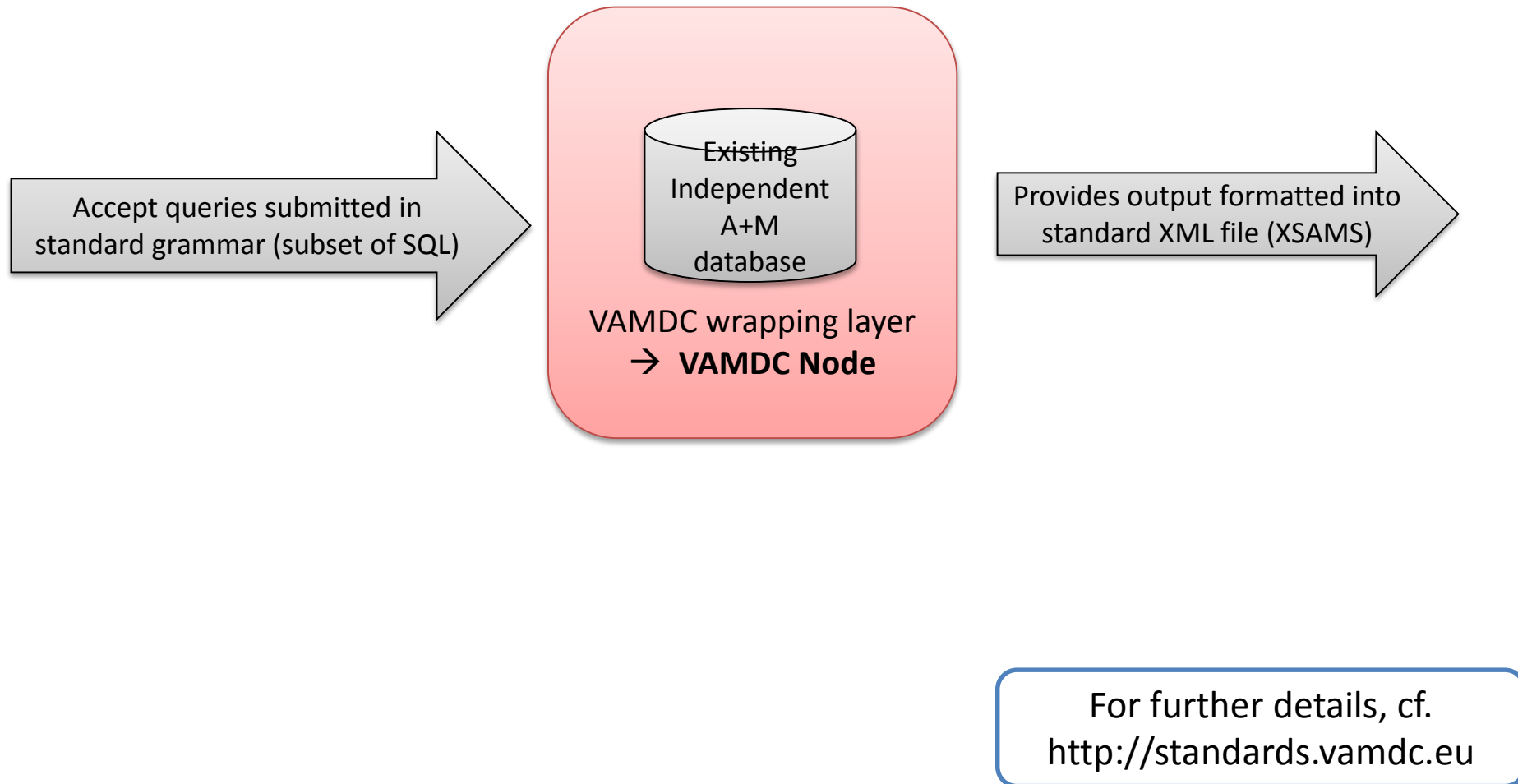
The VAMDC infrastructure technical architecture



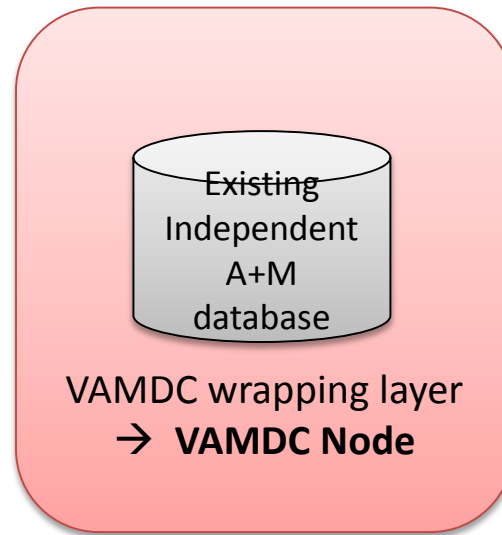
The VAMDC infrastructure technical architecture



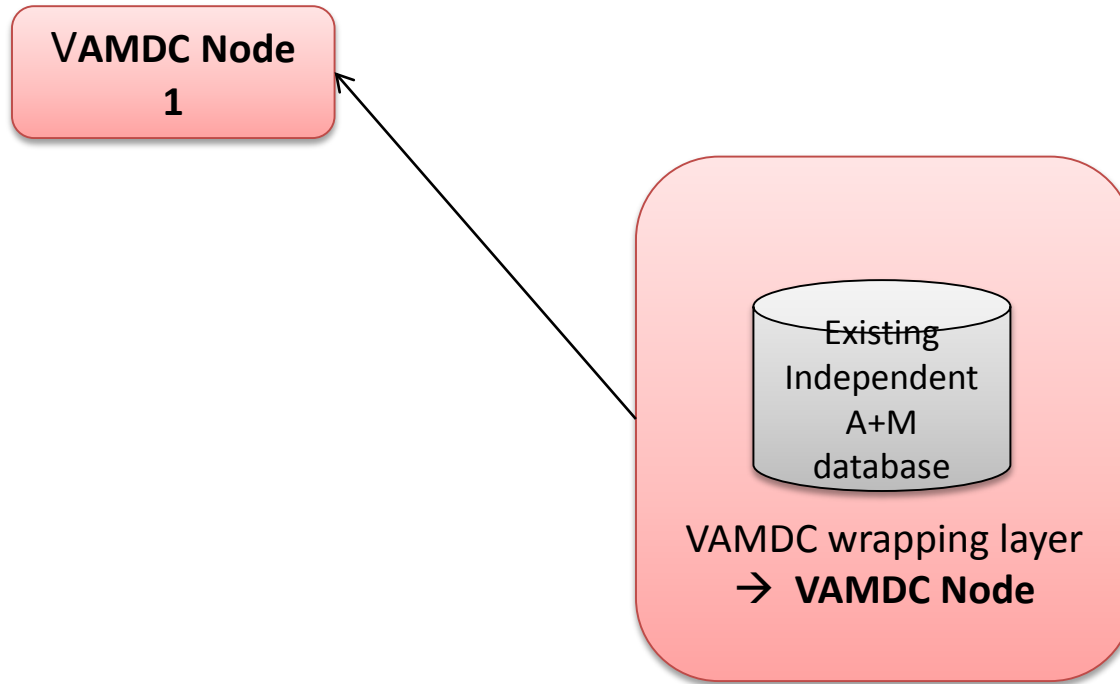
The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture

