Ref. Ares(2017)2229067 - 30/04/2017

ENVRI^{plus} DELIVERABLE

D3.3
Report for best practices on robust telecom/data transmission

WORK PACKAGE 3 – Improving measurement networks: common technological solutions

LEADING BENEFICIARY: University of BREMEN

Author():s	Beneficiary/Institution
Robert Huber	UniHB
Olivier Gilbert	CNRS
Alex Vermeulen	LU
Dario Papale	UNITUS
Jean-Francois Rolin	Ifremer
Christoph Waldmann	UniHB
Susanne Rohs	FZJ

Accepted by: Executive Board

Deliverable type: [REPORT]

Dissemination level: PUBLIC

Deliverable due date: 31.04.2017/M24

Actual Date of Submission: 30.04.2017/M24

ABSTRACT

Near-Real-Time (NRT) data transmission from distant observatory or sensor locations to data processing centre is a common feature shared by almost all ENVRIplus Research Infrastructures (RI) and is essential for rapid quality checking for the benefit of RI operators. NRT data transmission and quality control procedures are left to independent choices of the RIs responding to their specific needs and availability of infrastructural resources. However, a common approach to those NRT technologies would be extremely useful to coordinate response of RIs in case of environmental crisis.

This document summarizes some key technologies used for data transmission, their availability, capabilities (in terms of bandwidth) and associated costs. It analyses requirements of research infrastructures and gives a summary of RIs best practises with respect to data transmission. Guided by the ENVRI Reference Model, further, potential harmonisation and standardisation approaches are discussed und special consideration of available sensor metadata standards and data transmission formats. In addition, common near real time quality routines and algorithms are presented.

We will show in this deliverable that data transmission is not an easy task and hardly can be harmonized - in technical terms - across all ENVRIplus research infrastructures due to the diverse latency and bandwidth requirements and telecommunication availability e.g. at distant sites. However, emerging standards have a good potential to move the standardisation of data transmission and will move the standardisation level closer to the sensor which would in turn enable cross domain data processing services such as the identified common NRT quality control routines.

Project internal reviewer(s):

Project internal reviewer(s):	**Beneficiary/Institution**
Ari Asmi	UHEL

Document history:

Date	**Version**
2017-04-03	First draft version
2017-04-05	Final draft submitted for internal review
2017-04-30	Accepted

TERMINOLOGY

A complete project glossary is provided online here:
https://envriplus.manageprojects.com/s/text-documents/LFCMXHHCwS5hh

PROJECT SUMMARY

ENVRIplus is a Horizon 2020 project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a more coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe. It is driven by three overarching goals: 1) promoting cross-fertilization between infrastructures, 2) implementing innovative concepts and devices across RIs, and 3) facilitating research and innovation in the field of environment for an increasing number of users outside the RIs.

ENVRIplus aligns its activities to a core strategic plan where sharing multi-disciplinary expertise will be most effective. The project aims to improve Earth observation monitoring systems and strategies, including actions to improve harmonization and innovation, and generate common solutions to many shared information technology and data related challenges. It also seeks to harmonize policies for access and provide strategies for knowledge transfer amongst RIs. ENVRIplus develops guidelines to enhance transdisciplinary use of data and data-products supported by applied use-cases involving RIs from different domains. The project coordinates actions to improve communication and cooperation, addressing Environmental RIs at all levels, from management to end-users, implementing RI-staff exchange programs, generating material for RI personnel, and proposing common strategic developments and actions for enhancing services to users and evaluating the socio-economic impacts.

ENVRIplus is expected to facilitate structuration and improve quality of services offered both within single RIs and at the pan-RI level. It promotes efficient and multi-disciplinary research offering new opportunities to users, new tools to RI managers and new communication strategies for environmental RI communities. The resulting solutions, services and other project outcomes are made available to all environmental RI initiatives, thus contributing to the development of a coherent European RI ecosystem.

TABLE OF CONTENT

BEST PRACTISES ON ROBUST DATA TRANSMISSION

INTRODUCTION

Near-Real-Time (NRT) data transmission from distant observatory or sensor locations to data processing centre is a common feature shared by almost all ENVRIplus Research Infrastructures (RI) and is essential for rapid quality checking for the benefit of RI operators. NRT data transmission and quality control procedures are left to independent choices of the RIs responding to their specific needs and availability of infrastructural resources. However, a common approach to those NRT technologies would be extremely useful to coordinate response of RIs in case of environmental crisis.

Within the Research Data Lifecycle (Fig. 1) of the ENVRI Reference Model (ENVRI RM), data transmission is part of the **data acquisition phase**, during which the research infrastructure collects raw data from registered sources to be stored and made accessible within the infrastructure. From the ***Science Viewpoint*** of the ENVRI RM, raw data is collected by the **Data Acquisition Community** and (streams of) measurement are brought by this community into a system for e.g. further processing steps. Data transmission is not explicitly mentioned in the Science Viewpoint but we can assume that the transmission process is meant by the phrase '*bring into a system*'. As an automated process, data transmission is part of the ENVRI RM **Data Acquisition Subsystem**.

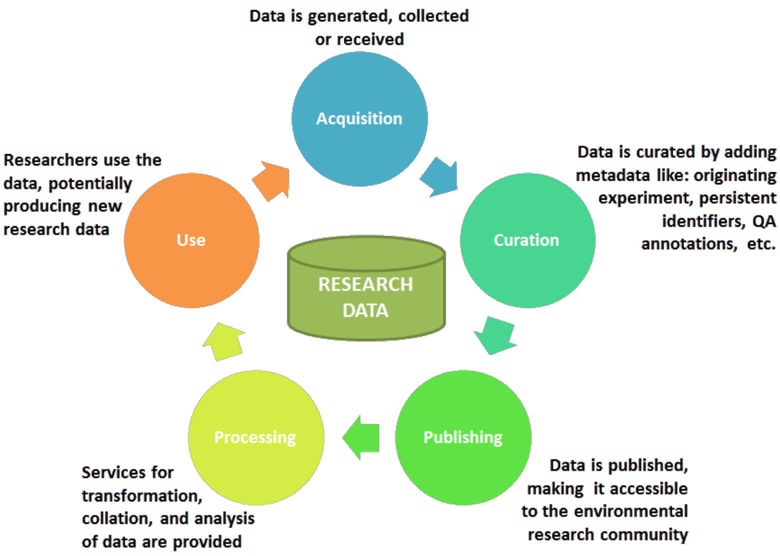


Fig. 1 The ENVRI Data Lifecycle

The transmitted raw_data is an Information Object Instance of the Information Object **PersistentData** within the ***Information Viewpoint*** of the ENVRI RM which somehow assumes that the all raw data is kept within an environmental information system which does not always apply for these systems. In the ***Computation Viewpoint*** of the ENVRI RM data transmission again is part of the **Data Acquisition Subsystem** where raw data is collected by **instrument controllers** which are managed and monitored by **acquisition services** which ensure proper data delivery into the infrastructure. To do this, acquisition services invoke dedicated **data transfer services** which   instantiate a **data transporter** or **raw data collector** which retrieve data from the instrument controller and cares for data import into the **Curation Subsystem**.

This complex model of objects, services and roles involved in data transmission processes illustrates the difficulties that are involved with harmonisation of RIs with respect to data transmission. It is deeply integrated within the Data Acquisition Subsystem which - by nature- is highly specific according to the diverse scientific purposes, objectives and needs of the associated research communities. However, the ENVRI RM also gives a clear indication where harmonisation efforts may succeed, for example at the data transfer services which mediate between instrumentation and Curation Subsystems. We will therefore in particular discuss this harmonisation potential in this document. However, because data transmission not only depends on computational aspects but in the first place depends on physical telecommunication infrastructures, we also will discuss and review key aspects and technologies of remote data transmission.

