



HAL
open science

RÉSIF-SI: A Distributed Information System for French Seismological Data

Catherine Péquegnat, Jonathan Schaeffer, Claudio Satriano, Helle A. Pedersen, Jérôme Touvier, Jean-Marie Saurel, Marie Calvet, Laurent Stehly, Patrick Arnoul, Philippe Bollard, et al.

► **To cite this version:**

Catherine Péquegnat, Jonathan Schaeffer, Claudio Satriano, Helle A. Pedersen, Jérôme Touvier, et al.. RÉSIF-SI: A Distributed Information System for French Seismological Data. *Seismological Research Letters*, 2021, 92 (3), pp.1832-1853. 10.1785/0220200392 . hal-03193314v2

HAL Id: hal-03193314

<https://hal.science/hal-03193314v2>

Submitted on 9 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RÉSIF-SI: A DISTRIBUTED INFORMATION SYSTEM FOR FRENCH SEISMOLOGICAL DATA

Catherine Péquegnat¹, Jonathan Schaeffer², Claudio Satriano^{3,*}, Helle Pedersen^{1,2}, Jérôme Touvier¹, Jean-Marie Saurel³, Marie Calvet⁴, Laurent Stehly^{1,2}, Patrick Arnoul⁵, Philippe Bollard², Olivier Dewée⁶, Fabien Engels⁷, Marc Grunberg⁷, Constanza Pardo³, David Wolyniec², Grégory Arneodo¹, Rima Bouazzouz¹, Olivier Geber³, Coralie Aubert¹, Mickaël Langlais¹, Sandrine Roussel¹, Florent Brenguier^{1,2}, Wayne C Crawford³, Arnaud Lemarchand³, Emeline Maufroy^{1,2}, Serge Olivier⁵, Diane Rivet⁸, Martin Vallée³, Jérôme Vergne⁹, Jérôme Chèze⁸, Éric Debayle¹⁰, Bertrand Delouis⁸, Philippe Guéguen¹, Sophie Lambotte⁹, Antoine Mocquet¹¹, Anne Paul¹, Fabrice Peix⁸, François Schindelé⁵, Nikolai Shapiro¹, Éléonore Stutzmann³, Pierre Volcke¹, and Andrea Walpersdorf^{1,2}

¹Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France

²Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

³Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

⁴Institut de Recherche en Astrophysique et Planétologie, Université Toulouse III Paul Sabatier, CNRS, CNES, Toulouse, France

⁵CEA/DAM/DIF, F-91297 Arpajon, France

⁶Division Technique de l'INSU, Meudon, France

⁷Université de Strasbourg, CNRS, EOST UMS 830, 67000, Strasbourg, France

⁸Université Côte d'Azur, CNRS, IRD, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France

⁹Université de Strasbourg, CNRS, IPGS UMR7516, 67000, Strasbourg, France

¹⁰Laboratoire de Géologie de Lyon : Terre, Planètes, Environnement, CNRS, UMR 5276, École Normale Supérieure de Lyon, Université de Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France

¹¹Laboratoire de Planétologie et de Géodynamique, Université de Nantes, Université d'Angers, UMR-CNRS 6112, Nantes, France

The authors acknowledge there are no conflicts of interest recorded.

*Corresponding Author

Claudio Satriano

Université de Paris, Institut de physique du globe de Paris, CNRS

1 rue Jussieu, F-75005 Paris, France

email: satriano@ipgp.fr

Keywords: seismological data; data management; data access; France

Fri 9th Apr, 2021, 16:07

1 **Abstract**

2 The Résif project, which started in 2008, aims at gathering under a common research infrastructure the
3 French seismological, GNSS and gravimeter permanent networks, as well as the mobile instrument
4 pools. A central part of Résif is its seismological information system, Résif-SI (started in 2012), which
5 is in charge of collecting, validating, archiving and distributing seismological data and metadata from
6 seven national centers. Résif-SI follows a distributed architecture, where the six data collection and
7 validation centers (A-nodes) send validated data and metadata to a national data center (Résif-DC)
8 which is the central point for data archiving and distribution. Résif-SI is based on international
9 standard formats and protocols and is fully integrated into European and international data exchange
10 systems (EIDA, EPOS, IRIS, FDSN). In this paper we present the organization of Résif-SI, the
11 technical details of its implementation and the catalog of services provided to the end users. The
12 paper is aimed both at seismologists, who want to discover and use Résif data, and at data center
13 operators, who might be interested in the technical choices made in the implementation of Résif-SI.
14 We believe that Résif-SI can be a model for other countries facing the problem of integrating different
15 organizations into a centralized seismological information system.

16 Introduction

17 Résif (Réseau sismologique et géodésique français – French seismological and geodetical network
18 – <https://www.resif.fr>) started as a project in 2008, with the aim of grouping into a single research
19 infrastructure all the major national and regional seismic, GNSS and gravimeter permanent networks
20 and mobile pools in France. Hand in hand with this objective was to improve and rationalize the data
21 distribution system. GNSS data from all Résif partners is distributed from the national Résif data
22 center for GNSS data (Résif, 2017), which also hosts the EPOS GNSS data gateway, while the Résif
23 data from permanent gravimeters are directly integrated into dedicated international data centers
24 (Voigt et al., 2016). In this article we focus on the seismological part of the Résif Information System
25 (Résif-SI), for which the architecture, technical implementation and governance structure has proven
26 to be adequate to address data distribution through a coordinated national cooperative. Note that the
27 present article refers to seismic waveform data and metadata: higher-level earthquake products in
28 France are now also being integrated into Résif, as discussed in Masson et al. (2021).

29 Prior to Résif, French seismic waveform data was distributed from five institutions:

- 30 • CEA (Commissariat à l'énergie atomique et aux énergies alternatives), commissioned by the
31 French ministry of the Interior to establish earthquake alerts towards the national authorities,
32 operated (and still operates) a dedicated seismic network with high availability data transmission
33 (Massinon and Plantet, 1976; Résif, 2018). Data from this network was available through direct
34 cooperation with CEA.
- 35 • ISTERre (then: LGIT) in Grenoble distributed data from the national strong motion network
36 (RAP – Résif, 1995) via the RAP National Data Centre (RAP-NDC – Pequegnat et al., 2008).
37 ISTERre also ensured national distribution of data from temporary seismic deployments using the
38 national instrument pool, but most research laboratories also ran their own, smaller instrument
39 pools.
- 40 • IPGP (Institut de physique du globe de Paris) distributed data from the GEOSCOPE network
41 (IPGP and EOST, 1982), via the GEOSCOPE data center, with a historical strong role of

42 GEOSCOPE in the promotion of open data distribution at a worldwide level (Roult et al., 1999).
43 IGP also operates three seismic networks in the Antilles (IPGP, 2008a,d,b), and one in
44 Reunion Island (IPGP, 2008c), and those datasets were at the time planned for integration in
45 the IGP data distribution.

46 • Géoazur and EOST (École et Observatoire des Sciences de la Terre): the metropolitan French
47 broadband and short period data were loosely integrated in a national cooperation related to
48 long term Earth Observations, but each regional network was operated independently, with no
49 joint instrumentation policy, and distributed mainly by these two institutions.

50 The principle of transparent access to data through the common international distribution tool
51 netdc (Casey, 1999) was agreed as a common goal, through an additional transparent layer to a
52 virtual data center, FOSFORE (Shapiro et al., 2008). Practically, the national French Broadband
53 network was underperforming, for a lack of sufficient station coverage and homogeneous instrumen-
54 tation policy. Also, the difference in technical implementation of data and metadata distribution, data
55 holdings and data center availability between the five locations meant that the federated approach in
56 practice needed to evolve to a new system.

57 When Résif was created as a concept, it moved under the direct supervision of the French
58 Ministry of higher education and research, and integrated 18 organizations into a national consortium,
59 which was signed in 2011. As a first success of this coordination effort, Résif obtained the funding
60 (9.3 M€, with approximately the same amount committed by the partners in terms of staff time) for a
61 project with four main components: the construction of a national broadband network; the extension
62 and renewal of the GNSS network and of the different mobile pools; the creation of a national Résif
63 information system. The constraints from labor policies and civil servants meant that the permanent
64 staff would transfer without relocation from the old system to the new, with additional temporary staff
65 provided to accompany the changes.

66 The result of this effort is presented in this article. The architecture of the Résif information
67 system (French: système d'information, hereinafter Résif-SI, re3data.org, 2016) is fully distributed,
68 matching the distributed architecture of the French seismic networks, which relies on CEA and on

69 regional operators (mainly Observatories hosted at universities, and with strong CNRS involvement)
70 for production and validation of data and metadata. These operators, through the so called "A-
71 nodes", have full responsibility of pushing data and metadata into the data distribution center, via
72 an automated procedure which allows for regular updates and automatic replacement in the data
73 bases, as instruments are replaced, errors (timing, instrument responses, . . .) are corrected, and
74 data gaps are filled. The system is at the same time fully integrated, with a national management
75 team (Executive Committee), a technical board which meets monthly and which adopts joint technical
76 solutions, a single national data distribution facility (hosted by the University of Grenoble), and a
77 coordinated participation in international and European and collaborative instances such as FDSN
78 (<https://www.fdsn.org>) and ORFEUS (<https://www.orfeus-eu.org>). In particular, the national data
79 distribution facility is part of the EIDA federation, a service within ORFEUS which gives federated
80 access to data from European data centers. EIDA is described in Strollo et al. (2021).

81 Résif-SI has provided, within the constraints given, remarkable results, with a stable and high
82 quality data distribution and a full integration into the international collaborations. The data holdings
83 and download statistics are on par with other large data centers in Europe.

84 We believe that the Résif-SI implementation can be a model for countries which face the challenge
85 of regrouping observation networks and data distribution, and particularly those countries where
86 the need for regional and institutional visibility and independence remain strong, but where each
87 institution alone does not have critical mass for running a high availability data center.

88 The paper is structured as follows: we start describing the overall [Organizational structure of](#)
89 [Résif-SI](#), for then discussing the [Data management: Résif-DC](#) system and in particular the technical
90 choices behind the Résif-SI data center and how they have been functional to integrate Résif-SI
91 into the international data exchange systems. The section "[New developments](#)" illustrates four
92 areas (Large-N experiments, Marine data, Building and infrastructural monitoring, and a StationXML
93 metadata editor) for which Résif-SI is at the forefront of current development in seismological data
94 management. We conclude by presenting the upcoming challenges and the strategy to meet them.
95 An Appendix ("[Appendix: Data access details and examples](#)") is provided with details and practical
96 examples on the different ways to access Résif-SI data.

97 **Organizational structure of Résif-SI**

98 Résif-SI archives and publishes seismological data from eleven permanent networks and approxi-
99 mately seventy temporary networks (Fig. 1).

100 Seismological data and metadata are collected and validated by six centers run by Résif partners
101 which are called “A-nodes” (Fig. 2), which are also responsible to secure a copy of the data for
102 two years. The data and metadata are then transmitted to the Résif seismological Data Centre
103 (Résif-DC –also called the “B-node”) hosted by the University of Grenoble Alpes for archiving (storage
104 and remote archive) and distribution (services, portal). Résif-DC is designed and operated by a
105 technical team from ISTERre and from the Observatoire des Sciences de l’Univers (OSUG). Table 1
106 illustrates the typology of data available from each of the 6 operational A-nodes, while Fig. 3 shows
107 the proportion of data archived at Résif-DC by type: accelerometric, velocimetric and other. A 7th
108 A-node, MARINE, operated by IPGP (Institut de physique du globe de Paris) and OCA (Observatoire
109 de la Côte d’Azur), is in its implementation phase, and will collect and validate the data from ocean
110 bottom seismometers (OBS).

111 Résif-DC, the national Résif data distribution center, is one of nineteen global centers distributing
112 data and metadata using formats and protocols which comply with International Federation of Digital
113 Seismograph Networks (FDSN –<https://www.fdsn.org>) standards. It is also one of the twelve nodes of
114 the European Integrated Data Archive (EIDA –<http://www.orfeus-eu.org/data/eida/>).