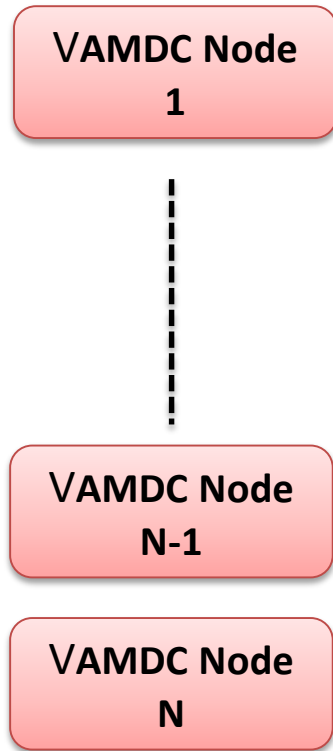


The VAMDC infrastructure technical architecture

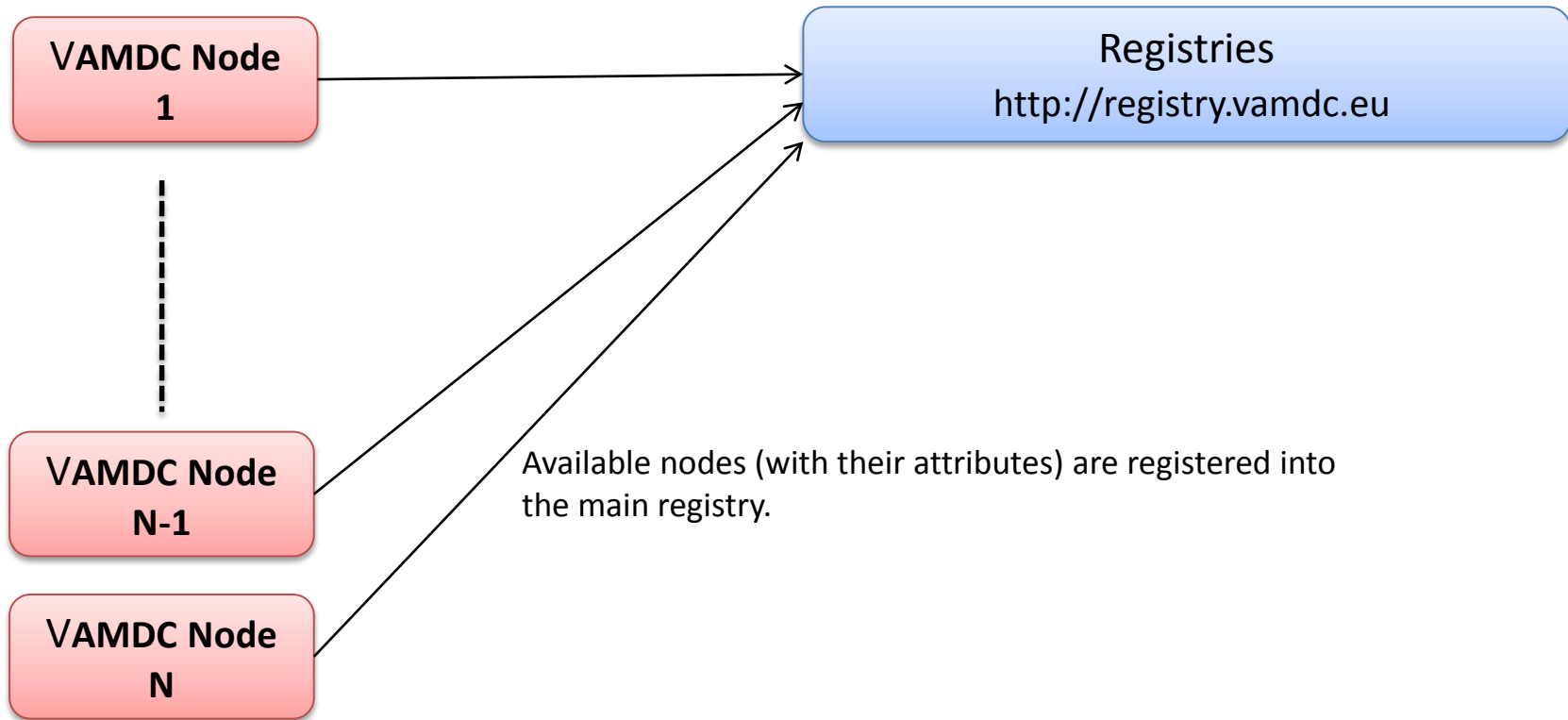
VAMDC Node

1

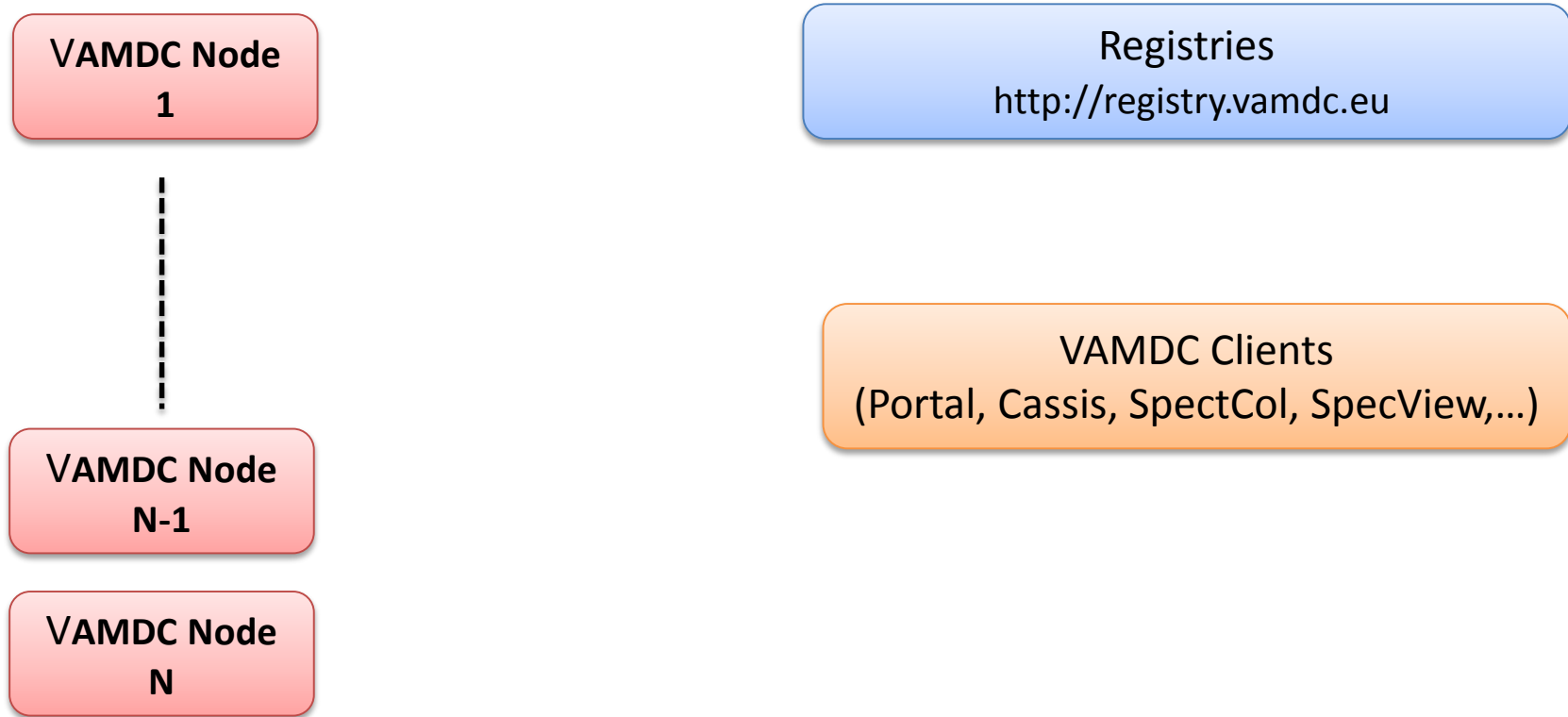
The VAMDC infrastructure technical architecture



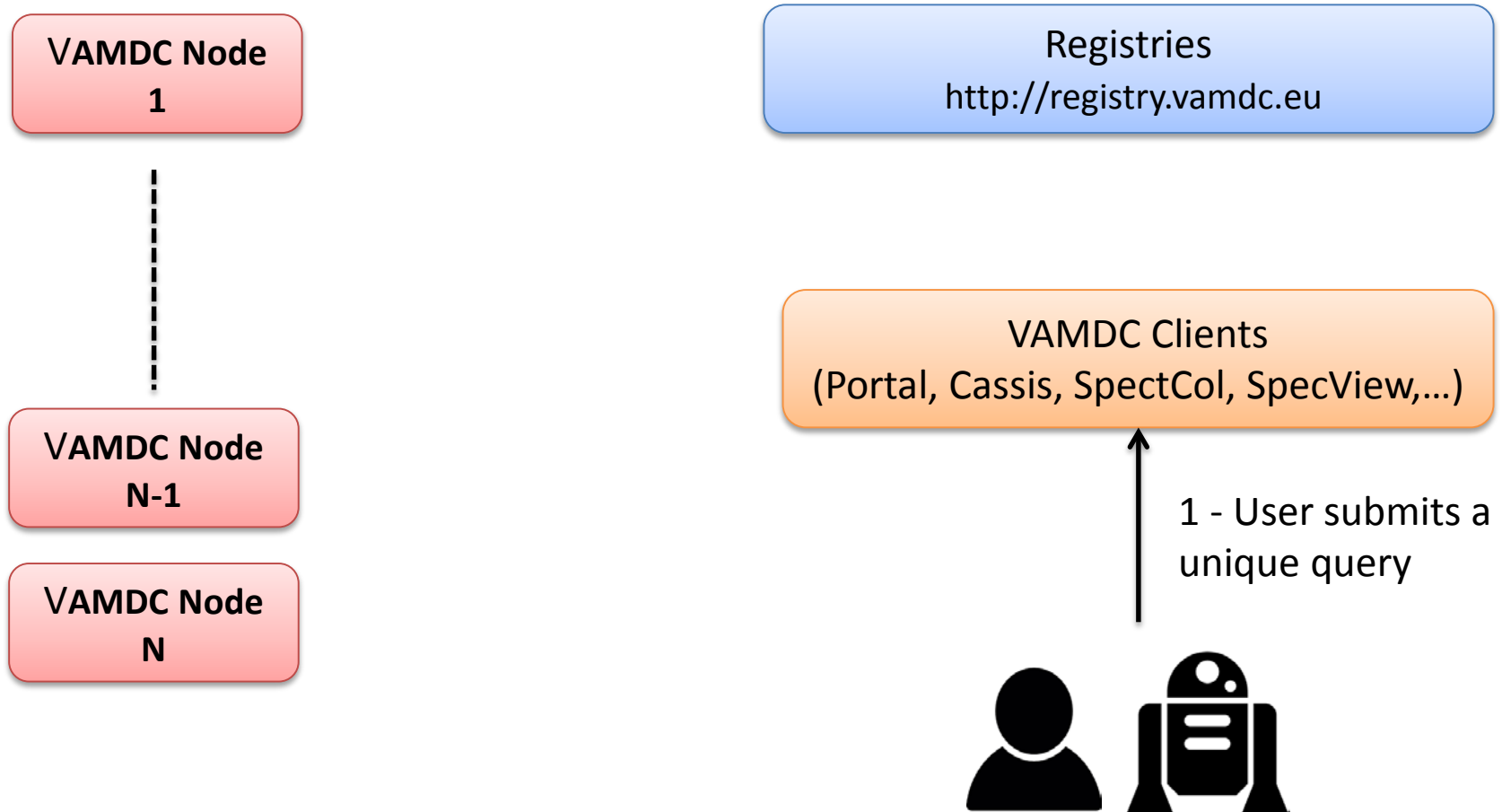
The VAMDC infrastructure technical architecture



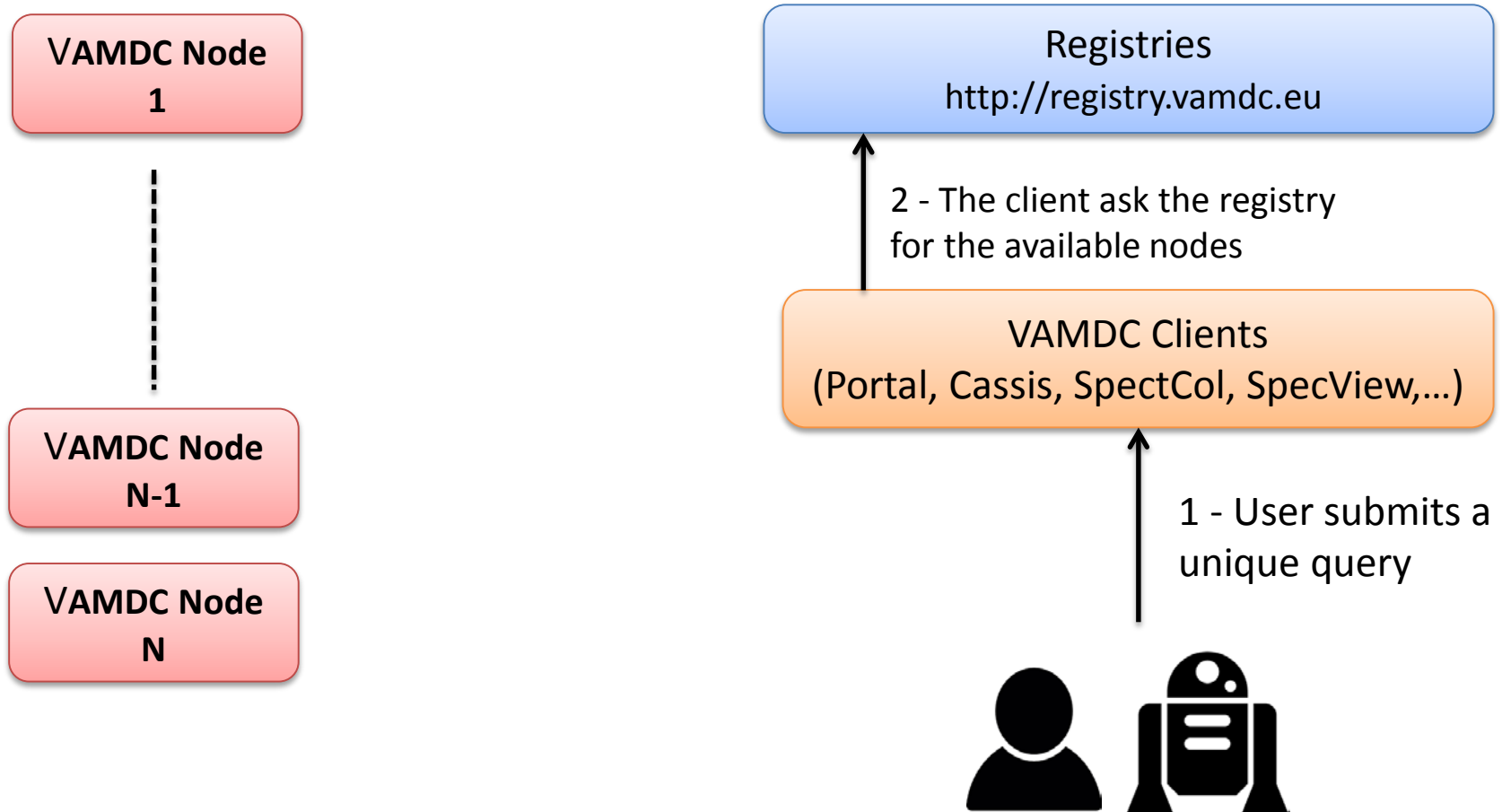
The VAMDC infrastructure technical architecture