As mentioned above, NRT data transmission enables NRT quality control which is essential for e.g. the reliability of environmental event tracking or hazard management systems. Task 3.3 has a special focus on NRT QC routines therefore we will further discuss the potential of harmonized NRT data transmission for NRT QC routines and potential applications including mapping of NRT quality checking procedures across RIs. We will further introduce first results of practical implementations of these harmonisation approaches during a demonstration action performed within the EMSO infrastructure.

CONNECTIVITY FOR EUROPEAN RESEARCH INFRASTRUCTURES

AVAILABLE NETWORKS FOR DATA TRANSMISSION

Robust data transmission relies on solid technological platforms for electronic communication provided by the diverse national and international telecommunications markets. In Europe, a broad spectrum of electronic communication networks is available that offer trustworthy operation, sufficient bandwidths and broad geographical coverage. As shown above, bandwidth requirements of European research Infrastructures are very diverse, depending on the scientific focus, used sensors as well as frequency of measurements. It is therefore beyond the scope of this document to recommend a specific technology or service provider.

The table below shows a summary of telecommunication options available on the Europe market and the broad range within bandwidth offered and associated costs.

	Costs per MB	**Fixed costs (subscription)**	**Max. data transfer rate [Kbps] (uplink)**
Satellite			
Iridium dialup	ca. 85,00 $	42,50 $	2.4
Iridium OpenPort	ca.		32-128
Globalstar (Transition Plan)	ca. 35,00 $	34,99 $	9.6
Fleet 77 (MPDS)	ca. 58,00 $	99,00 $	128
Fleet ONE Coastal	4,99 $	49,99 $	100
Fleet ONE Global	9,99 $	129,99 $	100
Fleet Broadband FB150	24,99 $	749,00 $	150
Fleet Broadband FB250	24,99 $	749,00 $	284

Fleet Broadband FB500	24,99 $	749,00 $	432
mini VSAT	ca. 0,99 $	49,00 $	128- 1000
Thuraya dialup	15 $	35 $	1.2
Thuraya GmPRS	5,5 $	50 $	1.9
ThurayaIP	4 $	55 $	55.5
UHF/VHF			
Raveon			ca. 19.2
Aselan			1.2 to 16
Mobile Internet			
GSM/GPRS	0,005 $ - 0.02 $		7
EDGE	0,005 $ - 0.02 $		100
UMTS	0,005 $ - 0.02 $		64
HSDPA	0,005 $ - 0.02 $		1400
LTE	0,005 $ - 0.02 $		50000
LTE Advanced	0,005 $ - 0.02 $		500000
Cable network			
DSL			1024
VDSL			10240
GEANT			100000000

Whereas satellite communication offers limited bandwidth at high costs but in global coverage, cabled telecommunication is cheapest but in many cases not available at locations where research Infrastructures operate. Mobile telecommunication is often a good compromise but again is not available everywhere with the desired bandwidth. In addition to these long distance communication networks, research infrastructures may make use of radio communications or use more exotic short to medium distance communication methods such as wifi networks or underwater acoustic modems as a bridge to cabled or mobile networks. The latter examples further shows that research infrastructures may use several communication methods either successively or in parallel. In any case it is impossible to favour one of the mentioned systems, the best practice in choosing an appropriate telecommunication system requires balancing between technological possibilities, availability and costs.

BROADBAND CONNECTIVITY

Because many European Research Infrastructures operate in a global context or in very distant areas such as the Arctic areas or the Oceans, satellite communication is often the only possibility to transmit data in real time. However, the related costs are significant and the bandwidth offered is limited. In contrast, research infrastructures located within the European continent and near shore areas benefit from a comparable good broadband availability.

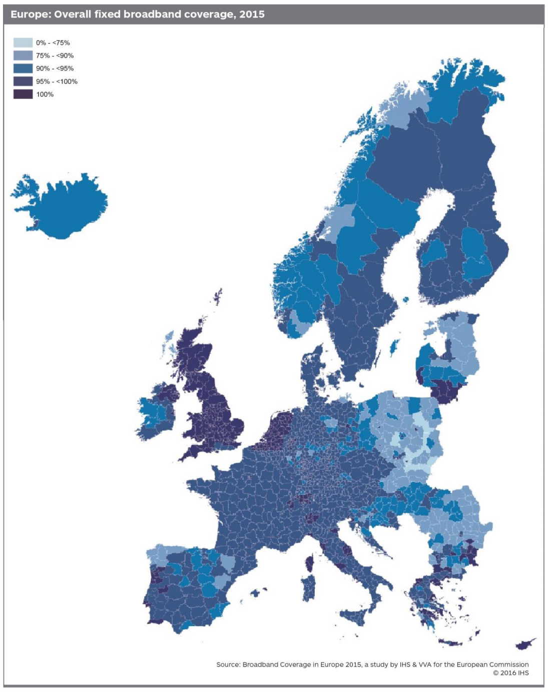
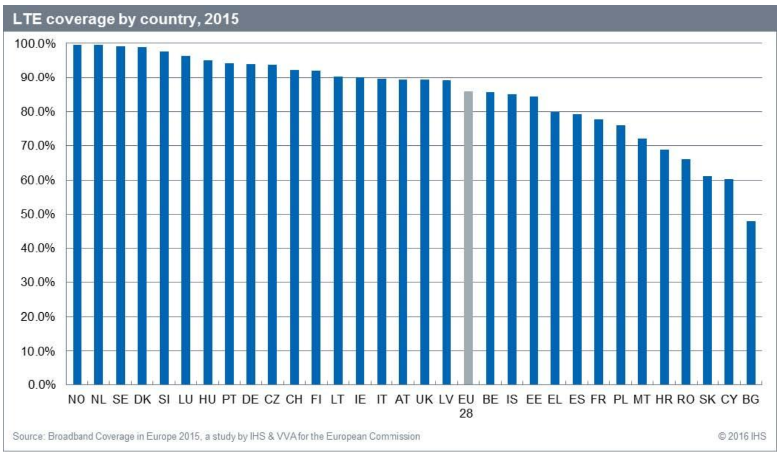


Fig. 2 Broadband coverage in Europe

In the past, the European Commission aimed at increasing the availability of broadband connections for all European countries. In its Europe 2020 strategy the commission has defined ambitious goals for 2020 within the Digital Agenda for Europe: *"(i) all Europeans have access to much higher internet speeds of above 30 Mbps and (ii) 50 % or more of European households subscribe to internet connections above 100 Mbps"*[1] and launched a series of supportive actions. This objective was reached in 2015 when 71 % of EU households had access to at least 30Mbps and almost 50% were provided with 100Mbps connectivity[2] (download capacity). While these results suggest a very good broadband coverage over Europe, it has to be mentioned that still significant differences exist between individual European countries and between rural and urban areas. In particular mobile broadband connectivity is significantly differing between European



countries.

[1] Digital Agenda for Europe: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=URISERV:si0016&from=DE

[2] Broadband Coverage in Europe 2015 https://ec.europa.eu/digital-single-market/en/news/broadband-coverage-europe-2015

Fig. 3 Availability of broadband connections in European households in 2015

Whereas Northern European countries like Norway, Netherlands and Sweden benefit from an almost 100 % LTE availability, coverages of some eastern European countries like Romania or Bulgaria range only between 50-60 %. The overall 100 Gbps coverage shows distinct differences between eastern European countries and central Europe, and distinct coverage gaps still exist in rural provinces all over Europe. In addition, coastal and shelf regions are only covered by mobile networks within a few kilometers from the coastline.