115 From the very beginning of Résif-SI in 2012, it was defined that the quality of data and metadata
116 and their rapid and regular availability was the responsibility of the A-nodes, while the quality and
117 continuity of data services depended on the national data center. This principle is still at the core of
118 the organization of Résif-SI, but all of Résif-SI jointly takes action to continuously improve data and
119 metadata quality.

120 **Data and metadata workflows**

121 A-nodes take care of data validation and production of associated metadata for the observatories and
122 instruments they are responsible of, according to their own workflows and objectives of completeness
123 and quality.

124 Data integration is initiated, managed and controlled by the A-nodes which can add, modify or
125 replace data and metadata (the removal being under the control of Résif-DC) at their own discretion
126 and according to their rhythm, in an autonomous way. To this purpose, Résif-DC maintains the
127 dedicated tool `ResifDataTransfer` (Volcke et al., 2013). An important and explicit rule of Résif-SI
128 is that Résif-DC does not in any way modify the data or metadata provided by A-nodes. The data and
129 metadata completion/validation process at A-Nodes is iterative. Since A-Nodes are responsible for
130 retaining all raw data collected by any means (including data in proprietary formats) for at least two
131 years, all or part of the processing phases can be easily reversed if necessary.

132 The integration protocol specifies:

- 133 • the *formats* allowed as input: *miniSEED* (Ahern et al., 2012) or *PH5* (Hess et al., 2018) for
134 data, *StationXML 1.1* (FDSN, 2019) for metadata;
- 135 • the transfer modalities –based on *rsync* (Tridgell and Mackerras, 1996) *push*– and acknowledg-
136 ment –based on *rsync get* or a *web service*;
- 137 • the *checks* carried out before the metadata and the data are integrated into the database and
138 the data archiving;
- 139 • the integration modalities: a product submission is a *transaction*;
- 140 • the *structure of the integration report* to the partner.

141 The integration controls are implemented at the data center level and have been jointly specified.

142 From the point of view of A-nodes, any product submission (simple or complex) is a transaction,
143 whose characteristics are preserved and can be retrieved by a web service. The integration of data
144 and metadata is asynchronous and the consistency between them is re-evaluated at the end of each

145 transaction. The submission of not-yet-described or incompletely described data is allowed, in order
146 to secure them as soon as possible and according to the needs of the A-nodes, but Résif-DC data
147 services do not deliver any data whose metadata is not available. This workflow is illustrated in
148 figure 4.

149 From the point of view of Résif-DC, an integration of data or metadata is a succession of complex
150 operations carried out by several independent workers (Fig. 5) that communicate with each other via
151 an AMQP (Advanced Message Queuing Protocol) system (ISO/IEC 19464, 2014). Four databases
152 are populated at this stage: (1) inventories of networks, stations, channels and responses; (2)
153 waveform inventories (PostgreSQL databases, PostgreSQL Global Development Group, 1996); (3)
154 metrics and data quality information (MongoDB - EIDA WFCatalog database, MongoDB, Inc., 2009;
155 Trani et al., 2017); (4) elements for monitoring the integration operations (PostgreSQL). Consistency
156 of information within and between databases (1), (2) and (3) is guaranteed by construction. Data is
157 archived in a SDS structure (GFZ and gempa GmbH, 2008, if miniSEED data) or as HDF5 record
158 (Folk et al., 2011, if PH5 data).

159 Résif-DC also computes on the fly, with each integration, some useful side-products. For instance,
160 Power Spectral Density plots (PSD, McNamara, 2004) for the data manager and the end user to
161 check the quality of the signal generated with PQLX (McNamara and Boaz, 2010), but also availability
162 plots generated with the help of a SEED data indexer (`seedtree5`, Volcke et al., 2012).

163 Résif-DC exposes two web services for A-Nodes to manage their data and transactions:

- 164 • <https://ws.resif.fr/resifws/transaction/1>: to retrieve transaction records, for example to ana-
165 lyze and correct rejected data or metadata;
- 166 • <https://ws.resif.fr/resifws/orphanfile/1>: to identify orphan data (without metadata) in the
167 Résif archive for corrective action.

168 Other operations have to be requested through a help desk because either they are rare, or there
169 is no safe way to automate (e.g., data removal).

170 **Metadata enhancement**

171 Résif-SI has progressively improved the quality and consistency of metadata by establishing a
172 controlled vocabulary and recommendations for standardizing the content of text fields in metadata.
173 Initially, the metadata of A-nodes were managed in local databases and formatted using the open
174 source PDCC tool by IRIS (Casey, 2016) or in-house tools. Each A-node has its specificity regarding
175 metadata. In particular, SISMOB and RAP have to deal with a large diversity of instruments and very
176 complex metadata. With this context, in-house tools to manage the metadata was mandatory. Indeed,
177 all the constraints on the metadata for Résif-SI networks could not be included into any standard
178 software (e.g., SeisComp, GFZ and GEMPA GmbH, 2008) at the time.

179 Before 2019, metadata were edited and submitted in SEED dataless format (Ahern et al., 2012)
180 and Résif-DC ingested the metadata and delivered it in the normalized StationXML format (FDSN,
181 2019). Today, all A-nodes produce StationXML metadata and most of Résif-SI is moving toward a
182 common toolbox and database for metadata edition. Section “Metadata editing made easy: YASMINE”
183 gives additional information on this project, which has high priority in Résif-SI, since StationXML
184 gives visibility and full acknowledgment of all dataset contributors –from station to data distribution
185 (see section “Data fairness”). This normalized information is automatically exported to the Résif-SI
186 web portal and to other services. Moreover, StationXML allows referencing to persistent network
187 identifiers to facilitate citation, and includes new fields necessary, for example, for OBS data (see
188 section “Marine data”).

189 **Data collection and completion**

190 The management of data flows is specific to each A-node. However, Résif-SI offers its own tools that
191 are generic enough to be deployed outside their initial context. An example is MORUMOTTO (MODular
192 aRchive bUilder from Multiple Origin Temporal Traces & Other stuff – Geber, 2019), developed by
193 IPGP for A-node VOLCANO. MORUMOTTO is used by network operators to quality control a data
194 archive, and in particular to detect and correct data gaps and overlaps. Data is regularly fetched from
195 a pool of different sources.

196 **Data validation**

197 Data validation procedures are specific to each A-node.

198 Data from networks FR and RA is qualified "M" ("data center modified, time-series values have
199 not been changed" –see [Ahern et al., 2012](#), page 108), and is made available 3 to 5 days after
200 collection by A-nodes, after having undergone the following checks:

- 201 • data has been completed as much as possible;
- 202 • there is no overlap, even at the sample level;
- 203 • instrument response validity epochs do not overlap;
- 204 • the power spectral density of the signal is in accordance with what is expected for each
205 particular station.

206 Data from networks G, GL, MQ, WI and PF is made available with the quality label "Q" ("quality
207 controlled data, some processes have been applied to the data" –see [Ahern et al., 2012](#), page 108),
208 and undergoes additional checks, namely:

- 209 • waveform modeling of teleseismic earthquakes of magnitude larger than 6 through the SCARDEC
210 method ([Vallée et al., 2011](#)) shows a good agreement between the observed and predicted
211 signals;
- 212 • for multi-parametric stations, the acceleration-converted seismometer and accelerometer
213 waveforms are identical for a selection of suitable earthquakes.

214 Because of the complexity of these additional operations, validated data from these networks is
215 available 6 to 12 months after its collection.

216 Availability of validated data from SISMOB temporary networks depends on the specificity of each
217 experiment and generally varies between 1 and 6 months (note that SISMOB data might be under
218 embargo for up to three years).

219 Table 2 summarizes the latency for the production of validated data by A-nodes, and their data
220 quality label. Note that, for networks collected in real-time (see section “Real-time: SeedLink”), raw
221 data is immediately available and eventually replaced with validated data.

222 **Quality of Service**

223 The Résif-SI organization does its best to implement a set of high quality services with high operational
224 performance, i.e. as continuously available and robust as possible. To achieve this, we developed a
225 culture of quality of service, based on a set of best practices, among which:

- 226 • a single help desk, accessible by email at resif-dc@univ-grenoble-alpes.fr. Users are typically
227 scientists searching for data, data or metadata producers needing help to access the services,
228 or network managers needing assistance.
- 229 • a clear presentation and documentation of our services through the website <https://ws.resif.fr>;
- 230 • an automated system which tests thoroughly our services and infrastructure, and monitors the
231 overall activity in order to alert the relevant personnel of Résif-SI as early as possible in case
232 any potential issues or anomalies are detected.;
- 233 • an RSS feed to publish news or scheduled downtime events.

234 The website <http://seismology.resif.fr> is also the place to browse all the available metadata, to get
235 information about citations, data and metadata access. It is also used for announcements regarding
236 operational news of Résif-DC, like maintenance down times or new services availability.

237 The monitoring system checks all the Résif-DC services, mimics user access for testing and
238 alerts the operational team if necessary. It also monitors lower levels of the infrastructure of Résif-DC.
239 The gathered information is used in reports to publish the overall availability of the data center.

240 Measuring and publishing the availability of our services is an ongoing project that will be
241 completed in 2021. Service level agreements (SLA) will be defined by the Executive Committee
242 for our main services and the data center’s yearly report will show the metrics. Table 3 provides a

243 rough estimate of service availability at Résif-DC in 2020, based on network reachability. Based on
244 feedback from the EIDA User Advisory Group, EIDA is presently working on a periodic test bench of
245 all the nodes based on random requests of data and metadata which will be more representative of
246 the services availability.

247 **Data management: Résif-DC**

248 The construction of Résif-SI is patterned on a Data Management Plan (e.g., [Michener, 2015](#)) describ-
249 ing the responsibilities of each partner, how long data should be stored, how metadata is managed,
250 how data is accessed. Résif-DC has a central role to play in this plan: it ensures that data is securely
251 stored and easily accessible.

252 **Data archive description**

253 Résif-DC stores the data in miniSEED (using the SDS structure, [Ahern et al., 2012](#); [GFZ and gempa](#)
254 [GmbH, 2008](#)) or PH5 format ([Hess et al., 2018](#)). The latter is designed by IRIS PASSCAL (Portable
255 Array Seismic Studies of the Continental Lithosphere), is based on HDF5 ([Folk et al., 2011](#)) and
256 widely used by the temporary experiments using dense arrays.

257 Résif-DC holds miniSEED data from 1986, all accessible through the FDSN `dataselect` web
258 services. Fig. 6 shows the volume of miniSEED and PH5 data, by network, held for each year as of
259 January 2021. We account for 68% of data from permanent networks versus 32% from temporary
260 experiments.

261 The increasing amount of data stored came from two major factors: first, the deployment of
262 approximately 150 new permanent broadband stations; second the increasing number of temporary
263 experiments associated with SISMOB. We anticipate a strong data growth in coming years due to
264 increasing number of experiments using dense arrays and the foreseen integration of data from
265 Distributed Acoustic Sensors, but the numbers are still difficult to quantify, as discussed in [Quinteros](#)
266 [et al. \(2021\)](#).

267 The data management description has been used to design an efficient central storage at Résif-
268 DC. Data is kept on two classes of storage, ensuring the same data security but with different levels
269 of performance. While seldom requested data is still quickly accessible, it is kept on a less powerful
270 and less expensive media. This storage tiering is one of the many features of our storage provider
271 SUMMER and is totally transparent for the data center systems, thanks to a storage abstraction layer
272 (`autofs`). Once a year, the data that can be transferred to less powerful storage media is selected
273 according to the seismic network's data management plan and manually moved.

274 **Securing the data**

275 When it comes to store the data securely, one has to identify and address several types of risks.
276 Résif-DC is concerned about:

- 277 1. hardware failures;
- 278 2. unintentional or malicious alteration/deletion of the data by an operator;
- 279 3. local disaster destroying physically the archive located in Grenoble (flood, earthquake, ...).

280 Risks [1](#) and [2](#) are addressed *de facto* by Résif-DC through the choice of storage provider
281 SUMMER ([Université Grenoble Alpes, 2016](#)), operated by the university of Grenoble Alpes, which
282 secures the data in three different data centers and offers a daily snapshot service. By using the
283 university storage service, we also ensure that all the data is kept on an academic and publicly funded
284 platform .