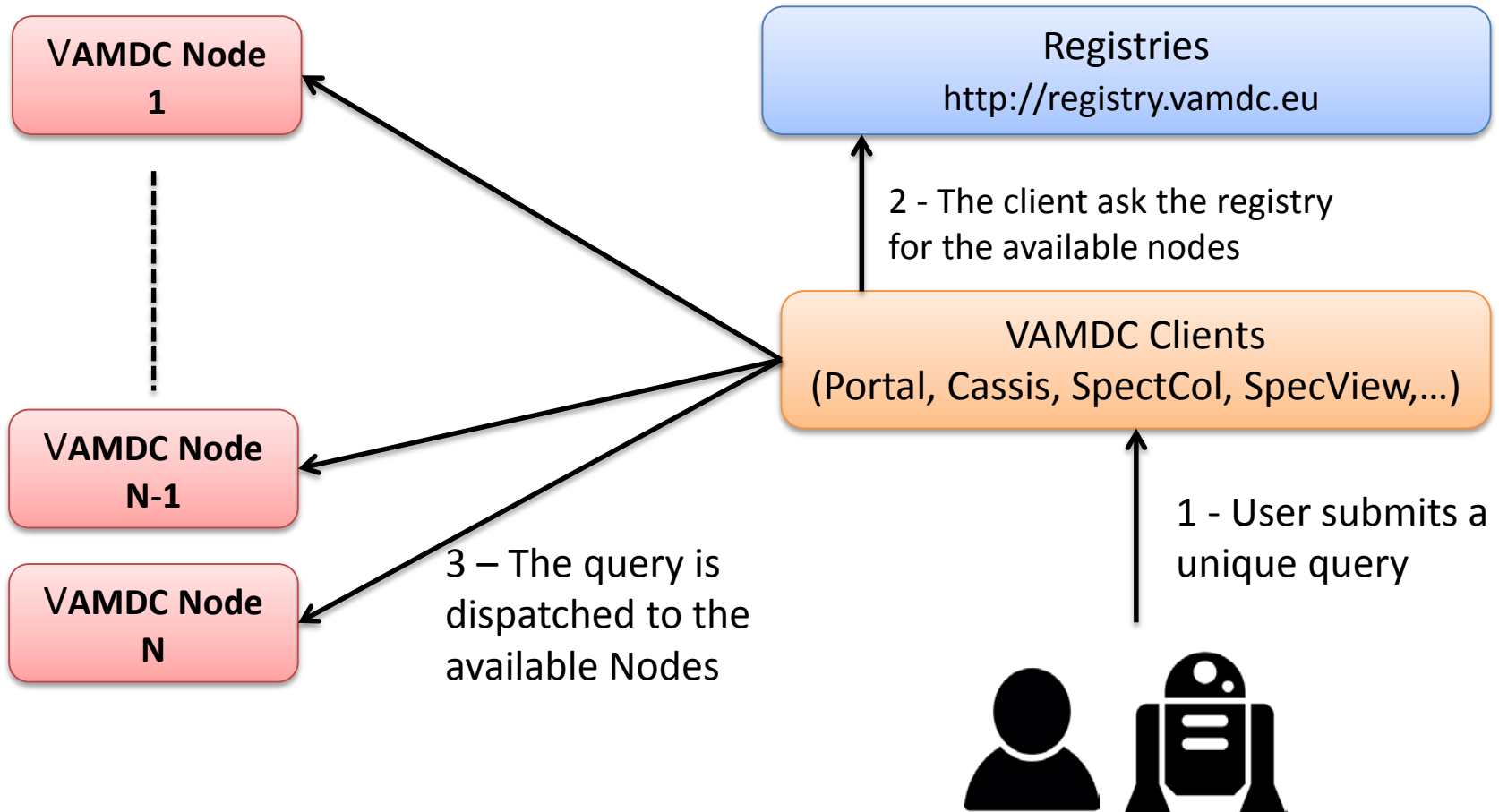
The VAMDC infrastructure technical architecture



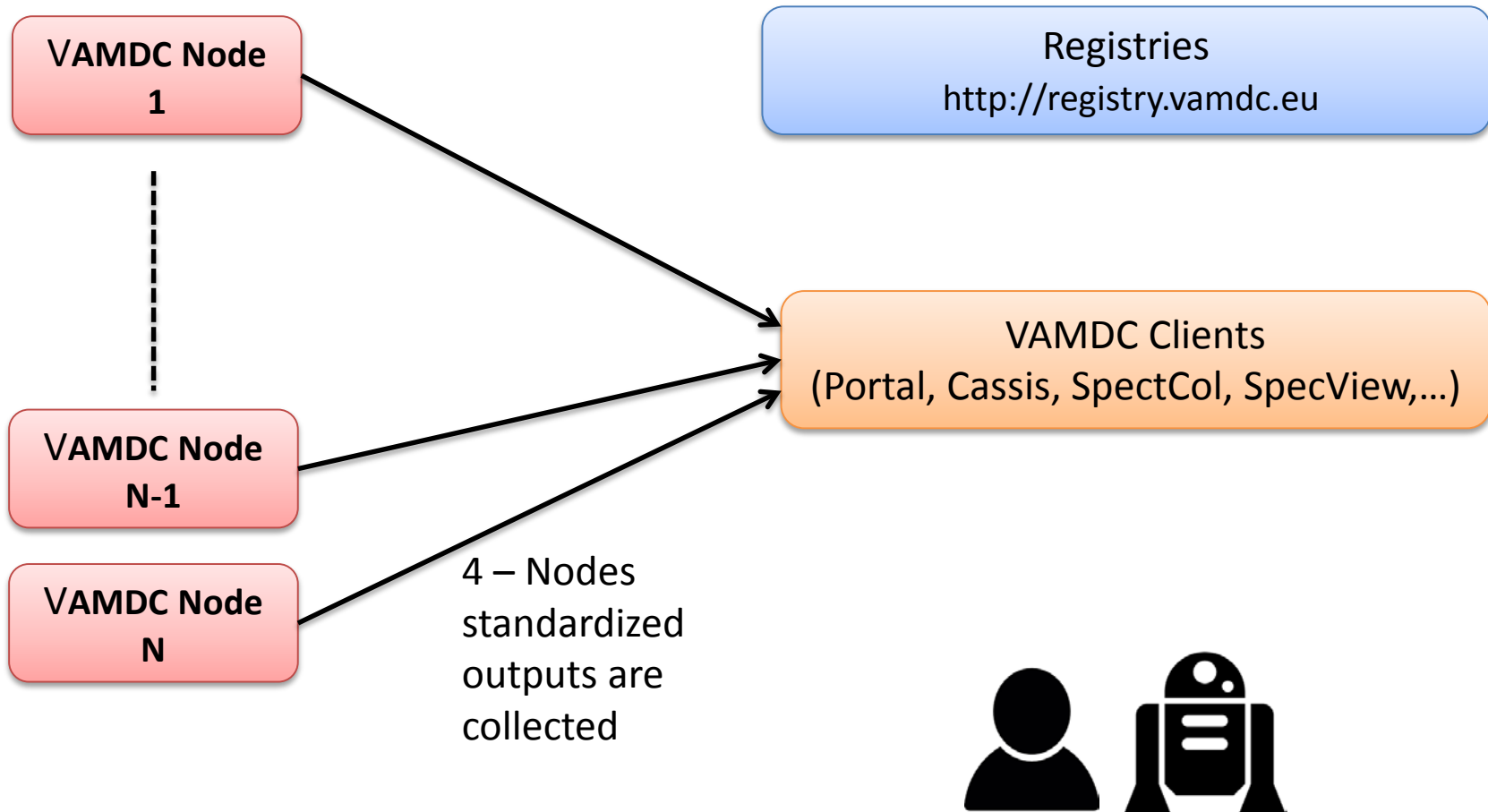
The VAMDC infrastructure technical architecture



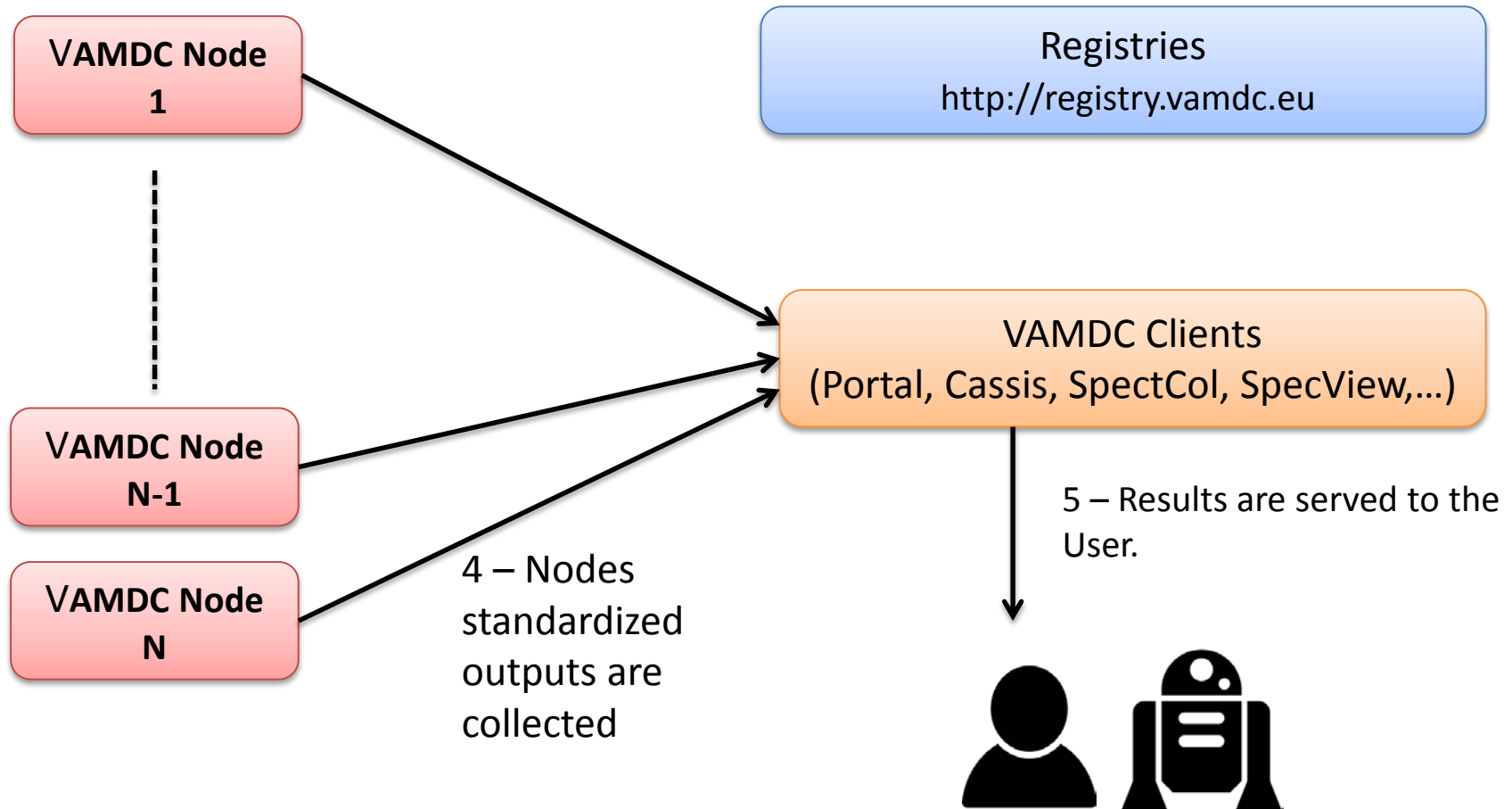
The VAMDC infrastructure technical architecture