It is clear that such incomplete European broadband coverage also affects the European research landscape. Distant, rural areas are typical locations for installations of environmental research infrastructures and therefore would strongly benefit from a better connectivity by reduced costs, latency and higher available bandwidth for data transmission. It is encouraging to see that the European Commission recently has defined even more ambitious goals regarding broadband connectivity and has defined as a strategic objective for 2025 that *"All European households, rural or urban, will have access to Internet connectivity offering a downlink of at least 100 Mbps, upgradable to Gigabit speed."*[3] However, a large portion of all European research infrastructures will not have access to this broadband infrastructure for their most distant sites, their installations at e.g. arctic, alpine, coastal and open marine locations.

BEST PRACTICES FOR DATA TRANSMISSION WITHIN RESEARCH INFRASTRUCTURES

STATUS QUO AND REQUIREMENTS

During the early phase of the project we prepared and circulated a questionnaire for all technical aspects of WP3 including energy production and robustness towards extreme conditions. The questionnaire further contained several questions focussing on data transmission issues. Answers have been collected using a Google forms template, in several cases partners additionally have been interviewed during phone calls. In total, 22 responses could be collected which gives an excellent overview on the telecommunication habits of ENVRIplus collaborating infrastructures. In addition we investigated the technical documentation (if available) with respect to the technological set-up of each infrastructure.

The questionnaire replies with respect to the used telecommunication networks reflect the constraints that the mostly distant location of the infrastructure stations have to encounter. Satellite communication is the most frequently used data transmission method (11 responses), followed by mobile networks (7) and radio and long range WIFI communication (3). Some infrastructures have to make use of more than one method, e.g a radio+satellite combination.

The reasons - as given in the questionnaire - for using the chosen communication network are pragmatic. For satellite communication there is mostly no alternative, e.g. in polar region. For most infrastructures, the provider choice is based on an estimate of the best cost vs. bandwidth compromise. Iridium was frequently chosen because of this reason and its global coverage. Iridium seems to be the satellite network of choice for most marine research infrastructures, in particular those using e.g. drifting floats or operate in high latitudes (e.g. EMSO, SIOS, Euro-ARGO). The latter also uses the Argo satellite network. Iridium is further used by terrestrial infrastructures operating in high arctic or antarctic regions (EPOS, SIOS), where this system is

[3] Connectivity for a Competitive Digital Single Market - Towards a European Gigabit Society http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0587&from=en

without alternative. Mobile communication is chosen where available for cost reasons as it offers the best cost / bandwidth ratio. Broadband networks (LTE, UMTS, 3G) are preferred where available. WIFI and radio connectivity is chosen to bridge the distance between station and e.g fibre network (e.g. SIOS, EISCAT-3D, ICOS).

Most research infrastructure make use of various telecommunication technologies and networks, the decision on a distinct telecommunication platform depends on individual requirements, availability of appropriate networks as well as their costs.

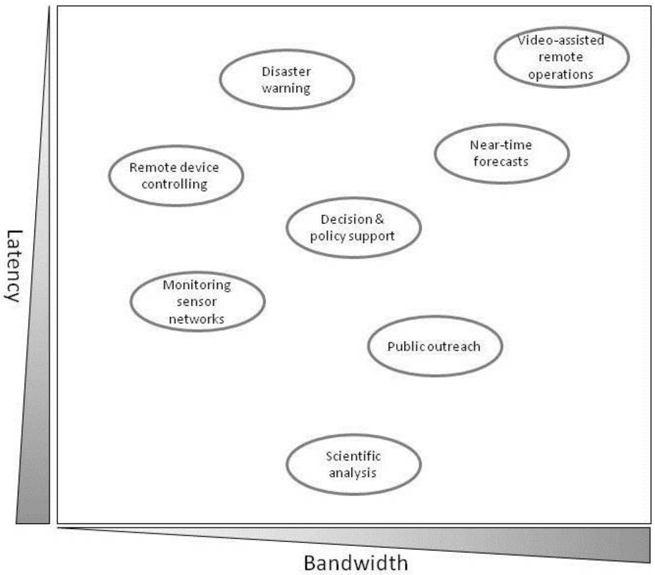


Fig. 4. Latency vs bandwidth requirements for some research infrastructure specific purposes

A good example for this multi-channel strategy is SIOS. In general data transmission in arctic and antarctic regions is a real challenge. The SIOS project has provided an excellent analysis[4] on this issue reflecting the decades of experiences with several transmission technologies ranging from ship radio to satellite links of several service providers. The raising demand regarding transmission rate, latency and bandwidth finally led to the establishment of a fibre network between Longyearbyen and Norway in 2003 which is capable to enable a Gbps connectivity. Several more distant SIOS arctic sites still have to rely on satellite or even ship radio networks. Arctic conditions also have to be managed by the INTERACT research infrastructure which operate in high latitudes or alpine regions, here data transmission is handled using the full range of available technologies depending on availability. For example WIFI and mobile network (GSM) connections are used in Siberian peatland while short wave radio connection or satellite communication is without alternative in e.g. Svalbard polar stations.

Marine research infrastructures such as FixO3, EMSO, EMBRC, Euro-GOOS, SeadataNet or JERICO, that operate at the open ocean or beyond the ocean surface, encounter many difficulties with respect to data transmission. Underwater operations need to ensure data transmission within the water column, which is either managed by underwater cables or acoustic modems. In some cases e.g. towed underwater observatories may make use of inductive modems. The latter

[4] Kaczmarska et al. (2013) D5.1 Energy and data connection strategies for the four main land-based platforms (including green energy options)

two examples only provide limited bandwidth and underwater cables are extremely expensive and rarely available. In general, marine infrastructures make use of any possible communication method and combine these if necessary. Whereas near coast stations may have the opportunity to use mobile networks (LTE, UMTS etc) or even are directly linked to terrestrial telecommunication networks via underwater cables, open sea operations still rely on satellite communication such as Iridium.

Such multi-channel communication strategies are further unavoidable for research infrastructures which operate in both the marine as well as the terrestrial domain. As described above, SIOS is one prominent example Similarly ICOS collects data from the marine as well as the atmospheric and ecosystem spheres. In particular the marine component of ICOS has to manage the same challenges as described above for marine infrastructures, because ICOS shares some marine observatories such as FRAM or PAP with EMSO and FiXO3. The ICOS series of terrestrial stations are in general connected either via cabled telecommunication networks or make use of fast mobile networks because a relatively large amount of data are submitted every day (up to hundreds of Mb). In general, ICOS makes use of a broad range of data transmission technologies.
This is also the case for the atmosphere focused ACTRIS[5] which states in its station requirements "Communication system is optimized for remote operations and the data will be stored on-board and transmitted in NRT using appropriate systems (Satellite, WiFi, WiMax, IR, …)".
Similar strategies are followed by LTER which integrates a large amount of environmental observatory stations all over Europe which partly benefit from the European mobile broadband coverage.

Special requirements exist for research infrastructures supporting hazard management and associated decision and planning processes such as earthquake and tsunami monitoring systems which are part of EPOS and EMSO. These networks need broadband connectivity, low latency and robust data transmission methods provided by terrestrial or mobile networks. While transmission of geophysical data is not within the scope of Task 3.3, it is worthwhile to mention that distant stations make use of e.g. private VSAT connection which provides high bandwidth satellite connectivity.