285 The risk [3](#) of a local destruction has been addressed by copying twice a year a snapshot of our
286 archive to tapes stored at a distant place. At this juncture, we have secured a collaboration with the
287 the Computing Centre of the National Institute of Nuclear Physics and Particle Physics (CC-IN2P3,
288 [IN2P3/CNRS, 1976](#)) located in Lyon, to serve as this remote site. Although the recovery plan is not
289 fully formalized, we guarantee that the data hosted at Résif-DC will survive the mentioned risks, as
290 long as it has been stored for at least six months. An additional copy for recent data is ensured by
291 A-nodes, which are committed to keeping a secure copy of their datasets for a duration of two years.

292 Altering or destroying the data hosted by Résif-DC would require an operator or an attacker to
293 get access to high privileges on three independent infrastructures, namely:

294 • Write access to the Résif-DC infrastructure would allow to delete live data in the archive. The
295 data can in that case be retrieved from SUMMER snapshots or backups;

296 • Write access to the SUMMER infrastructure would allow to destroy backups, snapshots and
297 live data. The data and metadata can in that case be recovered from a distant copy made twice
298 a year at CC-IN2P3;

299 • Write access to the CC-IN2P3 infrastructure would allow to destroy the distant copy.

300 Résif-SI considers that the scenario of the combined access to these three infrastructures is unlikely
301 to happen.

302 **Data distribution services**

303 Résif-SI data distribution is based on a small set of elementary and very robust data services, which
304 can be divided into: real-time data access, asynchronous data access, and web services. These
305 base services constitute the building blocks for higher-level data services and products provided by
306 Résif-SI, or by European and international partners. Résif-SI is fully integrated into EIDA through its
307 elementary data services. EIDA offers a wide set of additional services and tools, including smart
308 clients, federated archive access and web interfaces. A full description of EIDA can be found in [Strollo
309 et al. \(2021\)](#).

310 The Résif-SI web portal (<http://seismology.resif.fr>) provides information about Résif-SI, documen-
311 tation on how to use the elementary and additional services, licensing, etc. It also provides URL
312 builders for the web services and dynamic search options for browsing the seismic networks, station
313 metadata, data availability, data quality, etc.

314 Résif-DC has implemented most of its data models and workers in order to preserve the specificity
315 and accuracy of the metadata produced by A-nodes. In particular, care was taken to be able to
316 export the rich Résif-SI metadata. This *a priori* expensive solution proved to be profitable in the long

317 term, as it enabled Résif-SI to integrate from early on the description of non-seismological channels
318 (polynomial responses), and to manage all station and channel comments, as well as OBS metadata.
319 In this way, additional services are effective across all the data types managed by Résif-SI.

320 The two initial FDSN web services `station` and `dataselect` (FDSN, 2013) were put into
321 production in 2013. Their implementation is based on IRIS WebServiceShell (IRIS, 2016), which
322 we interfaced (via Python) with our databases and archives. All the other web services have been
323 implemented in Python and ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015)
324 within the FLASK framework (Ronacher, 2010) and their codes are fully available on public code
325 repositories (see [Data and Resources](#)).

326 The basic Résif-DC data services are:

- 327 • A SeedLink server (GFZ and gempa GmbH, 2008), accessible at `rtserve.resif.fr` (stan-
328 dard port 18000), for real-time access to miniSEED data. This service, based on IRIS
329 `ringserver` (Trabant, 2011), provides a single access point to all the real-time data streams
330 of the A-nodes for which metadata is publicly available;
- 331 • The standard `fdsn-dataselect` web service (FDSN, 2013) to access validated or near
332 real-time data –public or restricted– for all of the miniSEED archive.
- 333 • The FDSN-compliant `ph5-dataselect` web service (Résif, 2019a), which serves large-N data
334 (stored in PH5 archives) as miniSEED or SAC format on the fly (useful for small subsets of the
335 PH5 datasets);
- 336 • The standard `fdsn-station` web service (FDSN, 2013) to retrieve station metadata;
- 337 • The standard `fdsn-availability` web service (FDSN, 2013) to interrogate the miniSEED
338 data inventory;
- 339 • The FDSN-compliant `ph5-availability` web service (Résif, 2019a), to interrogate the PH5
340 data inventory;
- 341 • The `eidaws-wfcatalog` web service (Trani et al., 2017) to retrieve data availability as well as
342 other metrics (e.g., gaps, overlaps, RMS);

343 • The standard `fdsn-event` web service (FDSN, 2013), operated by BCSF-RéNaSS (Bureau
344 Central Sismologique Français, French Central Seismology Bureau – Réseau National de
345 Surveillance Sismique, National Seismic Monitoring Network), which gives access to event
346 parameters located by BCSF-RéNaSS (Masson et al., 2021).

347 Résif-DC also allows the retrieval of a complete dataset, by network, using `rsync` protocol. The
348 user is granted read-only access to the part of the archive exposing the dataset. This is particularly
349 useful to retrieve entire datasets in PH5 format. Fig. 7 shows the amount of waveform data served by
350 `fdsn-dataselect` and SeedLink.

351 Built on top of these base services, the following additional web services are available from
352 Résif-DC:

353 • The `timeseries` and `timeseriesplot` web services (Résif, 2019a) to obtain pre-processed
354 waveforms or plots;

355 • The `resp`, `sacpz` and `evalresp` web services (Résif, 2019a) to obtain or plot instrumental
356 responses;

357 A complete catalog of the Résif-DC web services is available on <https://ws.resif.fr>. More details
358 and usage examples are given in the “Appendix: Data access details and examples”.

359 Data fairness

360 Data fairness refers to the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles for
361 scientific data management and stewardship (Wilkinson et al., 2016). We here discuss how far we
362 are complying to those best practices.

363 The **findability** of our data and metadata is ensured by the international standard web service
364 `station` (FDSN, 2013), accessible via any web browser, http client, and the community’s most
365 popular tools like ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015), EIDA’s
366 `fdsnws_scripts` (Heinloo, 2018) or the Web Service Fetch scripts (Hutko, 2013). In addition, most

367 of the Résif datasets have DOIs (Digital Object Identifiers) associated with their seismic network,
368 according to the standard procedure approved in 2014 by the FDSN (Evans et al., 2015). Within
369 Résif-SI, much effort has been dedicated to manage the quality of the published metadata, as
370 stated in section “[Metadata enhancement](#)”. The seismological metadata is very descriptive, allowing
371 scientists to identify precisely the data needed. In addition, the `station` web service provides a
372 rich set of options to make selections in the metadata catalog, such as time period, geographical
373 region, or update time. Findability of Résif-SI data is also guaranteed by the ORFEUS European
374 metadata catalog (ORFEUS, 2020a), implemented by EIDA and maintained by ETH-Zürich, which
375 enables users to fetch metadata from all the EIDA’s data centers in one request, at one entry point.
376 Finally, within FDSN, a large effort in establishing data routing tables has been achieved. Now, each
377 seismological network has a well known reference data center, and this information is made publicly
378 available by each data center, and can be consulted for each network on the FDSN’s network details
379 page (FDSN, 2020).

380 The **accessibility** of the data is ensured by the international standard web service `dataselect`
381 (FDSN, 2013) that gives access to all Résif’s archive, except for data in PH5 format, which has its
382 own non-standard web service `ph5-dataselect` (Résif, 2019a). Authentication is also provided for
383 restricted data, with local methods or with EIDA’s authentication system.

384 The **interoperability** is ensured inside Résif by adopting a common vocabulary for StationXML
385 metadata, which makes it possible to expose on the the Résif portal significant parts of the StationXML
386 content coherently across all the operators. For the services to the user, interoperability is first and
387 foremost ensured by web services in strict adherence to FDSN standards when available. This
388 is the key for universal data and metadata access through interoperable additional layers across
389 many data centers offering the same services. At a service level, EIDA offers such interoperable
390 services (e.g. `WebDC3`, ORFEUS 2020b, which gives transparent access to the data of all the
391 EIDA nodes) while many users develop workflows based on the FDSN web services interfaces (e.g.,
392 [Zaccarelli et al., 2019](#); [MacCarthy et al., 2020](#)). Cross-disciplinary interoperability requires challenging
393 standardization of metadata vocabulary, both for data and services. The interoperability between the
394 different Résif data types (seismology, GNSS, gravimetry) is delegated to the EPOS infrastructure,

395 into which Résif is fully integrated. National technical discussions are undertaken across different
396 EPOS activities in France, rather than strictly within Résif perimeter, and the Résif seismology
397 and GNSS community actively contribute to EPOS. The EPOS cross-disciplinary interoperability is
398 achieved through a dedicated layer that connects different services through a metadata catalog which
399 uses standardized vocabulary to describe the services. Additionally, the scientific users of Résif data
400 are already observed to mix data and data products from different domains (e.g., seismic waveforms
401 with environmental data, such as weather condition or ocean wave activity) within smart clients that
402 they create, or by mixing data from local files downloaded from different sources with direct download
403 of waveform data. The main condition for the success of these applications is meticulous application
404 of domain standards and the effective data and metadata accessibility.

405 The data provided by Résif-SI is also made **reusable** in respect of the FAIR guidelines. The rule
406 of Résif-SI is to distribute all open data under the Creative Common Attribution 4.0 license ([Creative
407 Commons, 2013](#)) CC4.0:BY, coherent with French Law. StationXML specification (currently at version
408 1.1 –[FDSN 2019](#)) does not provide a field to indicate the data license, but we made this information
409 easily accessible by systematically exposing the DOI of a network via the <Identifier> tag in
410 StationXML metadata. As an example, metadata from RAP network (network code “RA”, accessible
411 from <https://ws.resif.fr/fdsnws/station/1/query?network=RA&level=network>) contains the following
412 tags:

```
413 _____  
414 <Description>RESIF-RAP Accelerometric permanent network</  
415 Description>  
416 <Identifier type="DOI">10.15778/RESIF.RA</Identifier>  
417 _____
```

418 License information is provided in the DataCite XML document ([DataCite, 2019](#)) of each network
419 via the <rightsList> tag. For instance, the “RA” network DataCite XML ([https://data.datacite.org/
420 application/vnd.datacite.datacite+xml/10.15778/RESIF.RA](https://data.datacite.org/application/vnd.datacite.datacite+xml/10.15778/RESIF.RA)) contains the following tags:

```
421 _____  
422 <rightsList>
```

```
423     <rights rightsURI="info:eu-repo/semantics/openAccess">Open
424         Access</rights>
425     <rights rightsURI="https://creativecommons.org/licenses/by
426         /4.0">Creative Commons By 4.0 Universal</rights>
427 </rightsList>
428
```

429 Citation instructions are available on the DataCite page (e.g., <https://search.datacite.org/works/10.15778/RESIF.RA>) and dynamically presented by several portals (e.g., FDSN network details, [FDSN 2020](#), or Résif seismic data portal, [re3data.org 2016](#)). Résif-DC provides a list of citation instructions in the portal's citation page ([Résif-SI, 2020](#)). We strongly encourage scientists to cite data producers and distributors in their papers.

434 The use of licenses for seismic waveform data is presently only at its beginning, with licenses
435 being put in place for many seismic networks in Europe. A strong motivation are the citations (see
436 above), but also to handle liability issues in the case of erroneous data or metadata. Résif-SI is
437 actively working, within the framework of a cooperation between ORFEUS and IRIS, to promote good
438 citation usage to scientific users of waveform data, and to inform publishers of these progresses, as
439 the introduction of licenses engages the liability of the journals and of the publishing scientists, even
440 though this liability is not yet reinforced.

441 An accompanying issue is that data centers who hold copies of the datasets need to expose
442 the license information in a proper way. For this to be practically possible, Résif-SI has, with the
443 agreement of the involved organizations, stopped all data copies to other data centers which were
444 previously put in place, and asked for the old copies to be deleted. The only exception is GEOSCOPE
445 data archived at the IRIS Data Management Center, due to a strong user base of GEOSCOPE data
446 through this data center. Overall, the seismological community worldwide still needs to efficiently
447 communicate license information to users, but if the citations (see above) are properly done, the
448 license CC4.0:BY is respected.