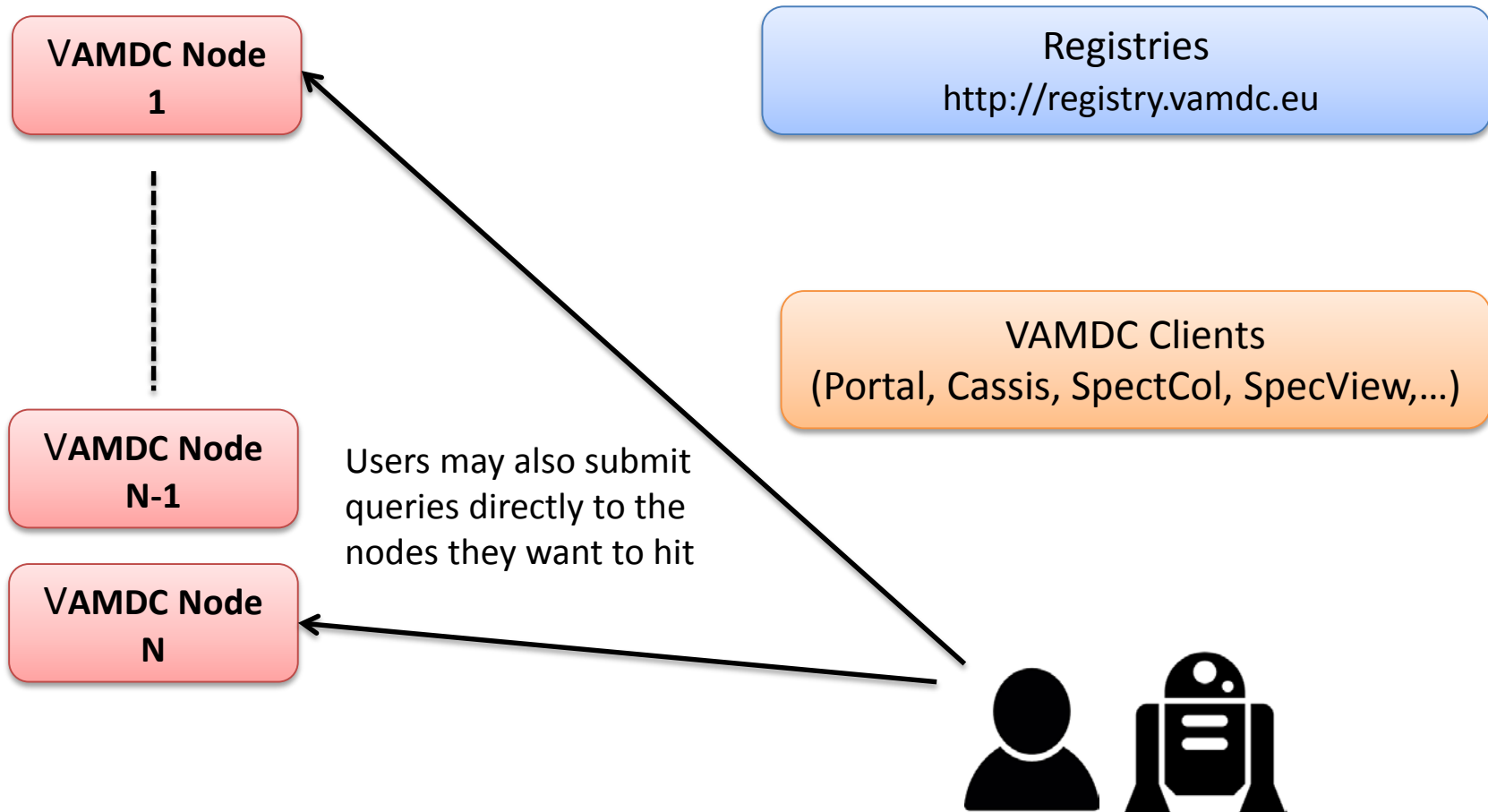
The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



The VAMDC infrastructure technical architecture



The Research Data Alliance and the Data Citation WG

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.

Goals of this WG are to create identification mechanisms that:

- allows us to identify and cite arbitrary views of data, from a single record to an entire data set in a precise, machine-actionable manner
- allows us to cite and retrieve that data as it existed at a certain point in time, whether the database is static or highly dynamic
- is stable across different technologies and technological changes

Solution: The WG recommends solving this challenge by:

- ensuring that data is stored in a versioned and timestamped manner.
- identifying data sets by storing and assigning persistent identifiers (PIDs) to timestamped queries that can be re-executed against the timestamped data store.

- The RDA recommendations come from standalone databases or warehouse.
- VAMDC is a distributed infrastructure, with no central management system.

Highlighting the main issues

How to build a Query Store in our distributed infrastructure?

- The solution belongs to a space with lot of constraints
 - Any choice will impact each of the ~30 databases federated by VAMDC.
 - Any technological change of the infrastructure must be validated by the majority of the members

The solution must cause least effects on the existing infrastructure and have minimal implementing cost for the database owners.

- This constraint suggest to fit the solution into the standard wrapping layer transforming an autonomous Database into a VAMDC node.

Sketching the solution strategy

Implementation is an overlay to the standard / output layer, thus independent from any specific data-node

Tagging versions of data

Two layers
mechanisms

1 → Fine grained granularity:
Evolution of XSAMS output
standard for tracking data
modifications*

2 → Coarse grained granularity:
At each data modification to a
given data node, the version of
the Data-Node changes

With the **second mechanism** we know that something changed : in other words, we know that the result of an identical query may be different from one version to the other. The detail of which data changed is accessible using the **first mechanisms**.

Query Store

Is built over the versioning of Data
(the coarse-grained mechanism)

Is plugged over the existing VAMDC
data-extraction mechanisms.

**Due to the distributed VAMDC
architecture, the Query Store may be
seen as a smart log-service.**

* <http://dx.doi.org/10.1016/j.jms.2016.04.009>
arxiv version at <https://arxiv.org/abs/1606.00405>

Let us focus on the query store:

The difficulties we had to cope with:

- Handle a query store in a distributed environment (RDA did not design it for these configurations).
- Integrate the query store with the existing VAMDC infrastructure.

The implementation of the query store is the goal of a joint collaboration between VAMDC and RDA-Europe 3.

- Development started during spring 2016.
- Final product released during 2017.

Collaboration with Elsevier for embedding the VAMDC query store into the pages displaying the digital version of papers.

Designing technical solution for

- Paper / data linking at the paper submission (for authors)
- Paper / data linking at the paper display (for readers)

Further Information

Technical details for data versioning:

New model for datasets citation and extraction reproducibility in VAMDC,

C.M. Zwölf, N. Moreau, M.-L. Dubernet,

In press *J. Mol. Spectrosc.* (2016), <http://dx.doi.org/10.1016/j.jms.2016.04.009>

Arxiv version: <https://arxiv.org/abs/1606.00405>

This mechanism should be included in the follow 2017 version of the standards.

Details about the implemented Query Store:

Video from the last week webinar: <https://youtu.be/OLe-qcqCcCw>

- Information
 - On the software architecture adopted for implementing the recommendation
 - On the implementing tricks

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.

Minimal impact
for federated
database owners
for dealing with
the Query Store

Database owners just need to install the latest
version of the VAMDC wrapping software

Data providers has to fill a “version” field (~ a simple string),
which is just the version label.

- When the database is modified and/or data node software changes, the version label should evolve.

Remarks before the live demo:

- We provided the VAMDC infrastructure with a working Query Store
- The concept adopted and the implemented code are quite generic and both can be adapted to other use-cases:
 - If it worked in our complex distributed case, it may work in many contexts
 - The cost for adapting an existing service/database to a VAMDC-type Query Store is minimal
 - All the complexity is handled and masked in our generic software.

Contact : carlo-maria.zwolf AT obspm.fr

Live demo link: <https://youtu.be/kDDWFpi22cU>



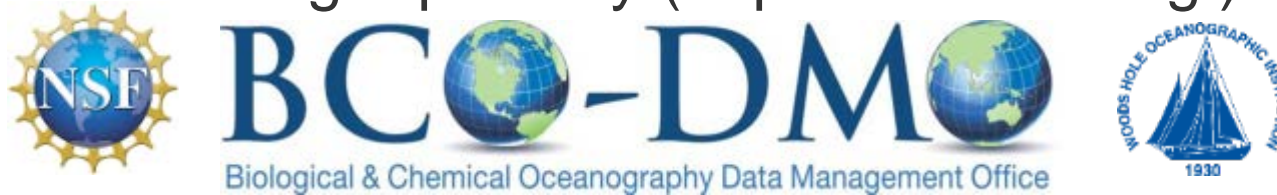
Adoption of Data Citation Outcomes by BCO-DMO

**Cynthia Chandler, Adam Shepherd, David Bassendine
Biological and Chemical Oceanography Data Management Office
Woods Hole Oceanographic Institution and Blue Dot Labs**

research data sharing without barriers
rd-alliance.org

A story of success enabled by RDA

- An existing repository (<http://bco-dmo.org/>)





- Marine research data curation since 2006
- Faced with new challenges, but no new funding
- e.g. data publication practices to support citation
- Used the outcomes from the RDA Data Citation Working Group to improve data publication and citation services

<https://www.rd-alliance.org/groups/data-citation-wg.html>


- BCO-DMO is a thematic, domain-specific repository funded by NSF Ocean Sciences and Polar Programs
- BCO-DMO curated data are
 - Served: <http://bco-dmo.org> (URLs, URIs)
 - Published: at an Institutional Repository (CrossRef DOI) <http://dx.doi.org/10.1575/1912/4847>
 - Archived: at NCEI, a US National Data Center <http://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:0078575>

Archival Copy

for Linked Data URI: <http://lod.bco-dmo.org/id/dataset/3046>

Version Date	Archive	Persistent Identifier	Date Assigned
2007-05-24	National Oceanographic Data Center (NODC)	0078575 	2011-10-26
2007-05-24	Marine Biological Laboratory/Woods Hole Oceanographic Institution Library (MBLWHOI DLA)	10.1575/1912/4847 	2011-10-12

BCO-DMO Dataset Landing Page (Mar '16) ⁵⁰



Biological & Chemical Oceanography Data Management Office

[HOME](#)[DATA](#)[RESOURCES](#)[ABOUT US](#)