Only few research infrastructures make use of one or a defined number of data transmission technologies. The highly standardized Euro-ARGO, that operates a large fleet of drifting measuring buoys makes use of satellite connection via the ARGOS (preferably) or the IRIDIUM network (with the capacity of bi directional transmission) , because it is unlikely that these floats will have the opportunity to connect to e.g. mobile networks while they drift at the open seas. Limited choices for real time data transmission also existed for IAGOS which installs sensor package modules onboard of passenger aircrafts. The most important requirement hereby was to exclude disturbances on the avionic systems of the aircraft. Therefore IAGOS has chosen to use a direct connection to SATCOM because it has the lowest impact on the avionic systems. A special example is EISCAT-3D which has an extremely high demand on bandwidth capabilities. Further its instrumentation is highly sensitive to interferences with e.g. radio frequencies, therefore the EISCAT-3D sites are planned to be fully connected via optical fibre links[6]. To set up this arctic network, more than 1000 km of new fibre cables will be maintained which will be

[5] http://fp7.actris.eu/Portals/97/deliverables/PU/WP6_D6.4_M12.pdf
[6] https://eiscat3d.se/sites/default/files/e3d_docs/WP8%20deliverables/Deliverable_8_3_080910.pdf

linked with the local Gbps network.

THE FIRST MILE BOTTLENECK

As mentioned above, data transmission follows a multi-stage strategy where the sensor platform is not directly connected to a broadband network but the distance to broadband connectivity can effectively bridged with mediating communication technologies such as e.g. acoustic underwater modems or radio transmitters. In these examples, bandwidth is often limited by these technologies, e.g. an acoustic modem has a data rate of only 1-10 Kbps. This is causing a data transmission bottleneck for these research infrastructures. For some scientific applications this is acceptable but for others transmission would be to slow or - if alternative technical solution would be put into place - too expensive.

An hypothetical example is the transmission of raw data derived from novel environmental DNA (eDNA) sensors which produce a massive amount of data. An individual measurement run can result in > 50 Gbyte of raw data. If the data would be transmitted via acoustic modem or radio, the transmission would be energy hungry and slow, transmitted via a satellite network this would create extreme costs of > €25.000 per run. In comparison, EISCAT estimated the cost of maintaining a fibre cable to be about €20.000 per kilometer.

Because maintaining own fibre networks is illusory for most distant stations, the best strategy for data transmission sometimes simply may be: not to transmit. This is done by a variety of research infrastructures which collect data for which a long latency is acceptable. For example long terrestrial or marine environmental time series which serve for scientific purposes which require years for data analysis anyway, it is highly acceptable to visit these sites ‘in persona’ and pick up the storage media in regular intervals. This might be also a suitable strategy for sites which potentially are subject to vandalism or theft. And in some cases transport of storage media it is simply represents the best balance between cost and scientific benefits.

Consequently, a considerable number of research infrastructure sites e.g. from FixO3, EMSO or LTER are not connected to a communication network at all. However, in the marine community, this is not considered as the Best Practice as the technical supervision is then reacting too late after the dysfunctioning. ESONET and FixO3 recommend (Label) to send news on the status of the equipment by acoustics or messenger floats. No transmission often means long embargoes on data and the old habits of non standard quality evaluation.

HARMONIZING ROBUST DATA TRANSMISSION

Following a roundtable discussion during the first ENVRIweek, it became clear that data transmission formats are very diverse. In general, data transmission at the sensor as well as platform level (Fig.5) largely depends on community specific needs and habits or simply on manufacturer specifications. Both result in proprietary or niche formats and protocols that require data to subsequently be processed by data transformation services before they can be delivered in a standardized format (Fig. 6).

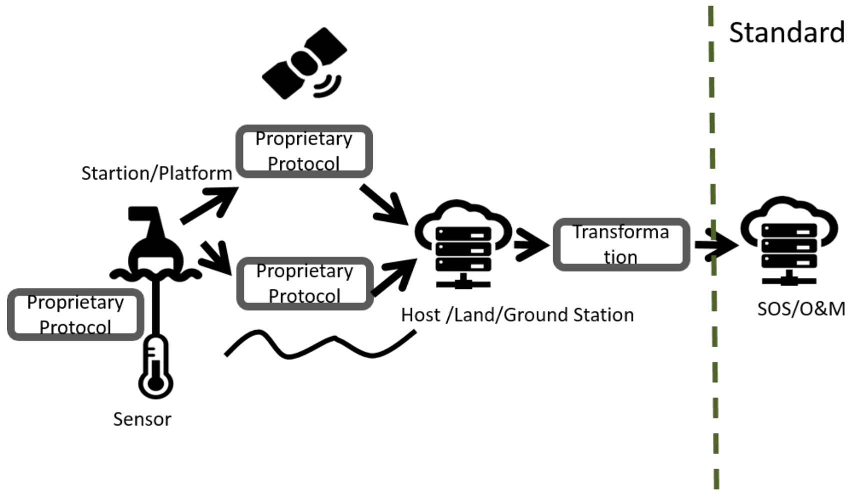


Fig. 5 Current situation of data transmission at an example marine infrastructure platform. Data is transmitted via a proprietary (manufacturer) format from sensors to the platform which transmits the data in another proprietary (community) format to a land station and its associated data management platform. Here, the data is transformed to a standards OGC compliant format and distributed via a standard protocol (Sensor Observation Service).

Following the ENVRI Reference Model, the status quo evaluation above shows, that **data Acquisition** services and in particular the preparation of data transfer (ENVRI RM: **prepare data transfer**) prior to data transmission are not yet sufficiently standardized. This hinders efficient, multi RI (Research Infrastructure) data processing routines such as Data Quality checking.

Since the above described technical and infrastructural constraints resulted in a very broad spectrum of physical data transmission technologies which therefore cannot be standardized across RIs, standardising data transmission formats is the most promising approach towards a common robust data transmission strategy. Because Task 3.3 further focusses on near real time quality control, special consideration is given in the following chapters to this aspect and the corresponding requirements for data transmission. However, the applicability of these standards still needs to be tested or proven in operational data transmission and subsequent NRT QC workflows. The following sections therefore focus on these applicability of some interdisciplinary candidate standards within the environmental RI community.

DATA TRANSMISSION REQUIREMENTS FOR CROSS RI NRT QUALITY CONTROL

NRT QC policies and best practices

Near Real Time quality control is strongly dependent on the scientific domain of each research infrastructure and not every RI already has a NRT policy in place. ARGO and ICOS which have already a NRT quality control policy, made the description of the QA/QC available to the ENVRIPlus community. Also, in the frame of ENVRIPlus, a quality control manual for IAGOS-ICH (water vapour measurements) was prepared. These documents allow for a deeper insight in the existing QA/QC-procedures.

ARGO has a very advanced QC-standard. A RTQC procedure has been agreed at international level and implemented in each Argo data assembly center. Nevertheless, it states that "because of the requirement for delivering data to users in NRT, the quality control procedures on the near-real-time data are limited and automatic." The same is valid for IAGOS-ICH. ARGO and IAGOS-ICH use around 20 QC routines. For IAGOS-ICH nearly half (9/20) of the implemented quality checks are specifically developed for the ICH sensor and, therefore, are not suitable for harmonization. The remaining tests are mostly the simple range-, outlier- or spike- tests which are mentioned above, where harmonization should be straightforward to adapt.

ICOS (Integrated Carbon Observation System) span across three main domains: atmosphere, ecosystem and ocean. This makes a harmonization of the Near Real Time Data Processing very complex: "Although many aspects of data management can and will be harmonized throughout the RI, there exists a broad range of tools and practices that are in parallel use, especially when comparing the details of the different thematic centres' work flow. The challenge is to bring together all outcomes under a common data curation scheme." For ENVRIPlus ICOS is a good show case for how to achieve a high standardization level agreed between different domains.

ARGO and ICOS are also part of the project AtlantOS, which developed recommendations for an automatic RT or NRT QC for selected variables (T&S, Current, Oxygen, CHla, Nitrate, Carbon, Sea level).

The U.S. Integrated Ocean Observing System (IOOS) established a Quality Control/Quality Assurance of Real-Time Oceanographic Data (QARTOD) Project Plan. It is a guide for establishing widely accepted procedures for the quality control of real-time data through the preparation of manuals for several of the 26 core variables identified in the plan. Currently, 17 federal organizations and 11 Regional Associations (RAs) are named as partners in U.S. IOOS.