449 Citation through the use of DOIs is only meaningful if the associated metadata is sufficiently rich

450 to include all the parties that contribute to the data production, management and distribution. Résif-SI
451 uses the `contributor` field of DataCite metadata ([DataCite, 2019](#)) to acknowledge different types
452 of contribution and roles within each seismic network, and more specifically the `contributorTypes`
453 *DataManager* (Résif-SI), *Distributor* (Résif Data Center), *HostingInstitution* (Université Grenoble
454 Alpes), *DataCurator* (the relevant A-Node), *DataCollector* (organisations that operate the seismic
455 network) and *Sponsor* (funding sources).

456 **From logging to usage statistics**

457 Data usage statistics are important information for the different persons associated with data produc-
458 tion, management and distribution. For example, the Résif-DC team needs to have real-time usage
459 statistics, correlated with IT system metrics in order to catch anomalies or anticipate requests growth
460 while data producers, project principal investigators and funding agencies need compiled usage
461 statistics in a larger view and for larger time span, such as overall statistics (for example number of
462 requests, users, shipped volumes, countries of request) or network relevant information (for example
463 network level download statistics, the most accessed stations, countries of requests).

464 In order to satisfy all the needs for statistics we built a system that concentrates the information,
465 builds statistics and gives access to dashboards (Fig. 8).

466 The constraints come from the variety of data access methods, each one having their own log
467 formats which are not always compatible with real-time processing. Another difficulty is to analyze
468 the data served through web services. It's easy to capture the quantity of data shipped by the web
469 server, but it's not possible to know what it is made of (which network, station, location or channel).
470 Consequently, we have to analyze each request and evaluate the typology of the response. All
471 the gathered information is stored in a PostgreSQL database, aggregated and anonymized. The
472 anonymization concerns the client's IP address and consists of hashing it as soon as it enters the
473 databases. Then, when the requests are aggregated, we only keep HyperLogLog objects ([Flajolet
474 et al., 2007](#)) in order to compute the cardinality of the clients. Therefore, our databases respect the
475 European legislation (General Data Protection Regulation, GDPR), as there is no way to retrieve an

476 IP address nor any kind of personal data.

477 The statistics can be accessed through several means: we provide a web service for end users
478 (see [Appendix: Data access details and examples](#)), and interactive dashboards for internal usage
479 (e.g., analyzing operational events in real time, usage and performance evaluation –Fig. 9). Fig. 10 is
480 an example of rendering statistics in a geographical map. Recently, we published a system allowing au-
481 tomatic creation of plots and usage statistics in form of a report (`resif-delivery-stats-plotter`,
482 [Bollard 2021](#)). This program is aimed at network managers or principal investigators seeking informa-
483 tion on data and metadata usage.

484 **New developments**

485 **Large-N data management**

486 Large-N data are produced by temporary experiments deploying small, self-contained seismometers
487 (the so called “nodes”) in dense arrays, with high sampling rate (i.e., above 100 samples per second,
488 e.g., [Brenguier et al., 2016](#); [Dougherty et al., 2019](#); [Gimbert et al., 2020](#)). The raw data generated by
489 those experiments is approximately 1 GB per day per node, and an experiment can use hundreds
490 of nodes for several months. In addition, the data management needs specific workflows and data
491 format treatments.

492 In order to host and distribute this new kind of data, the SISMOB A-node developed a specific
493 workflow which produces validated data in PH5 ([Hess et al., 2018](#)), a data format developed at
494 IRIS by PASSCAL and commonly used for Large-N data. In order to ingest this new data format,
495 Résif-DC also adapted the integration mechanisms and storage repositories, since PH5 data cannot
496 be indexed nor referenced in the same way as miniSEED data. Furthermore, in order to host the
497 foreseen volumes, a dedicated archive storage has been setup.

498 The metadata, on the other hand, is created in classical StationXML format, as other miniSEED
499 data, and submitted at Résif-DC using the standard procedures described above.

500 The PH5 archive is served to end users by two means:

- 501 • `rsync` access for a restricted set of users needing to download the entire dataset;
- 502 • the `ph5-dataselect` web service ([Résif, 2019c](#)) allowing the selection of a subset of the data
- 503 in miniSEED format with the same options and syntax as the FDSN `dataselect` web service
- 504 specification

505 Résif-DC also provides the `ph5-availability` web service ([Résif, 2019b](#)) for the users to get

506 the time span of the available data.

507 **Marine data**

508 Marine data collected on autonomous sensors such as ocean-bottom seismometers follow specific

509 guidelines developed and distributed through the FDSN “Mobile instrumentation” working group and

510 the European Union projects EPOS/ORFEUS, ENVRI-FAIR and SERA. These guidelines include:

511 data quality labels indicating whether the data were corrected or not for the instrument clock drift;

512 standards for post-implementing leap seconds; component code standards for horizontal channels

513 that are not geographically oriented; orientation/dip standards for pressure channels, and station

514 naming rules for repeated deployments at the same station (see [Clinton et al. 2018](#), Appendix

515 B, for the latest published version and [Crawford 2019](#) for the latest proposed version). These

516 guidelines have been adopted already by some OBS parks and EIDA data centers. The AlpArray

517 OBS component has been archived following these guidelines.

518 Résif has been working since 2017 on the integration of data from French Ocean Bottom

519 Seismometer (OBS) parks. This integration will be accomplished very soon by the commissioning

520 of the dedicated MARINE A-node. Within this frame, a system for for creating FDSN-standard

521 data and metadata for ocean bottom seismometers using standardized, easy-to-read information

522 files is currently in development ([Crawford et al., 2019](#)). A specific visual quality control is also in

523 development ([Goubier and Crawford, 2021](#)) to allow the instrument providers and scientific users to

524 verify instrument responses, noise levels and time corrections.

525 To date, Résif-DC distributes data from two temporary campaigns:

- 526 • RHUM-RUM ([Barruol et al., 2017](#)): 57 OBS for 2 months of data, 3 components at 50 Hz
- 527 • AlpArray ([AlpArray Seismic Network, 2015](#)): 8 OBS for 8 months of data, 3 components at 50
- 528 Hz.

529 Two other datasets are in the process of validation and integration:

- 530 • SISMANTILLES ([Laigle et al., 2007](#)): 20 OBS for 4 months, 3 components at 50 Hz.
- 531 • EMSO-MOMAR ([IPGP, 2007](#)): 5 OBS for 10 years 3 components at 50 Hz

532 OBSs cannot obtain precise GPS timing during their deployment. The instruments are equipped

533 with a very accurate clock to minimize the problem, however this clock drifts on the order of 1-2

534 seconds/year (e.g., [Loviknes et al., 2020](#)). The standard protocol for OBS time correction is to

535 synchronize an OBS with a GPS signal immediately before deployment and after recovery. The

536 measured timing deviation is assumed to have accumulated linearly over the deployment interval,

537 therefore the applied correction for time drift is linear. This assumption has been checked by [Hable](#)

538 [et al. \(2018\)](#) for the instruments used during the RHUM-RUM Experiment (French and German pools)

539 using ambient noise correlations.

540 Some users, however, prefer to have the data "unmodified", even if this means that there are

541 timing errors of the order of a second. That's why Résif-DC delivers both types of data, giving

542 quality label "Q" (Quality Controlled Data, some processes have been applied to the data –see [Ahern](#)

543 [et al., 2012](#), page 108) to those that are corrected and "D" (the state of quality control of the data is

544 indeterminate –see [Ahern et al., 2012](#), page 108) to those that are not (see "[Specificity of OBS data](#)

545 [and metadata](#)" in the Appendix).

546 **Organizing and improving archival of data from building and infrastructure monitor-**

547 **ing**

548 Since 2010, Résif has been involved in specifying metadata and solutions for building and infrastruc-

549 ture monitoring ([Clinton et al., 2018](#)). More and more high quality seismic sensors are deployed across

550 structures, with the same technical issues than for classical seismological networks (in large numbers,
551 with continuous recordings) and using therefore seismological standards. The StationXML metadata
552 specification ([FDSN, 2019](#)) can be used to include specific information on the structure, allowing
553 the engineering seismology community to integrate infrastructure monitoring data into standardized
554 practices. The building description, which is key for engineering purposes, can be included at two
555 levels: (1) following the European Macroseismic Scale typology of buildings, based on the material
556 of construction ([Grünthal, 1998](#)), and (2) through a full description of the building characteristics
557 according to the GEM Building taxonomy ([Brzev et al., 2013](#)).

558 To date, Résif-DC distributes seismic data from five instrumented buildings:

- 559 • City-hall of Grenoble, France ([Michel et al., 2009](#); [Guéguen et al., 2020](#)): 6 three-component
560 (3C) accelerometric sensors since 2004 (stations OGH1 to OGH6), and one additional 3C sensor
561 at intermediate height (station OGH7) and one weather station at the top since 2019 (station:
562 OGH8)
- 563 • Ophite Tower in Lourdes, France ([Michel and Guéguen, 2018](#)): 3 3C accelerometric sensors
564 and 15 1C accelerometric sensors distributed over the building height since 2008 (station:
565 PYT0) plus one temperature sensor at the top.
- 566 • Prefecture building in Nice, France ([Lorenzo et al., 2018](#)): 2 3C accelerometric sensors and
567 18 1C accelerometric sensors distributed over the building height since 2010 (station NCAD),
568 including one free-field station.
- 569 • Basse-Pointe College in Martinique Island: 2 3C accelerometric sensors and 18 1C accelero-
570 metric sensors distributed over the building height since 2010 (station: CGBP) including one
571 temperature sensor and one free-field station.
- 572 • Centre de découverte de la Terre in Martinique Island ([Gueguen, 2012](#)): 2 3C sensors in trigger
573 mode in a specific building with rubber bearing since 2005 (stations: CGCP and CGLR).

574 **Metadata editing made easy: YASMINE**

575 For almost 20 years, the IRIS Portable Data Collection Centers toolkit (PDCC, [Casey, 2016](#)) was
576 the main standalone GUI tool available to create and maintain station metadata in SEED dataless
577 format ([Ahern et al., 2012](#)). In 2017, several years after the adoption of the FDSN new StationXML
578 metadata standard ([FDSN, 2019](#)), there was still no solution widely available to edit and create native
579 StationXML files. This slowed down the adoption of StationXML by, at least, all Résif-SI contributing
580 A-nodes and forced us to use an interim SEED dataless to StationXML conversion solution. This
581 problem was hindering Résif-DC being able to export rich metadata to the user, including for example
582 the DOI of the network, or the contributing organizations.

583 Résif-SI first collected the metadata creation and edition needs among the French community
584 and closest partners. The requirements were then generalized so that they fulfill not only Résif-SI
585 needs but also a much broader international community.

586 In 2018, IRIS-DMC contracted Instrumental Software Technologies, Inc. (ISTI) to build a tool
587 for StationXML creation and editing and this first software was already satisfying some of Résif-SI
588 requirements. Résif-SI then contracted ISTI to continue the development of YASMINE (Yet Another
589 Station Metadata Information Editor) in order to satisfy our requirements. The results are two
590 independent pieces of software: `yasmine-GUI` and `yasmine-CLI`.

591 One of the main innovations was the introduction of the new Atomic Response Object Library
592 (AROL, [Wolyniec et al., 2019](#)). This library, written in YAML, is contributed by Résif-SI and is the
593 conversion of the long maintained PZ format library. Each stage of a device is only defined once and
594 then linked to the definition of the many different instrument configurations possibilities.

595 The web-based GUI `yasmine-GUI` offers the user the ability to create and edit StationXML
596 metadata. The user can create files from scratch through a guided wizard process and can import
597 response files from either NRL (Nominal Responses for seismic instruments Library, [Templeton,
598 2017](#)) or AROL. The user can also modify any part of the instrument response. A comparison mode
599 allows the user to compare the instrument response of two StationXML files sharing the same Station-
600 Channel-Network-Location (SCNL). Often used description elements (vault, geology, comments) can

601 be stored in the General ATomic llbrary of Tiny Objects (GATITO, in YAML, [Saurel et al., 2019](#)) to
602 simplify the task of homogenization of those elements over several StationXML files. Finally, the user
603 can store any of the Network, Station or Channel elements in a User Library that will store them as
604 templates for further re-use.