DATABASE


Programs	36
Projects	649
Deployments	2300
Datasets	7892
Instruments	395
Parameters	1366
People	2026
Affiliations	472
Funding	79
Awards	1304

Dataset: larval krill pigments

[Get Data](#)[Map It](#)

Project: U.S. GLOBEC Southern Ocean (SOGLOBEC)
Principal Investigator: Dr Robin Ross (University of California-Santa Barbara, UCSB) Dr Langdon Quetin (University of California-Santa Barbara, UCSB)
Contact: Dr Robin Ross (University of California-Santa Barbara, UCSB)
BCO-DMO Data Manager: Nancy Copley (Woods Hole Oceanographic Institution, WHOI BCO-DMO)
Validated: Yes
Data version: 2010-02-03
Version Date: 02/03/2010
Data URL: http://www.bco-dmo.org/dataset/3300/data ↗
Current State: Final no updates expected

GEOSPATIAL ACCESS



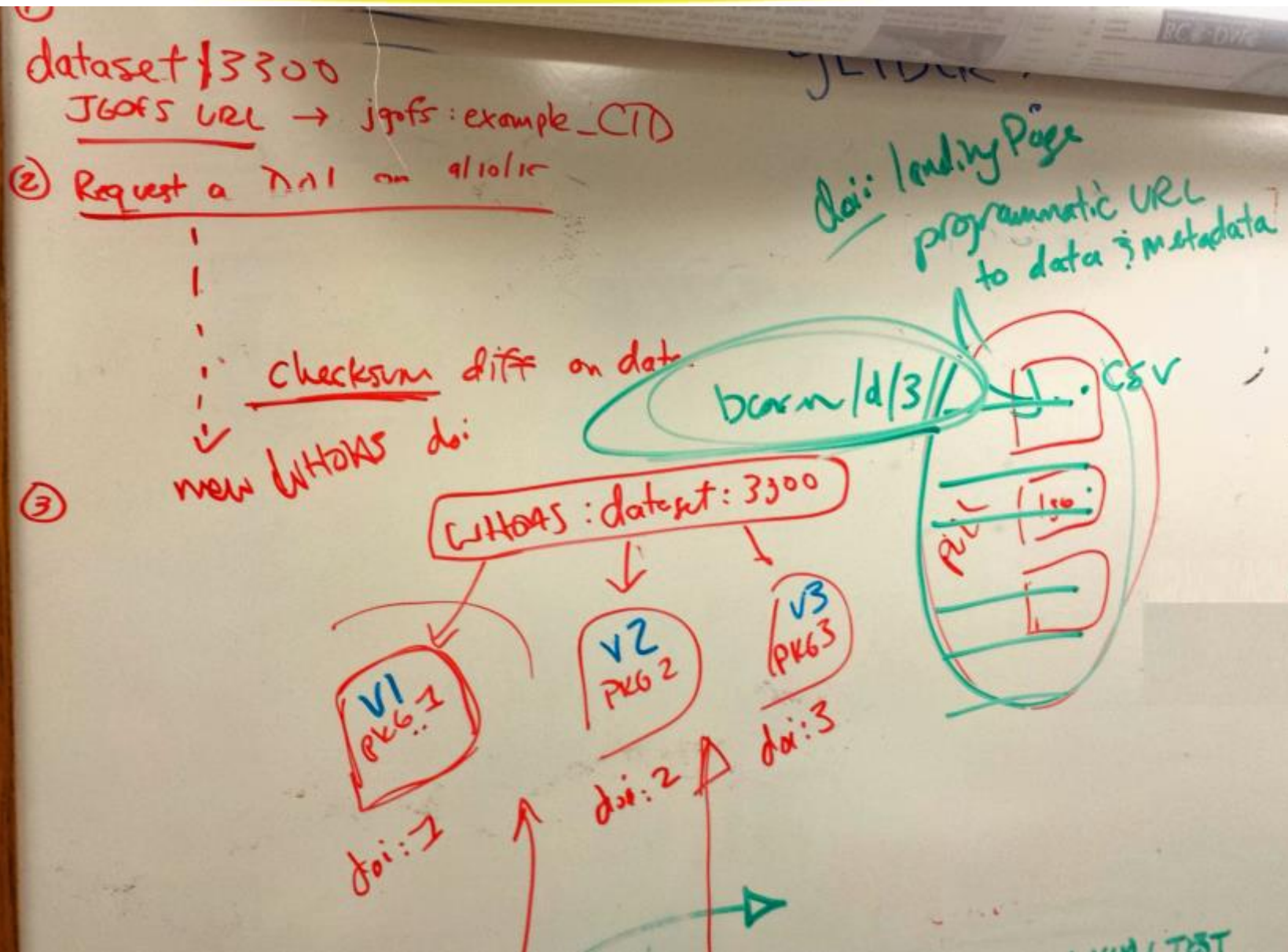
[Expand/Collapse All](#)

[Archival Copy](#)

Archive Location	Accession ID	Archive Copy URL
National Oceanographic Data Center (NODC)	0112635	http://accession.nodc.noaa.gov/0112635

Initial Architecture Design Considerations (Jan 2016)

51



Modified Architecture (March 2016)

52

BCO-DMO DATA PUBLICATION & CITATION

28 MAR 2016

Archive Record: WHOAS [mid = 567890]

DFD 90001 [ISO] downloaded: DOI
doi: 10.115/1111/bco-dmo.567890

DFD 90002 [DATA 1]

DFD 90003 [DATA N]

WHOAS

[ISO]

[1]

[N]

QUESTION:

supporting docs
are recorded as DFD
at WHOAS for
reuse w/ NCET

DATA FILE DESCRIPTOR [DFD]

OSPREY CONTENT TYPE

checksum: text
checksum algorithm: controlled vocabulary
filesize_in_bytes: integer
resolvable_url: URL
dataset: Dataset or Dataset Deployment
type: controlled vocabulary [Data File Type]

DATA File Type

name: text (ex: Tab Separated Values, ISO 11175-2)
data_one_identifier: text
data_one_type: controlled vocabulary
mime_type: text (text/csv, application/xml)

Archive Record: NCET [mid = 678901]

DFD 90001
DFD 90002
DFD 90003

Serialized from WHOAS,
Stored in WAF (HTTPS)
SHA 384 Manifest & Summary

Archive Record: DATAONE [mid = 789012]

DFD 91000 [ISO] downloaded: DOI
PROVISION # at NCET
DOI: 10.115/1111/bco-dmo.567890

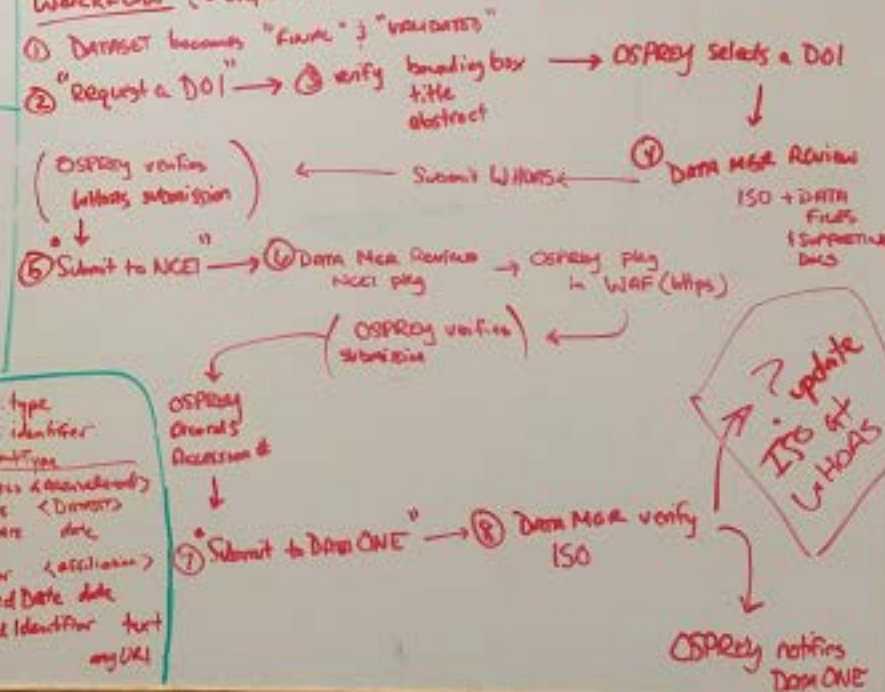
DFD 90002

DFD 90003

(stored locally
in OSPREY)
or update
versions?

(Serialized from
WHOAS)

WORKFLOW (8 steps for new DATASET VERSION)



<DATASET> archived By <Archive Record>

<Archive Record> has Bitstream
<Data File Descriptor>
checksum
checksum algorithm
fileInBitstream
downloadURL
serializes
type
isAbsolute

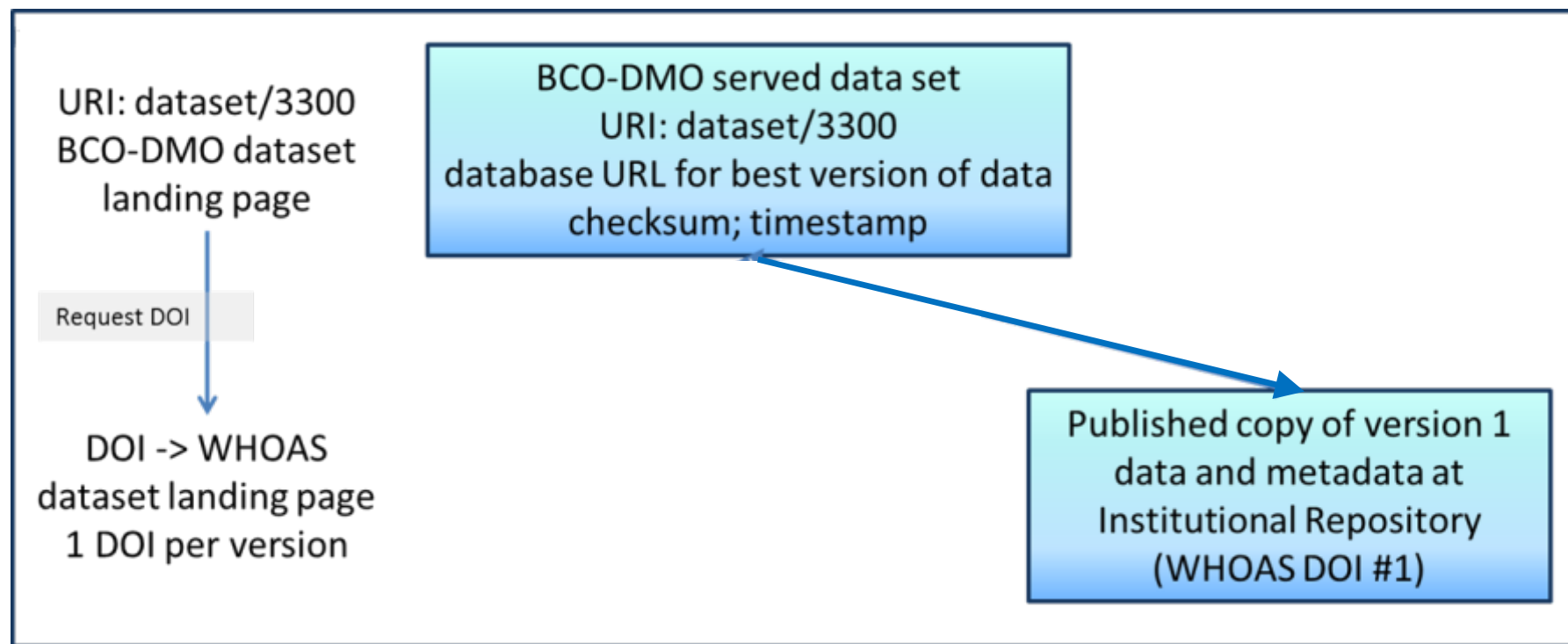
<Data File Descriptor>
checksum
checksum algorithm
fileInBitstream
downloadURL
serializes
type
isAbsolute