The ENVRIPlus community was requested to review this plan. Since, QARTOD gained extensive experience in the harmonization of QA/QC of different RIs, this plan is also summarized here.

The strategy of QARTOD is to provide "enough guidance to be meaningful, yet not overly prescriptive. The Project Plan framework allowed for innovative ideas to be discussed and adopted, while adhering to the general direction of the work plan." The initiative for QARTOD started in the year 2004. QUARTOD established seven Data Management Laws of QARTOD, which gave the fundament for the subsequent process. This yielded in an eight-page action plan in 2012 which outlines objectives for establishing QA/QC procedures for real-time data of the core variables. Nine manuals have been prepared to date, a tenth is in progress, and five of the nine manuals have been updated. "The lesson learned (or being learned) is that the introduction of the intention to proceed with a manual should not be underestimated. It takes time and effort to build the needed community support" (M. Bushnell, personal communication).

Cross domain NRT QC routines

We have collected information on commonly used NRT quality routines from ENVRIplus RIs during workshops organized during the last two ENVRIweek meetings as well as through responses to a dedicated questionnaire we sent to RI representatives.

.	EMSO	FixO3	EuroARGO	ACTRIS (surf.)	ACTRIS (lidar)	IAGOS	ICOS (ETC)	ICOS (OTC)	GROOM	Count
Data Integrity Test	-	-	-	x	-	-	X	-	-	1
Metadata Consistency Test	-	-	-	x	x	x	X	-	-	4
Platform/Sensor Identification	x	x	x	x	-	-	X	-	x	8
Date/Time Check	x	x	x	-	-	x	X	x	x	7
Location Check	-	x	x	-	-	x	-	x	x	5
Spike or Outlier Test	x	x	x	x	-	x	x	x	x	8
Gradient Test	x	x	x	-	-	-	-	-	x	4
Stuck or Constant Value Test	-	x	x	-	-	x	X	x	x	6
Digit Rollover Test	-	-	x	-	-	-	-	-	x	2
Gap Test	-	-	-	-	-	x	x	-	-	2
Rate of Change / Step Test	-	x	-	-	-	x	X	-	-	3
Range Test (Regionality/Past Values)	-	x	x	-	-	x	-	x	x	5
Range Test (Global)	-	x	x	-	-	x	X	-	-	4
Range Test (Instrument Limits)	-	-	-	x	-	x	x	x	-	4
Range Test (Implicit)	-	-	-	x	x	x	-	-	-	3

Not every ENVRIplus RI already has a NRT quality control policy in place, however we received a sufficient number of responses from those RIs involved in this task to identify clear communalities among those RIs with respect to NRT QC routines. Most commonly used are simple test such as outlier or spike detection, gradient or stuck value tests.

While these rather simple checks only require a more or less continuous stream of data, other tests have additional requirements we need to consider for our harmonisation attempts for example the frequently required platform sensor identification verification as well as the often used range tests. Both checks are impossible if the transmitted data stream is not properly connected to sensor information. It is therefore important to link this data with at least some essential metadata which allows to identify the delivering sensor and its constraints. For time series data, every measurement needs to be timestamped. It is further necessary to provide information regarding the observed property a measured value belongs to as well as its associated unit. Sensor specific standards already support these requirements. It could therefore be highly advantageous for environmental RIs to utilize these standards already for data transmission purposes.

BRINGING STANDARDS CLOSER TO THE SENSOR

As described above, the approach of this task group was to introduce standardized data transmission formats and protocol into the **Data Acquisition** workflow, thus to move the standardisation level closer to the sensor. Currently, two major standards exist which support handling and transmission of sensor data:

The **Sensor Web Enablement (SWE)**[7] family of standards published by the Open Geospatial Consortium (OGC). Here, in particular the sensor description format SensorML are relevant to encode sensor constraints required for NRT QC and the Sensor Observation Service (SOS) for transmission of observational data. SOS can be used as a Pull-Service to query and download e.g the latest data but also offers a Push-Service, via its Transactional SOS capabilities which allow to transmit data in a bandwidth friendly way.

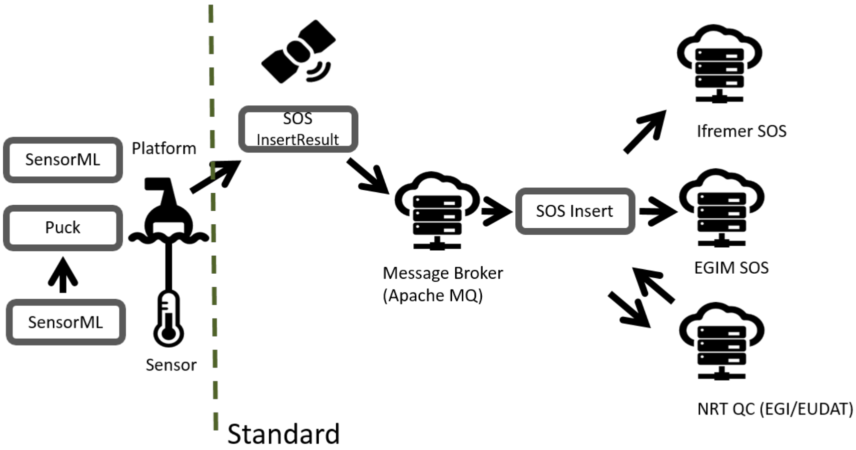


Fig 6. Standardisation options using the OGC SWE family of standards

The **Semantic Sensor Network Ontology (SSNO)** ontology defines a set of classes and properties which allow to effectively describe sensors as well as their output. SSNO can be encoded in various formats such as RDF or JSON-LD which can be used to stream data in a standardised way or to query this data via standard SPARQL from a dedicated triple store.

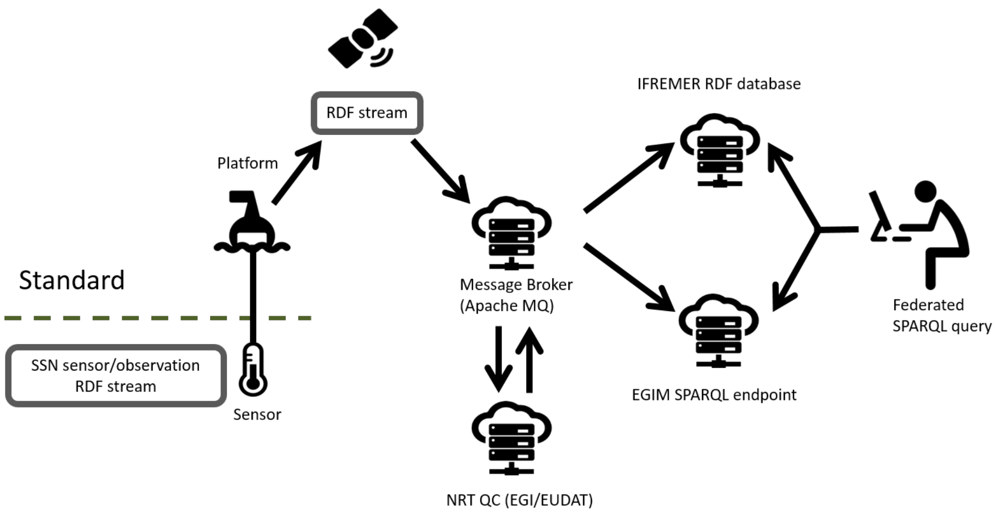


Fig 7. Standardisation options using the Semantic Sensor Network Ontology (SSNO)

Both standard families have been chosen to test future options for data transmission harmonisation initially with a special focus on OGC standards as these are the most commonly used or well known standards within the environmental research infrastructure community.