605 The command-line based `yasmine-CLI` allows the user to modify an existing StationXML file.
606 The user can add, delete or modify StationXML elements, except for the instrument response
607 elements. The user can also split a complex StationXML into multiple simpler files (e.g., split
608 a network StationXML into several per-station files) and print the instrument responses of each
609 channel/epoch contained in the StationXML. This tool will allow many automated StationXML file
610 modifications from scripts.

611 The two pieces of software are written in Python and rely heavily on ObsPy for the StationXML
612 content manipulation. They are both compatible with the latest StationXML 1.1 standard and are
613 released under the GNU GPL v3 license and distributed as Python packages. The Résif-SI installation
614 of YASMINE is reachable at <https://yasmine.resif.fr>.

615 **Conclusion and Future challenges**

616 The choice of a distributed yet strong architecture for the national Résif information system in seismol-
617 ogy, with clearly identified roles and responsibilities, has proven effective for France to deliver seismic
618 waveform data to users worldwide. The service to users has significantly improved in terms of features
619 offered and service robustness with minimal changes of running costs and keeping the same number
620 of permanent staff members. The advantage of the system is that it capitalizes on the distributed
621 human resources and competence nationwide, and maintains visibility for each involved institution.
622 The cost of this architecture is the continuous efforts needed to maintain the technical coherency and
623 cooperation across all institutions, and to overcome communication difficulties across the distributed
624 system. The choice made very early on to preserve original data models allowed us, despite the
625 initial cost, to respect the specific constraints on data and metadata for all the French networks,
626 an important consideration for the cohesion of Résif-SI and for the scientific users. Based on this

627 experience, the Résif board of directors has recently chosen a similar architecture for information on
628 earthquakes, with identified roles of different operators responsible for warning, bulletins, catalogs,
629 and shake maps, rather than each institution creating their own (Masson et al., 2021).

630

631 Résif faces future challenges, related to two main issues. First, the success of the system
632 at a national level means that the quantity and complexity of collected and distributed data is
633 ever-increasing. For example, all broadband Ocean Bottom Seismometer data will in the future be
634 integrated into Résif. The second issue is an order of magnitude change in data integration associated
635 with Large-N and distributed acoustic sensing (DAS) equipment. These challenges are international,
636 and Résif will continue to engage in international discussions on how to tackle them (see Quinteros
637 et al. 2021). We believe that the Résif architecture is sufficiently solid to accommodate the changes
638 through engagement of all the partners. The experience from Résif shows that a distributed but
639 strongly organized information system is a good architecture for countries that face the challenge of
640 integrating and distributing data across many organizations.

641 **Data and Resources**

642 All the software produced by Résif-SI is open source and available via <https://gitlab.com/resif> and
643 <https://gricad-gitlab.univ-grenoble-alpes.fr/OSUG/RESIF>.

644 **Acknowledgments**

645 We first of all thank all the network operators and scientists who entrust their datasets to Résif-SI, and
646 who work with us to ensure the highest possible level of data completeness, and data and metadata
647 quality.

648 We thank for their active participation in Résif-DC:

- 649 • our colleagues from the Direction générale des services informatiques of the Université Greno-

650 ble Alpes (DGDSI) (<https://services-numeriques-personnels.univ-grenoble-alpes.fr>) and the
651 Unité Mixte de Service GRICAD (<https://gricad.univ-grenoble-alpes.fr>): G. Bruno, G. Enderlé,
652 V. Louvet, A. Monot;

653 • the initiator of the SUMMER project: J. Le Tanou, and the technical committee of the SUMMER
654 storage infrastructure at Grenoble Alpes University (<https://summer.univ-grenoble-alpes.fr>);

655 • our colleagues from the Centre de données OSUG (OSUG-DC, [https://www.osug.fr/missions/
656 observation](https://www.osug.fr/missions/observation)): B. Bouterin and R. Cailletau;

657 • our colleagues from the Centre de Calcul de l'IN2P3 (CC-IN2P3, [https://cc.in2p3.fr/nos-services/
658 traitement-de-donnees](https://cc.in2p3.fr/nos-services/traitement-de-donnees)), and in particular J.-Y. Nief.

659 Résif-SI relies on a national and international collaborative framework and feedback from multiple
660 sources. We thank in particular:

661 • the European Technical Committee (ETC) of EIDA (<https://www.orfeus-eu.org/data/eida/structure>);

662 • Orfeus-DC (<https://www.orfeus-eu.org>) for its support in the early days of Résif-DC, and in
663 particular T. van Eck and R. Sleeman;

664 • the staff of IRIS-DS (<https://www.iris.edu/hq>) : T. Ahern, J. Carter, R. Casey, C. Trabant;

665 • The Résif-SI review committee (June 2014), for their precise and fruitful analysis: F. Genova
666 (president), A. Ferreira, G. Guyot, J.-Y. Nief, C. Rondenay, C. Surace.

667 The French seismological and geodetic network Résif-EPOS is a national facility cited in the
668 national Roadmap for Research Infrastructures published by the Ministry of Higher Education,
669 Research and Innovation. Résif-EPOS, which is coordinated by CNRS, receives funding, personnel
670 and other support from all the Consortium members. Résif-SI is also supported by the *System of
671 Observation and Experimentation for Research and the Environment* (SOERE) and by the Ministry
672 of Ecological Transition. Résif-SI has benefited from the Résif-CORE project (11-EQPX-0040),
673 funded by the national French funding program *Investissements d'Avenir* and managed by the French
674 National Research Agency (ANR).

675 We thank the two anonymous reviewers whose suggestions helped improve and clarify this
676 manuscript.

677 This paper is dedicated to the memory of Jocelyn Guilbert, former director of the “Laboratoire
678 de Détection et de Géophysique” (Laboratory of Detection and Geophysics) of CEA, deceased on
679 August 21, 2016. Jocelyn devoted a huge effort for the participation of CEA to Résif-SI and to the
680 construction of Résif in general.

681 **References**

- 682 Ahern, T., R. Casey, D. Barnes, R. Benson, T. Knight, and C. Trabant (2012). *SEED reference manual*.
683 Incorporated Research Institutions for Seismology (IRIS). <https://www.fdsn.org/publications>.
- 684 AlpArray Seismic Network (2015). AlpArray Seismic Network (AASN) temporary component.
685 doi:[10.12686/ALPARRAY/Z3_2015](https://doi.org/10.12686/ALPARRAY/Z3_2015).
- 686 Barruol, G., K. Sigloch, RHUM-RUM Group, and RESIF (2017). RHUM-RUM experiment, 2011-
687 2015, code YV (Réunion Hotspot and Upper Mantle – Réunion's Unterer Mantel) funded by ANR,
688 DFG, CNRS-INSU, IPEV, TAAF, instrumented by DEPAS, INSU-OBS, AWI and the Universities of
689 Muenster, Bonn, La Réunion. doi:[10.15778/RESIF.YV2011](https://doi.org/10.15778/RESIF.YV2011).
- 690 Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). Ob-
691 sPy: A Python Toolbox for Seismology, *Seismological Research Letters* **81**(3), 530–533.
692 doi:[10.1785/gssrl.81.3.530](https://doi.org/10.1785/gssrl.81.3.530).
- 693 Bollard, P. (2021). `resif-delivery-stats-plotter`. [https://gitlab.com/resif/](https://gitlab.com/resif/resif-delivery-stats-plotter)
694 [resif-delivery-stats-plotter](https://gitlab.com/resif/resif-delivery-stats-plotter). Accessed: 2021-02-15.
- 695 Brenguier, F., P. Kowalski, N. Ackerley, N. Nakata, P. Boué, M. Campillo, E. Larose, S. Rambaud,
696 C. Pequegnat, T. Lecocq, P. Roux, V. Ferrazzini, N. Villeneuve, N. M. Shapiro, and J. Chaput
697 (2015). Toward 4D Noise-Based Seismic Probing of Volcanoes: Perspectives from a Large-N
698 Experiment on Piton de la Fournaise Volcano, *Seismological Research Letters* **87**(1), 15–25.
699 doi:[10.1785/0220150173](https://doi.org/10.1785/0220150173).
- 700 Brenguier, F., D. Rivet, A. Obermann, N. Nakata, P. Boué, T. Lecocq, M. Campillo, and N. Shapiro
701 (2016). 4-D noise-based seismology at volcanoes: Ongoing efforts and perspectives, *Journal of*
702 *Volcanology and Geothermal Research* **321**, 182–195. doi:[10.1016/j.jvolgeores.2016.04.036](https://doi.org/10.1016/j.jvolgeores.2016.04.036).
- 703 Brzev, S., C. Scawthorn, A. W. Charleson, L. Allen, M. Greene, K. Jaiswal, and V. Silva (2013). GEM
704 Building Taxonomy (Version 2.0). Technical report, GEM Foundation.

705 Casey, R. (1999). Networked Data Centers (NetDC), *Data Services Newsletter* 1(1). [https://ds.iris.](https://ds.iris.edu/ds/newsletter/vol1/no1/225/networked-data-centers-netdc/)
706 [edu/ds/newsletter/vol1/no1/225/networked-data-centers-netdc/](https://ds.iris.edu/ds/newsletter/vol1/no1/225/networked-data-centers-netdc/), Accessed: 2020-09-24.

707 Casey, R. (2016). *Portable Data Collection Center (PDCC) v3.8.1 User Manual*. Incorporated
708 Research Institutions for Seismology (IRIS). [https://ds.iris.edu/ds/nodes/dmc/software/downloads/](https://ds.iris.edu/ds/nodes/dmc/software/downloads/pdcc)
709 [pdcc](https://ds.iris.edu/ds/nodes/dmc/software/downloads/pdcc).

710 Clinton, J., W. Crawford, C. Evangelidis, P. Kaestli, P. Gueguen, H. Pedersen, J. Quinteros, J.-M.
711 Saurel, R. Sleeman, and A. Strollo (2018). SERA Deliverable 4.2: Report on metadata challenges
712 and proposed solutions. Technical report, SERA. [http://www.sera-eu.org/en/Dissemination/](http://www.sera-eu.org/en/Dissemination/deliverables/)
713 [deliverables/](http://www.sera-eu.org/en/Dissemination/deliverables/).

714 Crawford, W. (2019). OBS data/metadata proposal. Technical report, FDSN. [http://www.ipgp.fr/](http://www.ipgp.fr/~crawford/documents.html)
715 [~crawford/documents.html](http://www.ipgp.fr/~crawford/documents.html).

716 Crawford, W., R. Bouazzouz, and L. Arean (2019). obsinfo. <https://gitlab.com/resif/obsinfo>. Ac-
717 cessed: 2021-02-15. See also <https://pypi.org/project/obsinfo/>.

718 Creative Commons (2013). Attribution 4.0 International – CC BY 4.0. [https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0)
719 [licenses/by/4.0](https://creativecommons.org/licenses/by/4.0). Accessed: 2020-10-08.

720 DataCite (2019). DataCite Metadata Schema. <https://schema.datacite.org>. Accessed: 2020-10-09.

721 Dougherty, S. L., E. S. Cochran, and R. M. Harrington (2019). The LARge-n Seismic Survey in
722 Oklahoma (LASSO) Experiment, *Seismological Research Letters*. doi:[10.1785/0220190094](https://doi.org/10.1785/0220190094).

723 Evans, P., A. Strollo, A. Clark, T. Ahern, R. Newman, J. Clinton, H. Pedersen, and C. Pequegnat
724 (2015). Why Seismic Networks Need Digital Object Identifiers, *Eos* 96. doi:[10.1029/2015eo036971](https://doi.org/10.1029/2015eo036971).

725 FDSN (2013). FDSN Web Services. <https://www.fdsn.org/webservices>. Accessed: 2020-10-05.

726 FDSN (2019). StationXML Schema 1.1. <http://www.fdsn.org/xml/station/>. Accessed: 2020-10-02.

727 FDSN (2020). Network Codes. <https://www.fdsn.org/networks>. Accessed: 2020-10-08.

728 Flajolet, P., Éric Fusy, O. Gandouet, and F. Meunier (2007). Hyperloglog: The analysis of a near-
729 optimal cardinality estimation algorithm. In *AofA'07: Proceedings Of The 2007 International*
730 *Conference On Analysis Of Algorithms*.