<Dataset>
dataset identifier
dataset type
dataset version
dataset date
dataset title
dataset order
dataset archived date
dataset archived identifier
dataset url

<Data File Descriptor>
data one identifier
data one type
mime type

BCO-DMO Data Publication System Components

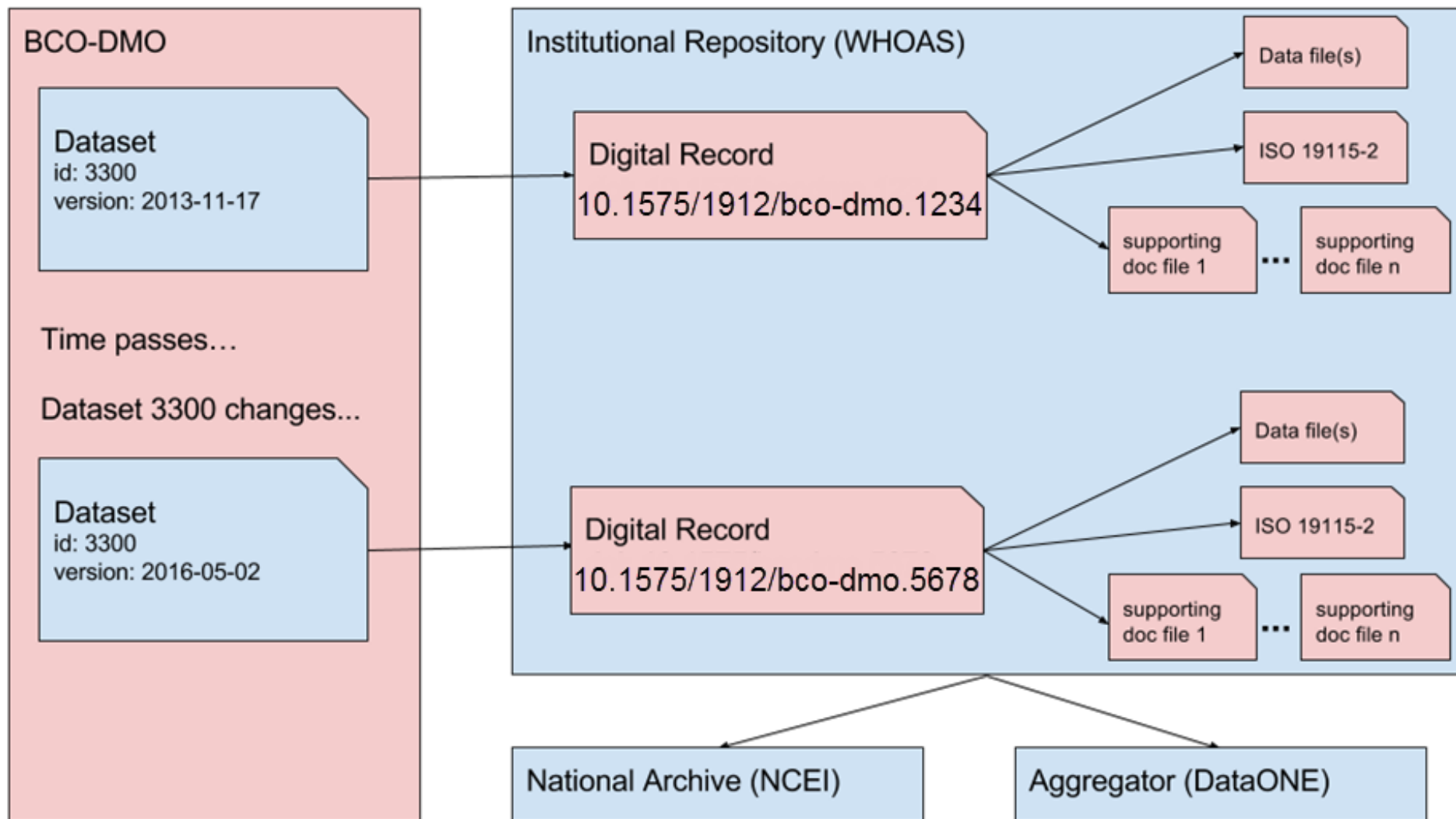
53



BCO-DMO publishes data to WHOAS and a DOI is assigned.
The BCO-DMO architecture now supports data versioning.


BCO-DMO Data Citation System Components

54



BCO-DMO Data Set Landing Page

55



Biological & Chemical Oceanography Data Management Office

[DATA](#) [RESOURCES](#) [ABOUT US](#)

DATABASE

- New Entry
- Programs 36
- Projects 696
- Deployments 2360
- Platforms 486
- Datasets 8079
- Instruments 408

Dataset: Cellular element quotas: Si in Synechococcus cells

[Get Data](#) [Map It](#)

Project: [Understanding the Role of Picocyanobacteria in the Marine Silicate Cycle \(Si_in_Syn\)](#)

Principal Investigator: [Dr Benjamin Twining](#) (Bigelow Laboratory for Ocean Sciences, Bigelow)

Contact: [Daniel Ohrenius](#) (Bigelow Laboratory for Ocean Sciences, Bigelow)

Data version: Final

Version Date: 05/06/2016

▼ [Archival Copy](#)

▼ Archival Copy

Version Date	Archive	Persistent Identifier	Date Assigned
2016-05-06	Marine Biological Laboratory/Woods Hole Oceanographic Institution Library (MBLWHOI DLA)	10.1575/1912/bco-dmo.651474 	2016-07-08

GEOSPATIAL ACCESS



References:

Twining, B.S., et al. Metal contents of phytoplankton and labile particulate material in the North Atlantic Ocean. *Prog. Oceanogr.* (2015)
<http://dx.doi.org/10.1016/j.pocean.2015.07.001>
https://www.researchgate.net/publication/282626294_Metal_contents_of_phytoplankton_and_labile_particulate_material_in_the_North_Atlantic_Ocean

▼ [More Information about this dataset](#)

Search



- ☒ Search WHOAS
☐ This Collection

BROWSE

All of WHOAS

Communities & Collections

Element quotas of individual *Synechococcus* cells collected during Bermuda Atlantic Time-series Study (BATS) cruises aboard the R/V Atlantic Explorer between dates 2012-07-11 and 2013-10-13

dc.description.sponsorship NSF Ocean Sciences (NSF OCE) OCE-1131139, NSF Ocean Sciences (NSF OCE) OCE-1335012, NSF Ocean Sciences (NSF OCE) OCE-1131046

dc.publisher Biological and Chemical Oceanography Data Management Office (BCO-DMO). Contact: bco-dmo-data@whoi.edu




dc.relation <http://lod.bco-dmo.org/id/dataset/644840>

By Issue Date

Authors

Titles

View Item

-  cellular-element-quotas-si-synechococcus-cells.tsv (10.09Kb)
 115-2.xml (88.65Kb)
 Names.pdf (12.10Kb)

31.6691


DOI
[10.1575/1912/bco-dmo.651474](https://doi.org/10.1575/1912/bco-dmo.651474)

Keyword

SXRF; Synchrotron radiation X-Ray Fluorescence; *Synechococcus*; si

BCO-DMO Data Set Landing Page

57



Biological & Chemical Oceanography Data Management Office

[DATA](#) [RESOURCES](#) [ABOUT US](#)

DATABASE

New Entry	
Programs	36
Projects	696
Deployments	2360
Platforms	486
Datasets	8079
Instruments	408
Parameters	1373

Dataset: Cellular element quotas: Si in Synechococcus cells

[Get Data](#) [Map It](#)

Project: Understanding the Role of Picocyanobacteria in the Marine Silicate Cycle (Si_in_Syn)
Principal Investigator: Dr Benjamin Twining (Bigelow Laboratory for Ocean Sciences, Bigelow)
Contact: Daniel Ohnemus (Bigelow Laboratory for Ocean Sciences, Bigelow)
Data version: Final
Version Date: 05/06/2016

▼ [Archival Copy](#)

Version	Archive	Persistent Identifier	Date

References:

Twining, B.S., et al. Metal contents of phytoplankton and labile particulate material in the North Atlantic Ocean. Prog. Oceanogr. (2015)

<http://dx.doi.org/10.1016/j.pocean.2015.07.001>

<https://www.researchgate.net/publication>

/282626294_Metal_contents_of_phytoplankton_and_labile_particulate_material_in_the_North_Atlantic_Ocean

are also provided.

GEOSPATIAL ACCESS



References:

Twining, B.S., et al. Metal contents of phytoplankton and labile particulate material in the North Atlantic Ocean. Prog. Oceanogr. (2015)

<http://dx.doi.org/10.1016/j.pocean.2015.07.001>





<https://www.researchgate.net/publication>


/282626294_Metal_contents_of_phytoplankton_and_labile_particulate_material_in_the_North_Atlantic_Ocean

▼ [More information about this dataset](#)


ScienceDirect

JournalsBooksSign in ?

 Download PDF  Add to online library  Export  Advanced search

**ELSEVIER**

Progress in Oceanography
Volume 137, Part A, September 2015, Pages 261–283



Metal contents of phytoplankton and labile particulate material in the North Atlantic Ocean

Benjamin S. Twining^a, Sara Rauschenberg^a, Peter L. Morton^{b, 1}, Stefan Vogt^c

[Show more](#)

[doi:10.1016/j.pocean.2015.07.001](https://doi.org/10.1016/j.pocean.2015.07.001) [Get rights and content](#)

Highlights

- First basin-wide measurements of plankton metal quotas in the N. Atlantic Ocean.
- Fe and Mn quotas significantly higher on western side of section.
- Cu and Ni quotas significantly elevated on eastern side of section.
- Evidence for Al scavenging by biogenic silica.
- Dissolved ratios not an accurate measure of cellular Fe quotas.