[7] http://www.opengeospatial.org/ogc/markets-technologies/swe

As mentioned before, Task 3.3 treats data transmission with special consideration of real time quality control aspects. It is therefore necessary to choose an appropriate data transmission format which is able to keep the instrument (sensor) context for each measurement in order to verify these against the sensor's constraints. Within the OGC SWE family of standards, **SensorML** is the format which is capable to describe sensor metadata, **OGC O&M (Observation & Measurement)**[8] is the preferred data format. Alternatively, the SOS offers a compact data format which is similar to a simple ASCII tabular format but also is able to define row and column separators.

However, OGC standards are quite flexible and require some further consolidation and community specific adaptations in order to provide a unambiguous format. Therefore, we have intensively worked on the realization of a joint marine SensorML profile in cooperation with Task 8.2 and TC_4 as a starting point. These activities are coordinated within the marine-swe-profiles group which uses a WIKI and mailing lists provided by 52°North to collect best practices and to moderate the discussion. This already led to some community consensus regarding the marine SensorML profile for example the use of distinct vocabularies within SensorML provided by Seadatanet (NERC).

These agreements made it possible to map some essential information held in the SensorML with SSNO. For example an appropriate ontology has been agreed on regarding measurement capabilities such as accuracy, resolution, range etc. which made it easy to map this with SSNO. Using this we defined a XSLT template transform SensorML XML information into a JSON-LD encoding of SSNO. As a first approach we followed the proposal of Calbimonte (2016) and distinguished between Dataset level information such as contributor, title, document validity etc. and core sensor information. An example SSNO sensor description can be found in the appendix.

Both, SensorML as well as SSNO JSON-LD files have been proven to be appropriate to describe sensors in the necessary detail required for data transmission and subsequent NRT QC and will be continuously used in the following months of the project for near real time quality control tests within this task and in cooperation with WP7 and WP9 during a use case which will be described later in this document.

TESTING HARMONIZED DATA TRANSMISSION

To test the suitability of the above mentioned standards for sensor data encoding (O&M, SSNO, SOS), we have chosen two approaches:

First, we initiated a demonstration action to test data transmission in a real life scenario. The MARUM waveglider, has been chosen as Tasks 3.3 demonstration action platform and had its first deployment near the Canary Islands in february 2017. The platform has been assembled by a Spanish company together with MARUM and is using software by UPC to collect and transmit data in a OGC transactional SOS compliant way as well as in binary format (for ACDP data).

The MARUM waveglider is equipped with the following sensors:

- Seabird Glider Payload CTD (GPCTD) with SBE 43F
- Turner C3 Submersible Fluorometer 2300-000 (clorophyll, turbidity and oil)

[8] http://www.opengeospatial.org/standards/om

- Sensorlab SP200-SM High Accuracy in Situ pH Sensor
- Teledyne Workhorse Monitor ADCP Model WHM300-1000

We provided a detailed documentation of the wavegliders instrumentation as well as the platform itself in OGC SensorML format. SensorML files are available here:

- http://dataportals.pangaea.de/sml/db/waveglider/ssw_58db9bd7d96d0.xml
- http://dataportals.pangaea.de/sml/db/waveglider/ssw_58db9cc87da06.xml
- http://dataportals.pangaea.de/sml/db/waveglider/ssw_58db9ccf4be4a.xml
- http://dataportals.pangaea.de/sml/db/waveglider/ssw_58db9cd6a7805.xml
- http://dataportals.pangaea.de/sml/db/waveglider/ssw_58db9cdb645f5.xml

During the test case it was planned to transmit data in an hourly interval via Iridium connection. During the first deployment data transmission was successfully tested, however due to yet unknown reasons data transmission was interrupted during the waveglider's autonomous cruise after some hours and could not be reestablished.

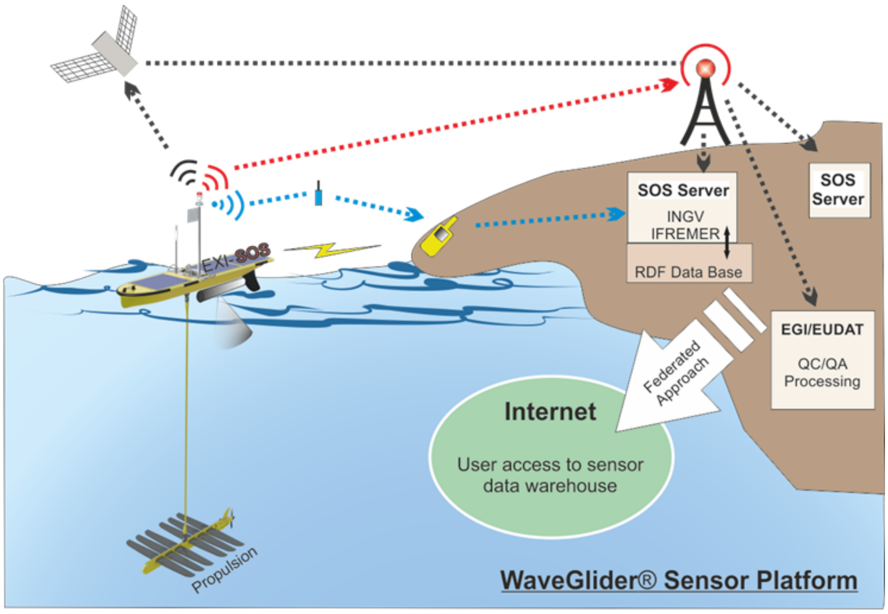


Fig. 8 Schematic representation of the Task 3.3 demonstration action for which the MARUM waveglider was chosen as platform.

We expect the return of the platform in May 2017 at the Canary Islands which will allow to fix the problem before the next mission. Despite these problems data transmission of transactional SOS data proved to be a true option for at least low-bandwidth data. In comparison with the simultaneously transmitted binary data, this format could easily be checked for completeness and correctness without proprietary software. The transactional SOS XML format proved to be less vulnerable to file corruption in comparison to the binary format. Further, this format could easily be ingested into a SOS Server operated by MARUM without further format conversion steps and thus could be earlier integrated into the data management workflow of MARUM and PANGAEA respectively. As mentioned earlier, due to a technical problem the demonstration action could only partially be completed we will therefore continue our tests during the next waveglider mission.

Second, we have used data archived in PANGAEA which originates from frequently used sensor types and transferred these data into the test candidates SSNO

(JSON-LD) as well as the Transactional SOS InsertResult (XML) as well as in OM-JSON.

Here are examples which encode a subset of data in these formats:

SSNO as JSON-LD:


```
{"@context": {
      "ssn": "http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#",
              "schema": "https://schema.org/"
  },"@graph":[
            {
              "@type":"ssn:Observation",
              "@id":"ex:ts:r:1:c:5",
              "ssn:observedBy":"http://dataportals.pangaea.de/sml/db/waveglider/ssw_3527945043_58b536f73e0cf.xml",
              "ssn:observedProperty":"http://vocab.nerc.ac.uk/collection/P01/current/ADEPZZ01",
              "ssn:observationResult":{
                "schema:value":3.0,
                "schema:unitCode":"m"
              },
              "ssn:observationResultTime":"2016-08-17T10:57"},
            {
              "@type":"ssn:Observation",
              "@id":"ex:ts:r:1:c:6",
              "ssn:observedBy":"http://dataportals.pangaea.de/sml/db/waveglider/ssw_3527945043_58b536f73e0cf.xml",
              "ssn:observedProperty":"http://vocab.nerc.ac.uk/collection/P01/current/PPSCZZ01",
              "ssn:observationResult":{
                "schema:value":3.0,
                "schema:unitCode":"dbar"
              },
              "ssn:observationResultTime":"2016-08-17T10:57"},

etc. etc.
```