731 Folk, M., G. Heber, Q. Koziol, E. Pourmal, and D. Robinson (2011). An overview of the HDF5
732 technology suite and its applications. In *Proceedings of the EDBT/ICDT 2011 Workshop on Array*
733 *Databases*, pp. 36–47.

734 Geber, O. (2019). MORUMOTTO. <https://github.com/IPGP/morumotto>. Accessed: 2020-10-02.

735 GFZ and gempa GmbH (2008). SeedLink. <https://www.seiscomp.de/doc/apps/seedlink.html>. Ac-
736 cessed: 2020-10-06.

737 GFZ and GEMPA GmbH (2008). The SeisComP seismological software package.
738 doi:10.5880/GFZ.2.4.2020.003.

739 GFZ and gempa GmbH (2008). Waveform archives. [https://www.seiscomp.de/doc/base/concepts/](https://www.seiscomp.de/doc/base/concepts/waveformarchives.html)
740 [waveformarchives.html](https://www.seiscomp.de/doc/base/concepts/waveformarchives.html). Accessed: 2020-10-02.

741 Gimbert, F., U. Nanni, P. Roux, A. Helmstetter, S. Garambois, A. Lecointre, A. Walpersdorf, B. Jourdain,
742 M. Langlais, O. Larman, F. Lindner, A. Sergeant, C. Vincent, and F. Walter (2020). The RESOLVE
743 project: a multi-physics experiment with a temporary dense seismic array on the Argentière Glacier,
744 French Alps. <https://arxiv.org/abs/2009.06321>.

745 Goubier, O. and W. Crawford (2021). visualQC. <https://gitlab.com/resif/visualqc>. Accessed: 2021-
746 02-15.

747 Grünthal, G. (1998). European macroseismic scale 1998. Technical report, European Seismological
748 Commission (ESC).

749 Gueguen, P. (2012). Experimental analysis of the seismic response of one base-isolation building
750 according to different levels of shaking: example of the Martinique earthquake (2007/11/29) Mw
751 7.3, *Bulletin of Earthquake Engineering* **10**(4), 1285–1298. doi:10.1007/s10518-012-9355-x.

752 Guéguen, P., F. Guattari, C. Aubert, and T. Laudat (2020). Comparing Direct Observation
753 of Torsion with Array-Derived Rotation in Civil Engineering Structures, *Sensors* **21**(1), 142.
754 doi:[10.3390/s21010142](https://doi.org/10.3390/s21010142).

755 Hable, S., K. Sigloch, G. Barruol, S. C. Stähler, and C. Hadziioannou (2018). Clock errors in
756 land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise
757 cross-correlations, *Geophysical Journal International* **214**(3), 2014–2034. doi:[10.1093/gji/ggy236](https://doi.org/10.1093/gji/ggy236).

758 Heinloo, A. (2018). `fdsnws_scripts`. https://geofon.gfz-potsdam.de/software/fdsnws_scripts. Ac-
759 cessed: 2020-10-05.

760 Hess, D., N. Falco, @rsdeazevedo, @damhuonglan, and K. Jacobs (2018). PIC-IRIS/PH5: V4.1.2.
761 doi:[10.5281/ZENODO.841332](https://doi.org/10.5281/ZENODO.841332).

762 Hutko, A. (2013). Web Service Fetch scripts. [https://seiscode.iris.washington.edu/projects/
763 ws-fetch-scripts](https://seiscode.iris.washington.edu/projects/ws-fetch-scripts). Accessed: 2020-10-05.

764 IN2P3/CNRS (1976). Centre de Calcul de l'IN2P3/CNRS - USR6402. <https://cc.in2p3.fr>. Accessed:
765 2020-10-07.

766 IPGP (2007). 4G (2007-2025): EMSO-MOMAR. https://www.fdsn.org/networks/detail/4G_2007/.
767 Accessed: 2020-09-30.

768 IPGP (2008a). GNSS, seismic broadband and strong motion permanent networks in West Indies.
769 doi:[10.18715/ANTILLES.WI](https://doi.org/10.18715/ANTILLES.WI).

770 IPGP (2008b). Seismic, deformation, gas, magnetic and weather permanent networks on Mount
771 Pelée volcano and Martinique Island. doi:[10.18715/MARTINIQUE.MQ](https://doi.org/10.18715/MARTINIQUE.MQ).

772 IPGP (2008c). Seismic, deformation, gas, magnetic and weather permanent networks on Piton de la
773 Fournaise volcano and La Reunion Island. doi:[10.18715/REUNION.PF](https://doi.org/10.18715/REUNION.PF).

774 IPGP (2008d). Seismic, deformation, gas, magnetic and weather permanent networks on Soufrière
775 volcano and Guadeloupe Island. doi:[10.18715/GUADELOUPE.GL](https://doi.org/10.18715/GUADELOUPE.GL).

776 IPGP and EOST (1982). GEOSCOPE, French Global Network of broad band seismic stations.
777 doi:[10.18715/GEOSCOPE.G](https://doi.org/10.18715/GEOSCOPE.G).

778 IRIS (2016). WebServiceShell. <https://github.com/iris-edu/webserviceshell>. Accessed: 2020-10-02.

779 ISO/IEC 19464 (2014). Information technology – Advanced Message Queuing Protocol (AMQP)
780 v1.0 specification. Standard, International Organization for Standardization, Geneva, CH.
781 (See also: <https://www.amqp.org>).

782 Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015).
783 ObsPy: a bridge for seismology into the scientific Python ecosystem, *Computational Science &*
784 *Discovery* **8**(1), 014003. doi:[10.1088/1749-4699/8/1/014003](https://doi.org/10.1088/1749-4699/8/1/014003).

785 Laigle, M., J. Lebrun, and A. Hirn (2007). SISMANTILLES 2 cruise, RV L'Atalante.
786 doi:[10.17600/7010020](https://doi.org/10.17600/7010020).

787 Lorenzo, G. W. F., M. P. S. d'Avila, A. Deschamps, E. Bertrand, E. D. Mercerat, L. Foundo-
788 tos, and F. Courboux (2018). Numerical and Empirical Simulation of Linear Elastic Seismic
789 Response of a Building: The Case of Nice Prefecture, *Earthquake Spectra* **34**(1), 169–196.
790 doi:[10.1193/042216eqs064m](https://doi.org/10.1193/042216eqs064m).

791 Loviknes, K., Z. Jeddi, L. Ottemöller, and T. Barreyre (2020). When Clocks Are Not Working: OBS
792 Time Correction, *Seismological Research Letters* **91**(4), 2247–2258. doi:[10.1785/0220190342](https://doi.org/10.1785/0220190342).

793 MacCarthy, J., O. Marcillo, and C. Trabant (2020). Seismology in the Cloud: A New Streaming
794 Workflow, *Seismological Research Letters* **91**(3), 1804–1812. doi:[10.1785/0220190357](https://doi.org/10.1785/0220190357).

795 Massinon, B. and J. Plantet (1976). A large-aperture seismic network in France: Description and
796 some results concerning epicenter location and upper-mantle anomalies, *Physics of the Earth and*
797 *Planetary Interiors* **12**(2-3), 118–127. doi:[10.1016/0031-9201\(76\)90041-8](https://doi.org/10.1016/0031-9201(76)90041-8).

798 Masson, F., S. Auclair, D. Bertil, M. Grunberg, B. Hernandez, S. Lambotte, G. Mazet-Roux, L. Provost,
799 J.-M. Saurel, A. Schlupp, and C. Sira (2021). The transversal seismicity action RESIF: a tool to
800 improve the distribution of the French seismicity products, *Seismological Research Letters*.

801 McNamara, D. E. (2004). Ambient Noise Levels in the Continental United States, *Bulletin of the*
802 *Seismological Society of America* **94**(4), 1517–1527. doi:10.1785/012003001.

803 McNamara, D. E. and R. I. Boaz (2010). PQLX: A seismic data quality control system description,
804 applications, and users manual, *US Geol. Surv. Open-File Rept* **1292**, 41.

805 Megies, T., M. Beyreuther, R. Barsch, L. Krischer, and J. Wassermann (2011). ObsPy – What can it
806 do for data centers and observatories?, *Annals of Geophysics* **54**(1). doi:10.4401/ag-4838.

807 Michel, C. and P. Guéguen (2018). Interpretation of the velocity measured in buildings by seismic
808 interferometry based on Timoshenko beam theory under weak and moderate motion, *Soil Dynamics*
809 *and Earthquake Engineering* **104**, 131–142. doi:10.1016/j.soildyn.2017.09.031.

810 Michel, C., P. Guéguen, S. E. Arem, J. Mazars, and P. Kotronis (2009). Full-scale dynamic response
811 of an RC building under weak seismic motions using earthquake recordings, ambient vibrations
812 and modelling, *Earthquake Engineering & Structural Dynamics*, 419–441. doi:10.1002/eqe.948.

813 Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan, *PLOS*
814 *Computational Biology* **11**(10), e1004525. doi:10.1371/journal.pcbi.1004525.

815 MongoDB, Inc. (2009). MongoDB. <https://www.mongodb.com>. Accessed: 2020-10-02.

816 ORFEUS (2020a). EIDA Federator. <https://www.orfeus-eu.org/data/eida/nodes/FEDERATOR>. Ac-
817 cessed: 2020-10-05.

818 ORFEUS (2020b). EIDA webdc3. <http://orfeus-eu.org/webdc3>. Accessed: 2020-10-05.

819 Pequegnat, C., P. Guéguen, D. Hatzfeld, and M. Langlais (2008). The French Accelerometric Network
820 (RAP) and National Data Centre (RAP-NDC), *Seismological Research Letters* **79**(1), 79–89.
821 doi:10.1785/gssrl.79.1.79.

822 PostgreSQL Global Development Group (1996). PostgreSQL. <https://www.postgresql.org>. Accessed:
823 2020-10-02.

824 Quinteros, J., J. A. Carter, J. Schaeffer, C. Trabant, and H. A. Pedersen (2021). Exploring Approaches
825 for Large Data in Seismology: User and Data Repository Perspectives, *Seismological Research*
826 *Letters*. doi:[10.1785/0220200390](https://doi.org/10.1785/0220200390).

827 re3data.org (2016). RESIF Seismic Data Portal. doi:[10.17616/R37Q06](https://doi.org/10.17616/R37Q06).

828 Ronacher, A. (2010). Flask. <https://flask.palletsprojects.com>. Accessed: 2020-10-02.

829 Roult, G., J.-P. Montagner, E. Stutzmann, S. Barbier, and G. Guiveneux (1999). The GEO-
830 SCOPE program: its data center, *Physics of the Earth and Planetary Interiors* **113**(1-4), 25–43.
831 doi:[10.1016/s0031-9201\(99\)00024-2](https://doi.org/10.1016/s0031-9201(99)00024-2).

832 Résif (1995). RESIF-RAP French Accelerometric Network. doi:[10.15778/RESIF.RA](https://doi.org/10.15778/RESIF.RA).

833 Résif (2017). RESIF-RENAG French national Geodetic Network. doi:[10.15778/RESIF.RG](https://doi.org/10.15778/RESIF.RG).

834 Résif (2018). CEA/DASE broad-band permanent network in metropolitan France.
835 doi:[10.15778/RESIF.RD](https://doi.org/10.15778/RESIF.RD).

836 Résif (2019a). Résif Web Services. <http://ws.resif.fr>. Accessed: 2020-10-06.

837 Résif (2019b). Webservice ph5-availability. <http://ws.resif.fr/resifws/ph5-availability/1/>. Accessed:
838 2020-10-05.

839 Résif (2019c). Webservice ph5-dataselect. <http://ws.resif.fr/resifws/ph5-dataselect/1/>. Accessed:
840 2020-10-05.

841 Résif-SI (2020). Networks citation page. http://seismology.resif.fr/#CMSConsultPlace:DOI_
842 [INVENTORY](https://doi.org/10.15778/RESIF.SI). Accessed: 2020-10-05.

843 Saurel, J.-M., J. Schaeffer, and C. Péquegnat (2019). GATITO: Generic ATomic llbrary of Tiny
844 Objects: Library of metadata objects for solid earth science metadata, mostly enumerations.
845 <https://gitlab.com/resif/gatito>. Accessed: 2020-10-12.