Recommended articles

[Evaluation of approaches to estimate biogenic ...](#)
2015, Marine Chemistry [more](#)

[High-frequency sea level oscillations in the Me...](#)
2015, Progress in Oceanography [more](#)

[Composition of metals in suspended particulate...](#)
2016, Regional Studies in Marine Science [more](#)

[View more articles »](#)

Citing articles (2)

Related book content

New Capabilities ... BCO-DMO becoming a DataONE Member Node

59

<https://search.dataone.org/>

The screenshot displays the DataONE search interface. At the top, the DataONE logo is on the left, and navigation links (About, News, Participate, Resources, Education, Data) are in the center. A search bar with the text "DATAONE SEARCH:" and buttons for "Search", "Summary", "Jump to:", "DOI or ID", and "Go" is on the right. A "Sign in" or "Sign up" button is also present. Below the navigation bar, a "Clear all filters" link is visible. The main content area is divided into three sections: a left sidebar, a central results list, and a right map panel. The sidebar includes a "Search" section with a search phrase input and a "My Search" section with a "BCO-DMO" filter. Below these are "Filter by:" options: Data attribute, Data files, Member Node, Creator, Year, Identifier, and Taxon. The central results list shows "Datasets 1 to 25 of 149" and lists three datasets by Dr. David Siegel and Dr. Kenneth O. Buesseler (2009) and Dr. Carl Lamborg and Dr. Kenneth O. Buesseler (2009), and Dr. Uta Passow (2014). Each entry includes a DOI and a link to the dataset file. The right map panel shows a map of the United States with several regions highlighted in blue and numbered (2, 4, 19, 1, 25, 1). The map includes a "Hide Map" button, a "Satellite" button, a "Terrain" button, and a "Google" logo. At the bottom, it says "Map data ©2016 Google, INEGI" and "1000 km" scale bar.

DataONE

About News Participate Resources Education Data

DATAONE SEARCH: Search Summary Jump to: DOI or ID Go Sign in or Sign up

✕ Clear all filters

Search ?

Search phrase

My Search

BCO-DMO ✕

Filter by:

- ▶ Data attribute
- ▶ Data files
- ▶ Member Node
- ▶ Creator
- ▶ Year
- ▶ Identifier
- ▶ Taxon

Datasets 1 to 25 of 149

1 2 3 ... 6 Next Sort by Most recent

Dr David Siegel and Dr Kenneth O. Buesseler. 2009. urn:nod:mTestBCODMO. <http://lod.bco-dmo.org/id/dataset-file/555907>.

Dr Carl Lamborg and Dr Kenneth O. Buesseler. 2009. urn:nod:mTestBCODMO. <http://lod.bco-dmo.org/id/dataset-file/555902>.

Dr Uta Passow. 2014. urn:nod:mTestBCODMO. <http://lod.bco-dmo.org/id/dataset-file/555897>.

DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S.

Hide Map »

2 4 19 1 25 1


Satellite Terrain

Google

Map data ©2016 Google, INEGI 1000 km Terms of Use

New Capabilities ... BCO-DMO Data Set Citation

60


Biological & Chemical Oceanography Data Management Office

DATARESOURCESABOUT US

Enter search terms

DATABASE

Programs37

Projects724

Dataset: Prochlorococcus_narB

Get DataMap It

Cite This Dataset

Data Citation:

Chisholm, Sallie (2016) The abundance of Prochlorococcus cells containing the nitrate reductase gene (narB) at the HOT and BATS sites in the Pacific and Atlantic Oceans between October 2005 and January 2008. Biological and Chemical Oceanography Data Management Office (BCO-DMO) Dataset version 2016-03-31 [if applicable, indicate subset used]. doi:10.1575/1912/bco-dmo.641735 [access date]

Terms of Use

All data sets are licensed under a [Creative Commons Attribution 4.0 International License](#) (CC BY 4). Per the CC BY 4 license it is understood that any use of the data set will properly acknowledge the individual(s) listed above using the suggested data citation. If you wish to use this data set, it is highly recommended that you contact the original principal investigator(s) (PI). Should the relevant PI be unavailable, please contact BCO-DMO (info@bco-dmo.org) for additional guidance. For general guidance please see the BCO-DMO [Terms of Use](#) document.

Awards1412

Current State: Final no updates expected

Archival Copy

GEOSPATIAL ACCESS



Version Date	Archive	Persistent Identifier	Date Assigned
2016-03-31	Marine Biological Laboratory/Woods Hole Oceanographic Institution Library (MBLWHOI DLA)	10.1575/1912/bco-dmo.641735	2016-04-04

Thank you ...

- To the Data Citation Working Group for their efforts
<https://www.rd-alliance.org/groups/data-citation-wg.html>
- RDA US and MacArthur Foundation for funding this adoption project
- TIMELINE:
 - Redesign/protoype completed by 1 June 2016
 - New citation recommendation by 1 Sep 2016
 - Report out at RDA P8 (Denver, CO) September 2016
 - Final report by 1 December 2016

Cyndy Chandler @cynDC42

ORCID: 0000-0003-2129-1647

@bcodmo

cchandler@whoi.edu

- Removed these to reduce talk to 10-15 minutes

■ Evaluation

- Evaluate recommendations (done December 2015)
- Try implementation in existing BCO-DMO architecture (work began 4 April 2016)

■ Trial

- BCO-DMO: R1-11 fit well with current architecture; R12 doable; test as part of DataONE node membership; R13-14 are consistent with Linked Data approach to data publication and sharing

NOTE: adoption grant received from RDA US (April 2016)

- **DC goals:** to create identification mechanisms that:
 - allow us to identify and cite arbitrary views of data, from a single record to an entire data set in a precise, machine-actionable manner
 - allow us to cite and retrieve that data as it existed at a certain point in time, whether the database is static or highly dynamic
- **DC outcomes:** 14 recommendations and associated documentation
 - ensuring that data are stored in a versioned and timestamped manner
 - identifying data sets by storing and assigning persistent identifiers (PIDs) to timestamped queries that can be re-executed against the timestamped data store

<https://www.rd-alliance.org/groups/data-citation-wg.html>

- Preparing the Data and the Query Store
 - R1 - Data Versioning
 - R2 - Timestamping
 - R3 - Query Store Facilities
- Persistently Identifying Specific Data Sets
 - R4 - Query Uniqueness
 - R5 - Stable Sorting
 - R6 - Result Set Verification
 - R7 - Query Timestamping
 - R8 - Query PID
 - R9 - Store the Query
 - R10 - Automated Citation Texts
- Resolving PIDs and Retrieving the Data
 - R11 - Landing Page
 - R12 - Machine Actionability
- Upon modifications to the Data Infrastructure
 - R13 - Technology Migration
 - R14 - Migration Verification

»» Data Versioning: For retrieving earlier states of datasets the data need to be versioned. Markers shall indicate inserts, updates and deletes of data in the database.

»» Data Timestamping: Ensure that operations on data are timestamped, i.e. any additions, deletions are marked with a timestamp.

»» Data Identification: The data used shall be identified via a PID pointing to a time-stamped query, resolving to a landing page.

Oct 2015 version w/ 14 recommendations

DC WG chairs: Andreas Rauber, Ari Asmi, Dieter van Uytvanck

New capability (implemented)

procedure: when a BCO-DMO data set is updated ...

- A copy of the previous version is preserved
- Request a DOI for the new version of data
- Publish data, and create new landing page for new version of data, with new DOI assigned
- BCO-DMO database has links to all versions of the data (archived and published)
- Both archive and published dataset landing pages have links back to best version of full dataset at BCO-DMO
- BCO-DMO data set landing page displays links to all archived and published versions

- Extended description of recommendations
**Identification of Reproducible Subsets for Data Citation,
Sharing and Re-Use**

Andreas Rauber
Vienna University of Technology, Austria
rauber@ifs.tuwien.ac.at

Dieter van Uytvanck
CLARIN ERIC, Utrecht, Netherlands
dieter@clarin.eu

Ari Asmi
University of Helsinki, Finland
ari.asmi@helsinki.fi

Stefan Pröll
SBA Research, Vienna, Austria
sproell@sba-research.org

- Altman and Crosas. 2013. “Evolution of Data Citation ...”
- CODATA-ICSTI 2013. “Out of cite, out of mind”
- FORCE11 <https://www.force11.org/about/mission-and-guiding-principles>
- R. E. Duerr, et al. “On the utility of identification schemes for digital earth science data”, ESI, 2011.



Citing Dynamic Datasets at NICT

Yasuhiro Murayama
murayama@nict.go.jp

research data sharing without barriers
rd-alliance.org

Citing dynamic datasets for Sensing Big Data at NICT

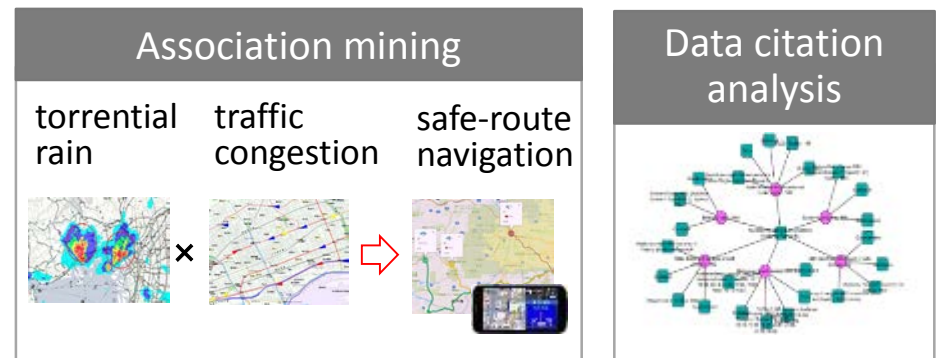
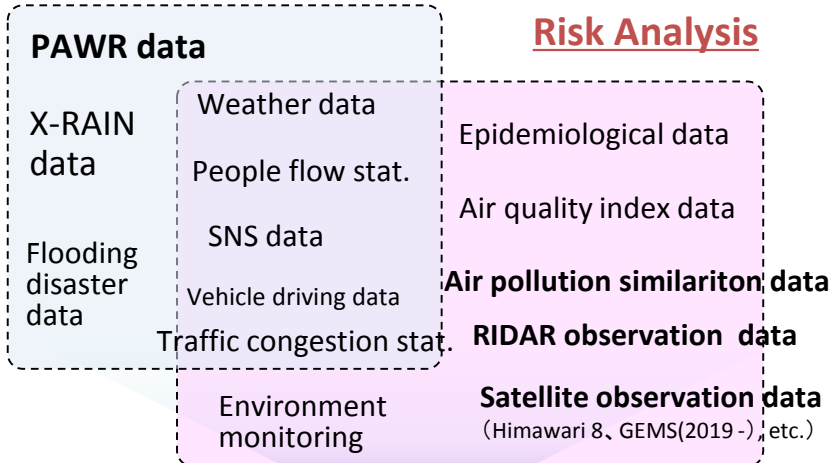
Yasuhiro Murayama & Koji Zettsu

National Institute of Information & Communications Technology, Japan

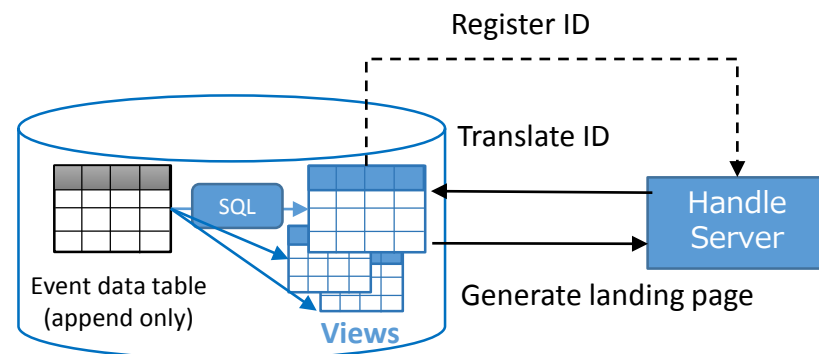
- Superior discoverability of individually-disseminated sensing data for natural disaster events
- Analysis of data reputation for incentive reward in participatory sensing
- Data provenance for quality assurance of data mining results