OM-JSON:


```
{"type" : "http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_DiscreteTimeSeriesObservation",
              "observedProperty": {"href":"http://vocab.nerc.ac.uk/collection/P01/current/ODSDM021"},
              "procedure":
{"href":"http://dataportals.pangaea.de/sml/db/waveglider/ssw_3527945043_58b536f73e0cf.xml"},
              "result": {
                "defaultPointMetadata": {
                  "uom" : "1"
                },
                "points":[{"time":  {"instant":"2016-08-17T10:57"},"value":30.0838},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0822},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0755},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0742},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0677},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0508},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0548},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0623},
            {"time":  {"instant":"2016-08-17T10:57"},"value":30.0693}]}}
```


SOS InsertResult:


```
<?xml version="1.0" encoding="UTF-8"?><sos:InsertResult xmlns:sos="http://www.opengis.net/sos/2.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" service="SOS" version="2.0.0"
xsi:schemaLocation="http://www.opengis.net/sos/2.0 http://schemas.opengis.net/sos/2.0/sos.xsd">
<sos:template>291EC6DA-43F8-46D3-86F5-628FBF955261</sos:template>
<sos:resultValues>2016-08-17T10:57#3.0#3.0#40.7980#30.0838@2016-08-17T10:57#3.2#3.2#40.7788#30.0822@2016-08-
17T10:57#3.4#3.4#40.7069#30.0755@2016-08-17T10:57#3.6#3.6#40.6992#30.0742@2016-08-
17T10:57#3.8#3.8#40.6478#30.0677@2016-08-17T10:57#4.0#4.0#40.6172#30.0617@2016-08-
17T10:57#4.2#4.2#40.5555#30.0540@2016-08-17T10:57#4.4#4.4#40.5534#30.0613@2016-08-
17T10:57#4.6#4.6#40.5724#30.0689@2016-08-17T10:57#4.8#4.8#40.5915#30.0765@</sos:resultValues>
</sos:InsertResult>
```


For this experimental setup we have linked these data formats with the above mentioned SensorML sensor metadata of the MARUM waveglider use case. For SSNO as well as OM-JSON this could easily be achieved by linking the SensorML file via the ‘ssn:observedBy’ or ‘procedure’ variable respectively. The compact SOS InsertResult format however, does not allow specifying sensor information. Instead, a ‘sos:template’ element has to be used to link with another XML document, the ‘sos:ResultTemplate’:

```
<sos:proposedTemplate>
    <sos:ResultTemplate>
      <swes:identifier>291EC6DA-43F8-46D3-86F5-628FBF955261</swes:identifier>
      <sos:offering>Conductivity_Temperature_Depth_Ph</sos:offering>
      <sos:observationTemplate>
        <om:OM_Observation gml:id="sensor2obsTemplate">
          <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_ComplexObservation"/>
          <om:phenomenonTime nilReason="template"/>
          <om:resultTime nilReason="template"/>
          <om:procedure
xlink:href="http://dataportals.pangaea.de/sml/db/waveglider/ssw_3527945043_58b536f73e0cf.xml"/>
etc.etc..
```

While for all formats the encoding rules are straightforward and could be implemented without problems, the differences between these three formats are apparent from the given examples and most important: significantly differ in string size.

For our test we have encoded example sensor data to estimate the potential transmission file size. We have chosen the typical outputs of the sensor ‘Seabird Glider Payload CTD (GPCTD) with SBE 43F’ which is installed at the MARUM waveglider. This sensor delivers pressure, temperature, conductivity as well as oxygen concentration as output of one measurement run.

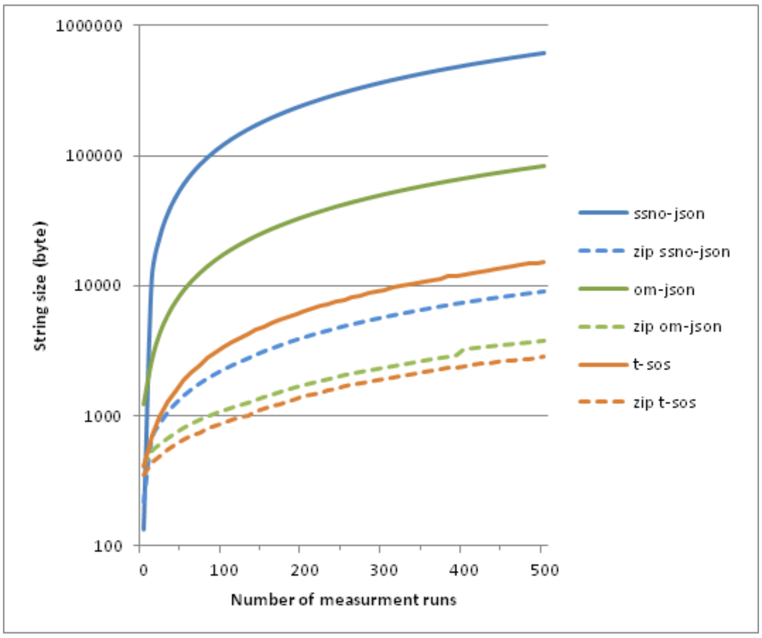


Fig. 9 String sizes vs. encoded data rows (measurement runs) of the test data encoding formats. Please note the logarithmic scale of the y-axis.

In steps of 10 measurements runs, each resulting in 40 individual measurements, we have produced outputs for each of the three formats. In addition, we have compressed each output string using the ZIP algorithm to estimate the compressed file sizes.

Obviously, The SSNO-JSON encoded data results in largest file sizes. In comparison to the SOS (InsertResult) encoded data, the resulting SSNO-JSON string is about fourty times (!) larger which is a result of the necessary repetition of sensor id as well as parameter id for each measurement which semantically is correct but causes the observed string sizes. The OM-JSON string still is about seven times larger in comparison to the SSNO-JSON string but only five times larger than the SOS string. Our test also showed that file sizes can significantly be reduced by compressing algorithms. For example, the zipped SSNO-JSON can be reduced to about 14% of the original size. As a result, zipped SSNO-JSON files are about 2.4 times as big as zipped OM-JSON files which are 1.2 times larger in comparison to zipped SOS files.

In general, the SOS InsertResult encoding is the by far most compact format available but a real disadvantage is that sensor information is not directly connected to the delivered data. OM-JSON results in significantly larger file sizes but keeps the link between sensor information and data. Whereas the very large SSNO semantically ideally ensures the link between sensor information and each measurement but associated file sizes will be unacceptable for many applications. In comparison, OM-JSON may be the best compromise in terms of file size vs. semantics, however the compact format of SOS InsertResult encoding still may serve as the most pragmatic approach which allows to transfer standard compliant data at lowest possible costs.

CONCLUSIONS

We have shown in this deliverable that data transmission is not an easy task and hardly can be harmonized - in technical terms - across all ENVRIplus research infrastructures due to the very different requirements regarding e.g. bandwith and latency as well as due to the lack of broadband connectivity at distant sites. However, we also showed that emerging standards would enable ENVRIplus RIs to standardise data transmission formats which would in turn enable cross domain data processing services such as common NRT quality control routines. During the remaining months of the ENVRIplus project we will investigate this in more detail and in particular will make first steps towards the realisation of these common services.

Following the recommendations of the task group which have been agreed on during the second ENVRIweek meeting, we therefore applied for a use case in WP9 which proposed to use standards compliant data transmission technologies very early in the data workflow, to use appropriate formats (OGC transactional SOS, SSNO RDF stream etc.) for data transmission, and to use these standardized data transmission formats for cross-domain NRT quality checks. The use case (IC_14) was positively evaluated and we could start to implement the proposed EGI technology framework on the EGI cloud, which is an excellent opportunity for us to evaluate new approaches in this environment. EGI already has provided an Apache Storm installation on their cloud cluster in order to support the implementation our planned approach. EGI has also installed Apache Cassandra on their cloud, which is expected to serve as a (temporary) cache for quality-annotated data. During the next months we will implement a dedicated Storm Topology for our use case, with appropriate spouts and bolts. The next step involves testing a QC routine on actual sensor data delivered by the MARUM waveglider during its next deployment.