846 Shapiro, N., V. Douet, C. Pardo, C. Péquegnat, S. Barbier, C. Maron, M. Grunberg, M. Schaming,
847 et al. (2008). FOSFORE: portal for distributions of French seismological data, *AGUFM* **2008**,
848 S43D–1909.

849 Strollo, A., D. Cambaz, J. Clinton, P. Danecek, C. P. Evangelidis, A. Marmureanu, L. Ottemöller,
850 H. Pedersen, R. Sleeman, K. Stammler, D. Armbruster, J. Bienkowski, K. Boukouras, P. L. Evans,
851 M. Fares, C. Neagoe, S. Heimers, A. Heinloo, M. Hoffmann, P. Kaestli, V. Lauciani, J. Michalek, E. O.
852 Muhire, M. Ozer, L. Palangeanu, C. Pardo, J. Quinteros, M. Quintiliani, J. A. Jara-Salvador, J. Scha-
853 effer, A. Schloemer, and N. Triantafyllis (2021). EIDA: The European Integrated Data Archive and
854 Service Infrastructure within ORFEUS, *Seismological Research Letters*. doi:[10.1785/0220200413](https://doi.org/10.1785/0220200413).

855 Templeton, M. E. (2017). IRIS Library of Nominal Responses for Seismic Instruments.
856 doi:[10.17611/S7159Q](https://doi.org/10.17611/S7159Q).

857 Trabant, C. (2011). ringserver - Generic packet ring buffer with network interfaces. [https://github.com/](https://github.com/iris-edu/ringserver)
858 [iris-edu/ringserver](https://github.com/iris-edu/ringserver). Accessed: 2020-10-06.

859 Trani, L., M. Koymans, M. Atkinson, R. Sleeman, and R. Filgueira (2017). WFCatalog:
860 A catalogue for seismological waveform data, *Computers & Geosciences* **106**, 101–108.
861 doi:[10.1016/j.cageo.2017.06.008](https://doi.org/10.1016/j.cageo.2017.06.008).

862 Tridgell, A. and P. Mackerras (1996). The rsync algorithm. Technical Report TR-CS-96-05, Australian
863 National University, Department of Computer Science. (See also: <http://rsync.samba.org>).

864 Université Grenoble Alpes (2016). SUMMER: Stockage Unifié Mutualisé Massif Evolutif et Réparti.
865 <https://summer.univ-grenoble-alpes.fr>. Accessed: 2020-10-07.

866 Vallée, M., J. Charléty, A. M. G. Ferreira, B. Delouis, and J. Vergoz (2011). SCARDEC: a new
867 technique for the rapid determination of seismic moment magnitude, focal mechanism and source
868 time functions for large earthquakes using body-wave deconvolution, *Geophysical Journal Interna-*
869 *tional* **184**(1), 338–358. doi:[10.1111/j.1365-246x.2010.04836.x](https://doi.org/10.1111/j.1365-246x.2010.04836.x).

870 Voigt, C., C. Förste, H. Wziontek, D. Crossley, B. Meurers, V. Pálinkáš, J. Hinderer, J.-P. Boy, J.-P.

871 Barriot, and H. Sun (2016). Report on the Data Base of the International Geodynamics and Earth
872 Tide Service (IGETS). Technical report. doi:[10.2312/GFZ.B103-16087](https://doi.org/10.2312/GFZ.B103-16087).

873 Volcke, P., R. Bouazzouz, J. Schaeffer, and C. Péquegnat (2013). ResifDataTransfer. <https://gitlab.com/resif/resif-data-transfer>. Accessed: 2020-10-06.
874

875 Volcke, P., J. Schaeffer, and C. Péquegnat (2012). seedtree5. <https://gitlab.com/resif/seedtree5>.
876 Accessed: 2020-10-06.

877 Weertman, B. (2010). Web Services at the DMC, *Data Services Newsletter* **12**(3). [http://ds.iris.edu/](http://ds.iris.edu/ds/newsletter/vol12/no3/44/web-services-at-the-dmc)
878 [ds/newsletter/vol12/no3/44/web-services-at-the-dmc](http://ds.iris.edu/ds/newsletter/vol12/no3/44/web-services-at-the-dmc), Accessed: 2020-10-06.

879 Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg,
880 J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas,
881 I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth,
882 C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher,
883 M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A.
884 Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der
885 Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and
886 B. Mons (2016). The FAIR Guiding Principles for scientific data management and stewardship,
887 *Scientific Data* **3**(1). doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

888 Wolyniec, D., J. Schaeffer, J.-M. Saurel, and C. Péquegnat (2019). AROL: Atomic Response
889 Objects Library containing metadata descriptions of earth science observation instruments. <https://gitlab.com/resif/arol>.
890 Accessed: 2020-10-12.

891 Zaccarelli, R., D. Bindi, A. Strollo, J. Quinteros, and F. Cotton (2019). Stream2segment: An Open-
892 Source Tool for Downloading, Processing, and Visualizing Massive Event-Based Seismic Waveform
893 Datasets, *Seismological Research Letters*. doi:[10.1785/0220180314](https://doi.org/10.1785/0220180314).

894 **Full mailing address for each author**

895 **Catherine Péquegnat** – catherine.pequegnat@univ-grenoble-alpes.fr

896 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
897 38000 Grenoble, France

898 **Jonathan Schaeffer** – jonathan.schaeffer@univ-grenoble-alpes.fr

899 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

900 **Claudio Satriano** – satriano@ipgp.fr

901 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

902 **Helle Pedersen** – helle.pedersen@univ-grenoble-alpes.fr

903 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
904 38000 Grenoble, France

905 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

906 **Jérôme Touvier** – jerome.touvier@univ-grenoble-alpes.fr

907 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
908 38000 Grenoble, France

909 **Jean-Marie Saurel** – saurel@ipgp.fr

910 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

911 **Marie Calvet** – marie.calvet@irap.omp.eu

912 Institut de Recherche en Astrophysique et Planétologie,
913 Université Toulouse III Paul Sabatier, CNRS, CNES, Toulouse, France

914 **Laurent Stehly** – laurent.stehly@univ-grenoble-alpes.fr

915 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
916 38000 Grenoble, France

917 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

918 **Patrick Arnoul** – patrick.arnoul@cea.fr

919 CEA/DAM/DIF, F-91297 Arpajon, France

920

921 **Philippe Bollard** – philippe.bollard@univ-grenoble-alpes.fr
922 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

923 **Olivier Dewée** – olivier.dewee@cnrs.fr
924 Division Technique de l'INSU, Meudon, France

925 **Fabien Engels** – fabien.engels@unistra.fr
926 Université de Strasbourg, CNRS, EOST UMS 830, 67000, Strasbourg, France

927 **Marc Grunberg** – marc.grunberg@unistra.fr
928 Université de Strasbourg, CNRS, EOST UMS 830, 67000, Strasbourg, France

929 **Constanza Pardo** – pardo@ipggp.fr
930 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

931 **David Wolyniec** – david.wolyniec@univ-grenoble-alpes.fr
932 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

933 **Grégory Arneodo** – gregory.arneodo@ac-grenoble.fr
934 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
935 38000 Grenoble, France

936 **Rima Bouazzouz** – bouazzouz.rima@hotmail.fr
937 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
938 38000 Grenoble, France

939 **Olivier Geber** – olivier.geber@gmail.com
940 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

941 **Coralie Aubert** – coralie.aubert@univ-grenoble-alpes.fr
942 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
943 38000 Grenoble, France

944 **Mickaël Langlais** – mickael.langlais@univ-grenoble-alpes.fr
945 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
946 38000 Grenoble, France

947

948 **Sandrine Roussel** – sandrine.roussel@univ-grenoble-alpes.fr
949 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
950 38000 Grenoble, France

951 **Florent Brenguier** – florent.brenguier@univ-grenoble-alpes.fr
952 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
953 38000 Grenoble, France
954 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

955 **Wayne C Crawford** – crawford@ipgp.fr
956 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

957 **Arnaud Lemarchand** – arnaudl@ipgp.fr
958 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

959 **Emeline Maufroy** – emeline.maufroy@univ-grenoble-alpes.fr
960 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
961 38000 Grenoble, France
962 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

963 **Serge Olivier** – serge.olivier@cea.fr
964 CEA/DAM/DIF, F-91297 Arpajon, France

965 **Diane Rivet** – diane.rivet@geoazur.unice.fr
966 Université Côte d'Azur, CNRS, IRD, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France

967 **Martin Vallée** – vallee@ipgp.fr
968 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

969 **Jérôme Vergne** – jerome.vergne@unistra.fr
970 Université de Strasbourg, CNRS, IPGS UMR7516, 67000, Strasbourg, France

971 **Jérôme Chèze** – cheze@geoazur.unice.fr
972 Université Côte d'Azur, CNRS, IRD, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France
973

974 **Éric Debayle** – eric.debayle@ens-lyon.fr
975 Laboratoire de Géologie de Lyon : Terre, Planètes, Environnement, CNRS, UMR 5276,
976 École Normale Supérieure de Lyon, Université de Lyon, Université Claude Bernard Lyon 1,
977 Villeurbanne, France

978 **Bertrand Delouis** – delouis@geoazur.unice.fr
979 Université Côte d'Azur, CNRS, IRD, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France

980 **Philippe Gueguen** – philippe.gueguen@univ-grenoble-alpes.fr
981 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
982 38000 Grenoble, France

983 **Sophie Lambotte** – sophie.lambotte@unistra.fr
984 Université de Strasbourg, CNRS, IPGS UMR7516, 67000, Strasbourg, France

985 **Antoine Mocquet** – antoine.mocquet@univ-nantes.fr
986 Laboratoire de Planétologie et de Géodynamique, Université de Nantes, Université d'Angers,
987 UMR-CNRS 6112, Nantes, France

988 **Anne Paul** – anne.paul@univ-grenoble-alpes.fr
989 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
990 38000 Grenoble, France

991 **Fabrice Peix** – fabrice.peix@geoazur.unice.fr
992 Université Côte d'Azur, CNRS, IRD, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France

993 **François Schindelé** – francois.schindele@cea.fr
994 CEA/DAM/DIF, F-91297 Arpajon, France

995 **Nikolai Shapiro** – nikolai.shapiro@univ-grenoble-alpes.fr
996 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
997 38000 Grenoble, France

998 **Éléonore Stutzmann** – stutz@ipgp.fr
999 Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France
1000

1001 **Pierre Volcke** – pierre.volcke@isere.fr
1002 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
1003 38000 Grenoble, France
1004 **Andrea Walpersdorf** – andrea.walpersdorf@univ-grenoble-alpes.fr
1005 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre,
1006 38000 Grenoble, France
1007 Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France
1008

Table 1: Typology of stations by networks within Résif A-nodes.

RLBP: French broadband permanent network (FR) co-hosted by École et Observatoire des Sciences de la Terre (EOST) and Observatoire de la Côte d'Azur (OCA); also receives data from partner permanent networks (CL, ND, MT).

RAP: Permanent Accelerometric Network (RA, FR), hosted by Observatoire des Sciences de l'Univers de Grenoble (OSUG).

GEOSCOPE: GEOSCOPE Global Observatory (G), hosted by Institut de physique du globe de Paris (IPGP).

VOLCANO: Seismological instruments from the Observatories of Piton de la Fournaise (Reunion Island, PF), Guadeloupe and Martinique (GL, MQ, WI), hosted by Institut de physique du globe de Paris (IPGP).

SISMOB: Land mobile instruments, hosted by ISTerre; it also receives data from partner laboratories. Data can be embargoed up to three years after the experiment start.

CEA: CEA broadband stations (RD).

NRT: number of stations with at least one near-real time stream. BB: number of stations with at least one broad band channel (NRT or not). SP: number of stations with at least one short period channel (NRT or not). SM: number of stations with at least one accelerometric channel (NRT or not). Others (NRT or not): tiltmeter, weather, hydrophone, wind, rotational sensor, mass position, etc. Note that for SISMOB the "Others" category includes very short period "node" geophones (channel codes starting with "DP").