Torrential Rain Risk Analysis

Air Pollution Risk Analysis

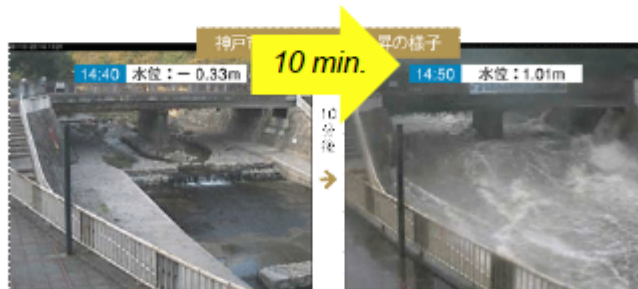


Citable mechanism and experiment



Localized Torrential Rain Disaster

- < 100 mm/h heavy rain within 10km² area in short time (< 1 hour).
- Unpredictable. Urban heat island as a major cause
- 200+ times per year in Japan. 180 billion JPY damage for past 10 years in Tokyo



Flash flooding (Kobe, 2008)



Flood in subway (Fukuoka, 1999)



Landslide (Hiroshima, 2014)

Example of Meteorological Disaster Event List

(Japan Meteorological Agency)

- 2014-2015年度の豪雨災害事例(関西)

神戸地方気象台, <http://www.jma-net.go.jp/kobe-c/knowledge/saigai/saigai.html>

- 平成26年 8月8日～10日 台風第11号による大雨と暴風
- 平成26年 8月15日～18日 兵庫県の大雨
- 平成26年 8月24日～8月25日 兵庫県の大雨
- 平成26年 9月10日～9月11日 兵庫県の大雨
- 平成26年 10月12日～14日 台風第19号による大雨と暴風について
- 平成27年 7月16日～18日 台風第11号による大雨と暴風

- これらの事例に該当する以下のデータにdDOIを設定

- フェーズドアレイ気象レーダ(PANDA)降水量データ
- XRAIN降水量データ

- 事例報告書等からメタデータを作成

Weather Radars

**Metadata from past Gov.
Reports**

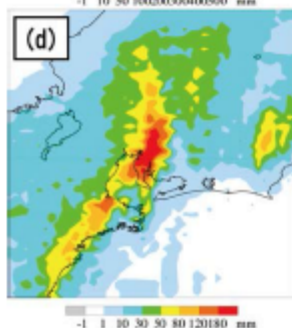
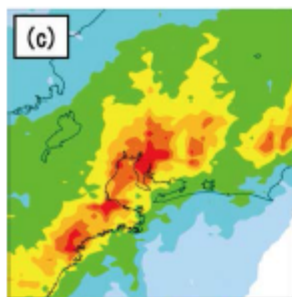
Query example of Torrential Rainfall Event

- 以下のフローに沿って、PANDAデータ及びXRAINデータから抽出
 - 抽出条件に基づいてメタデータを生成

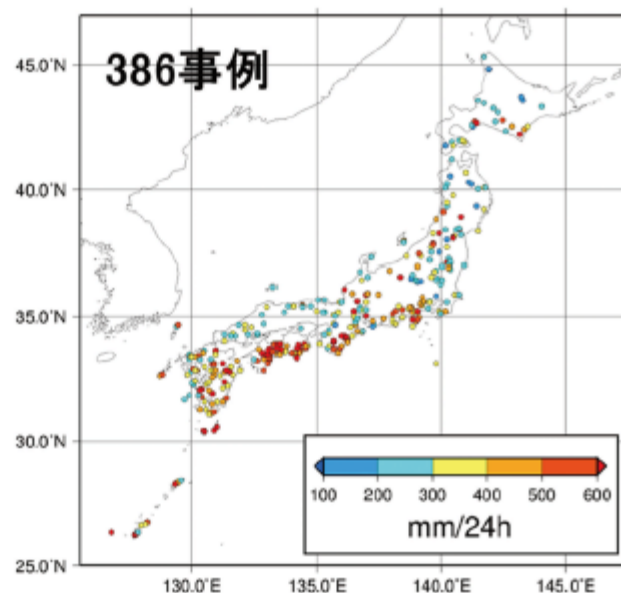


抽出された集中豪雨事例の例

上段: 24時間積算降水量、
下段: 前3時間積算降水量



1995～2009年4月～11月の期間
で抽出した例



第5図 抽出された集中豪雨事例の分布。各点は、各集中豪雨事例における最大前24時間積算降水量の点を示す。陰影は前24時間積算降水量を表す。

2017/3/14

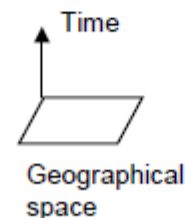
出典: 津口裕茂, 加藤輝之: 集中豪雨事例の客観的な抽出とその特性・特徴に関する統計解析, 日本気象学会, 天気61. 6, pp.455-469 (2014年6月)

Discovery of spatial, temporal, thematic associations from heterogeneous sensing data


Torrential rain disaster in Tokyo (Sep. 2014)



- **Flood damage events:** heavy rain & flood-related tweets
- **Traffic damage events:** heavy rain & traffic-related tweets
- **Refuge events:** heavy rain & tweets at shops, restaurants



 Weather sensor

 Precipitation sensor

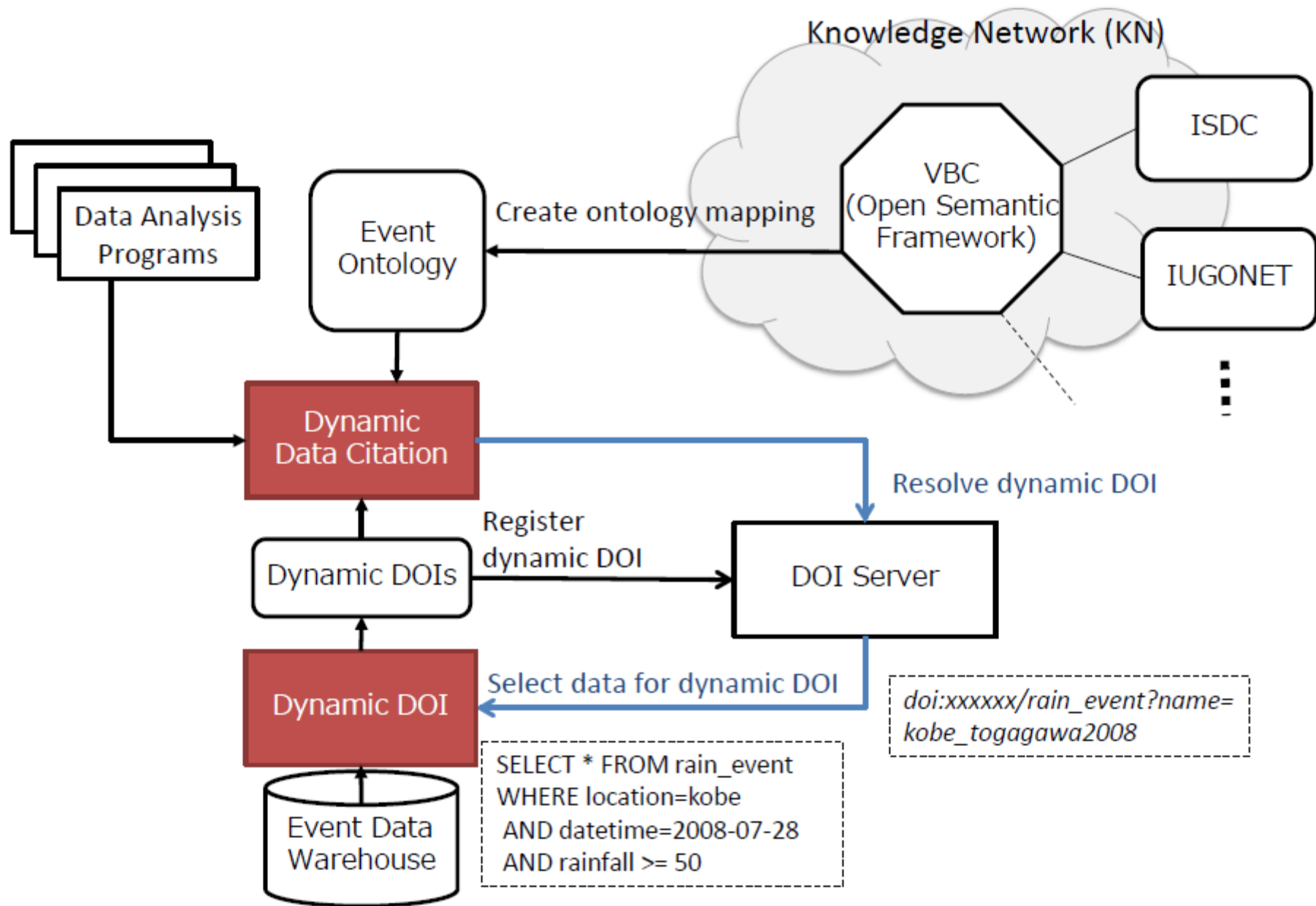
 SNS

 Traffic information

Sensing Data Statistics (Part)

Sensing Data	Type	Records	Bytes	Period
PAWR precipitation data (2km grid)	Live	19.4 billion 16.8 million/day	8.7TB (6.9GB/day)	2013.5 - current
PAWR precipitation data (250m grid)	Live	(2,800/day)	(3.5GB/day)	(preparing)
XRAIN precipitation data (250m grid)	Archive	(14,400/day)	(5.9GB/day)	(preparing)
Soratena weather station data (3,000 stations nationwide)	Live	1.3 billion	479GB	2013.9 – current
AEROS atmospheric environmental observation data (Soramame-kun)	Live	36.4 million	37GB	2013.9 – current
Twitter SNS data (25 keywords) 2014	Archive	88 million	23GB	2014.4 - 2015.3
Twitter SNS data (geo-tagged) 2014	Archive	95 milion	25GB	(preparing)
Twitter SNS 2015	Archive			(preparing)

- PAWR precipitation data, http://pawr.nict.go.jp/index_en.html
- Soratena weather sensing data, <http://soratena.live-e.org/?file=top>
- AEROS (Soramame-kun), <http://soramame.taiki.go.jp/>



- 14:00 Introduction, Welcome
- 14:10 Short description of the WG recommendations
- 14:25 Report on Standardisation & Adoption
- 14:30 Brief reports on use cases
 - Virtual and Atomic Molecular Data Center (C.M. Zwolf)
 - EURO ARGO (TBA)
 - BCO-DMO (Slides from C. Chandler, presented by Asmi)
 - Pilot project in Japan (Yasuhiro Muyarama)
 - Others
- 15:10 Collaboration with other WGs and IGs
- 15:20 Other issues, future of the WG

Next Steps

- IG on Data Versioning (today, 16:00, MR1)
- Work with Domain IGs
- Support in adoption: what kind of support is needed?
(in the end it all boils down to money, but apart from this...)
 - Webinars: generic
 - Focused workshops for individual pilots
 - Joint projects: proposals, ...
- Dissemination of information from on-going pilots
 - Structuring: contact, descriptions, results, lessons learned
 - Outcomes: reports, slides, publications, code, discussions
 - Summary paper on pilots
- Anything else? AOB? Wishes?

Thanks

79

Thanks!
And hope to see you at the
next meeting
of the
WGDC