IMPACT ON PROJECT

ENVRIplus with its outstanding consortium consisting of an exceptional mixture of engineering, IT people and scientists provides an excellent opportunity to explore innovative solutions for common RI challenges such as data transmission. This task started as a rather conventional evaluation of data transmission technologies, their capabilities, constraints and costs but soon it became clear that the exclusive consideration of the physical layer of data transmission is not sufficient. Stimulated by the NRT quality control focus of Task 3.3 the standardisation potential of all data transmission related aspects was considered - guided by the ENVRI Reference Model. The deliverable therefore nicely illustrates the impact engineering decisions and the choice of technology platforms may have on software and related data and metadata formats and vice versa and how the ENVRI RM can contribute to identify standardisation options which may stimulate similar approaches within the project. The work presented here also has strong impact on the project as we have chosen to draft a WP9 use case (IC_14) to continue our work and to contribute to other workpackages by implementing test case scenarios for NRT QC on the EGI infrastructure. The document might also have impact on the work of WP7 where for example the benchmarks of data transmission formats may be useful to identify performance optimisation potentials.

REFERENCES

Calbimonte, J-P, 2016: Toward Semantic Sensor Data Archives on the Web. In: Proceedings of the 2nd Workshop on Managing the Evolution and Preservation of the Data Web colocated with ESWC 2016. https://infoscience.epfl.ch/record/218545/files/semantic-sensor-data(3).pdf

APPENDIX

Example JSON-LD representation of a SSNO sensor description as a result of a XSLT transformation from an OGC SensorML file:

```
{
    "@context": {
        "dcat": "http://www.w3.org/ns/dcat#",
        "ssn": "http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#",
        "sml": "http://www.opengis.net/sensorml/2.0",
        "DataScientist": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0006/",
        "Manufacturer": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0001/",
        "Operator": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0003/",
        "Owner": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0002/",
        "PrincipalInvestigator": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0004/",
        "TechnicalCoordinator": "http://vocab.nerc.ac.uk/collection/W08/current/CONT0005/",
        "schema": "https://schema.org/",
        "LongName": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0002/",
        "ShortName": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0006/",
        "SerialNumber": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0005/",
        "ApplicationDomain": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0011/",
        "CallSign": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0010/",
        "ICESCode": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0001/",
        "ModelName": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0003/",
        "ModelNumber": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0004/",
        "UUID": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0007/",
        "UniqueID": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN008/",
        "Version": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0013/",
        "WMOPlatformNumber": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0009/",
        "ManufacturerName": "http://vocab.nerc.ac.uk/collection/W07/current/IDEN0012/",
        "HousingMaterial": "http://vocab.nerc.ac.uk/collection/W05/current/CHAR0005/",
        "DataStorage": "http://vocab.nerc.ac.uk/collection/W05/current/CHAR0006/",
        "TransmissionMode": "http://vocab.nerc.ac.uk/collection/W05/current/CHAR0007/",
        "DampingRatio": "http://vocab.nerc.ac.uk/collection/W04/current/CAPB0002/",
        "OperatingDepth": "http://vocab.nerc.ac.uk/collection/W04/current/CAPB0006/",
        "SurvivalDepth": "http://vocab.nerc.ac.uk/collection/W04/current/CAPB0013/",
        "hasInstrumentType": "http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/",
        "hasPropertyValue": "https://schema.org/PropertyValue"
    },
    "@graph": [
        {
            "@type": "dcat:Dataset",
            "dcat:title": "SBE 37-SM MicroCAT C-T (P) Recorder",
            "dcat:description": "Moored Conductivity, Temperature, and (optional) Pressure measurements, at user-
programmable and and ' intervals. RS-232 or RS-485 interface, internal memory, and internal battery pack.",
            "ShortName": "SBE 37-SM MicroCAT C-T (P)",
            "LongName": "SBE 37-SM MicroCAT C-T (P) Recorder",
            "ManufacturerName": "Sea-Bird Electronics, Inc.",
            "dcat:contactPoint": [
                {
                    "@type": [
                        "Operator",
                        "schema:Person"
                    ],
                    "schema:name": "Test Mueller & sons"
                },
                {
                    "@type": "schema:Person",
                    "schema:name": "Robert Huber"
                }
            ]
        },
        {
            "@type": "ssn:SensingDevice",
            "@id": "ssw_3527945043_57c00bde7f5a4",
            "hasInstrumentType": "http://vocab.nerc.ac.uk/collection/L05/current/130/",
            "ssn:hasMeasurementCapability": [
                {
```

```
            "ssn:hasMeasurementProperty": {
                "@id": "http://vocab.ndg.nerc.ac.uk/term/w04/current/CAPB0001",
                "hasPropertyValue": {
                    "schema:value": "0.0003",
                    "schema:unitCode": "S/m"
                }
            },
            "ssn:forProperty": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/PCONZZ01"
        },
        {
            "ssn:hasMeasurementProperty": {
                "@id": "ssn:Resolution",
                "hasPropertyValue": {
                    "schema:value": "0.00001",
                    "schema:unitCode": "S/m"
                }
            },
            "ssn:forProperty": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/PCONZZ01"
        },
        {
            "ssn:hasMeasurementProperty": {
                "@id": "ssn:MeasurementRange",
                "hasPropertyValue": {
                    "schema:value": "0 7",
                    "schema:unitCode": "1"
                }
            },
            "ssn:forProperty": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/PCONZZ01"
        },
        {
            "ssn:hasMeasurementProperty": {
                "@id": "ssn:Accuracy",
                "hasPropertyValue": {
                    "schema:value": "0.002",
                    "schema:unitCode": "degC"
                }
            },
            "ssn:forProperty": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/TEMPPR01"
        },
        {
            "ssn:hasMeasurementProperty": {
                "@id": "ssn:Resolution",
                "hasPropertyValue": {
                    "schema:value": "0.0001",
                    "schema:unitCode": "degC"
                }
            },
            "ssn:forProperty": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/TEMPPR01"
        }
    ],
    "ssn:hasOperatingProperty": [
        {
            "@id": "urn:ogc:def:property:Datastorage",
            "hasPropertyValue": {
                "schema:value": "530000",
                "schema:unitCode": "samples"
            }
        },
        {
            "@id": "schema:height",
            "hasPropertyValue": {
                "schema:value": "0.2",
                "schema:unitCode": "m"
```

```
                    }
                },
                {
                    "@id": "schema:width",
                    "hasPropertyValue": {
                        "schema:value": "0.15",
                        "schema:unitCode": "m"
                    }
                },
                {
                    "@type": "OperatingDepth",
                    "hasPropertyValue": {
                        "schema:value": "350",
                        "schema:unitCode": "m"
                    }
                },
                {
                    "@type": "SurvivalDepth",
                    "hasPropertyValue": {
                        "schema:value": "500",
                        "schema:unitCode": "m"
                    }
                }
            ],
            "ssn:observes": [
                {
                    "@id": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/PCONZZ01",
                    "schema:unitCode": "mS/cm"
                },
                {
                    "@id": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/TEMPPR01",
                    "schema:unitCode": "degC"
                },
                {
                    "@id": "http://vocab.ndg.nerc.ac.uk/collection/P01/current/PPSCZZ01",
                    "schema:unitCode": "degC"
                }
            ]
        }
    ]
}
```