<i>Résif A-node</i>	<i>FDSN network codes</i>	<i>stations</i>	<i>NRT</i>	<i>BB</i>	<i>SP</i>	<i>SM</i>	<i>Others</i>
RLBP	FR	167	145	167	6		3
	CL	39	15	37	19	4	3
	ND	10	1	10		7	
	MT	16	10	16	5		1
RAP	FR	49	39			49	
	RA	232	93			232	
GEOSCOPE	G	54	30	54		10	36
VOLCANO	WI	15	11	15		15	
	PF	46	36	29	20		
	GL	22	9	11	18		
	MQ	14	12	8	8		
SISMOB	70 temporary network codes	2482		1271	277	63	1022
CEA	RD	19	13	19		1	
	<i>Total</i>	<i>3165</i>	<i>414</i>	<i>1637</i>	<i>353</i>	<i>381</i>	<i>1065</i>

Table 2: Latency for the production of validated data by A-nodes

<i>A-nodes</i>	<i>data quality</i>	<i>latency</i>
RLBP	M	3-5 days
RAP	M	3-5 days; 1 year for the data from stations in buildings and borehole
GEOSCOPE	Q	6 to 12 months
VOLCANO	Q	6 to 12 months
SISMOB	M	1 to 6 months

Table 3: Average service availability for year 2020 based on network reachability. Availability of 0.1% corresponds to 9 hours.

<i>service</i>	<i>availability (%)</i>
fdsnws-dataselect	99.84
fdsnws-station	99.62
fdsnws-availability	99.60
real time data	99.87
resifws-timeseries	99.66
resifws-timeseriesplot	99.65
eidaws-wfcatalog	99.29

1009 **List of Figure Captions**

1010 Figure 1 Seismic stations distributed by Résif-SI worldwide, in metropolitan France and
1011 in overseas Réunion and Lesser Antilles regions (location of these regions is indicated
1012 by the red boxes on the world map). Symbols and colors according to the network;
1013 number of stations for each network indicated in parentheses (see Table 1 for details
1014 on the network codes). The “Temp.” label includes all temporary deployments (land
1015 mobile instrument pool SISMOB and ocean bottom seismometer instrument pool). The
1016 three pink boxes on the “Réunion” map are deployments of 100 sensors each (see
1017 Brenquier et al., 2015). Note that the maps do not show all the Résif-SI data holdings
1018 worldwide and other stations are available from permanent or temporary deployments
1019 in Europe, Chile and Pacific Ocean. Note also that the maps include stations which
1020 are temporary or definitively closed. 49

1021 Figure 2 Structural architecture of Résif-SI. The A-nodes collect and validate data and
1022 metadata from permanent and temporary networks. They are also responsible to
1023 secure a copy of the data for two years. Data and metadata are then submitted
1024 to the B-node Résif-DC, which is responsible for long-term storage and distribution.
1025 Résif-DC also computes side-products (e.g., Power Spectral Density plots, McNamara,
1026 2004), and is member of EIDA. Overall interoperability between data from different
1027 observations within a large part of Solid Earth in Europe is ensured by EPOS. 50

1028 Figure 3 Distribution of data volumes in Résif archive, as of January 1st, 2021. Veloci-
1029 metric data represents 62.3%. Accelerometric data represents 30.2%, Data from very
1030 short period "node" geophones represents 7.1% and other data, like meteorological
1031 time series, accounts for 0.4%. 51

1032 Figure 4 Technical architecture of Résif-SI. A-nodes collect and validate the raw data,
1033 manage real-time data flow, edit the metadata and submit them to the B-node Résif-DC.
1034 Résif-DC concentrates the real-time data flow which is used by the national seismic
1035 alert system operated by the CEA (on top of their own dedicated data flow for the
1036 networks they manage), stores data and metadata in the long term, distributes it
1037 through standardized and specific media, develops side-products. Federated into
1038 EIDA, open data and metadata are accessible by end-users through a federator. . . . 52

1039 Figure 5 Validated data and metadata are integrated at Résif-DC by passing through a
1040 series of workers which ensure that the data and metadata conforms to the specifica-
1041 tions, register it in the archive and in the database, index it and create the side-products.
1042 Raw data is concentrated, registered and made available through standardized web
1043 services with very low latency (around 10 seconds). Data, metadata and other assets
1044 are dumped on a distant tape system at the Computing Centre of the National Institute
1045 of Nuclear Physics and Particle Physics (CC-IN2P3, Lyon) to secure them from a local
1046 disaster. 53

1047 Figure 6 Current validated data holdings at Résif-DC grouped by year. This figures
1048 illustrates the growth of the data produced by seismological networks since 1998. . . . 54

1049 Figure 7 Amount of waveform data distributed yearly through SeedLink (real-time) and
1050 FDSN `dataselect`. Note that statistics from the now retired EIDA Arlink services
1051 are not included. 55

1052 Figure 8 The logging system analyzes events from heterogeneous sources (e.g., web
1053 access logs, web services requests, server logs); the statistics are computed by
1054 different analyzers and aggregated in one central database (Postgresql). The statistics
1055 can be queried by various interfaces intended for different audiences; in particular a
1056 public web service `resifws-statistics` is available. 56

1057 Figure 9 This dashboard view is an example of the web interface presenting some of the
1058 Résif-DC distribution metrics. Here, the figures focuses on the `fdsnws-dataselect`
1059 web service during 2020, showing the volume sent (51.34 TB), the number of suc-
1060 cessful requests (33.02 millions) and the distribution of user agent requests for all the
1061 requests (83.75 millions). This last number accounts for the successful requests to
1062 the `/query` and `/queryauth` methods (HTTP 200), the requests returning no data
1063 (HTTP 204 and 404), the requests to other methods (documentation, `/auth`, wadl file),
1064 the requests producing errors due to improper parameters (HTTP 400 and 401) or
1065 insufficient permissions (HTTP 403). 57

1066 Figure 10 Geographical distribution of data requests to Résif-DC in 2020. 58

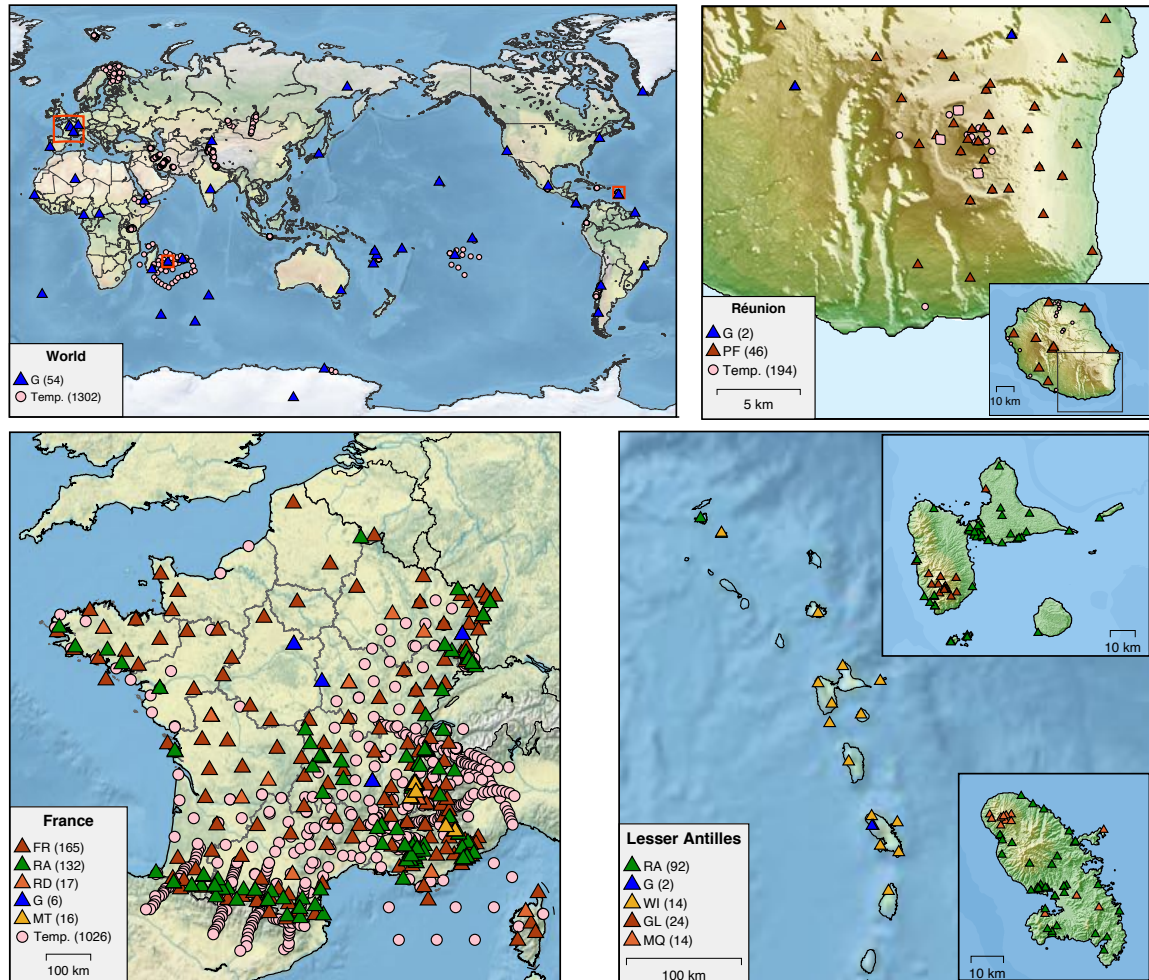


Figure 1: Seismic stations distributed by Résif-SI worldwide, in metropolitan France and in overseas Réunion and Lesser Antilles regions (location of these regions is indicated by the red boxes on the world map). Symbols and colors according to the network; number of stations for each network indicated in parentheses (see Table 1 for details on the network codes). The “Temp.” label includes all temporary deployments (land mobile instrument pool SISMOB and ocean bottom seismometer instrument pool). The three pink boxes on the “Réunion” map are deployments of 100 sensors each (see [Brenquier et al., 2015](#)). Note that the maps do not show all the Résif-SI data holdings worldwide and other stations are available from permanent or temporary deployments in Europe, Chile and Pacific Ocean. Note also that the maps include stations which are temporary or definitively closed.

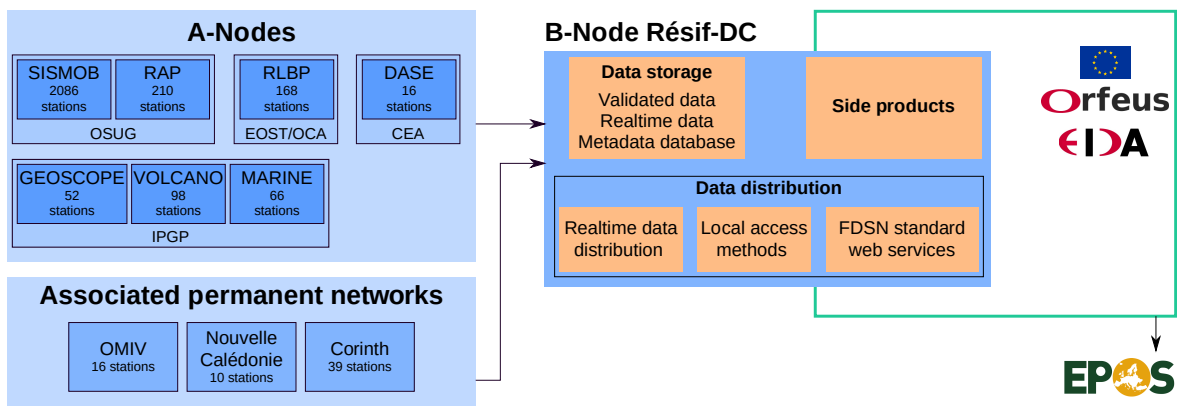


Figure 2: Structural architecture of Résif-SI. The A-nodes collect and validate data and metadata from permanent and temporary networks. They are also responsible to secure a copy of the data for two years. Data and metadata are then submitted to the B-node Résif-DC, which is responsible for long-term storage and distribution. Résif-DC also computes side-products (e.g., Power Spectral Density plots, [McNamara, 2004](#)), and is member of EIDA. Overall interoperability between data from different observations within a large part of Solid Earth in Europe is ensured by EPOS.

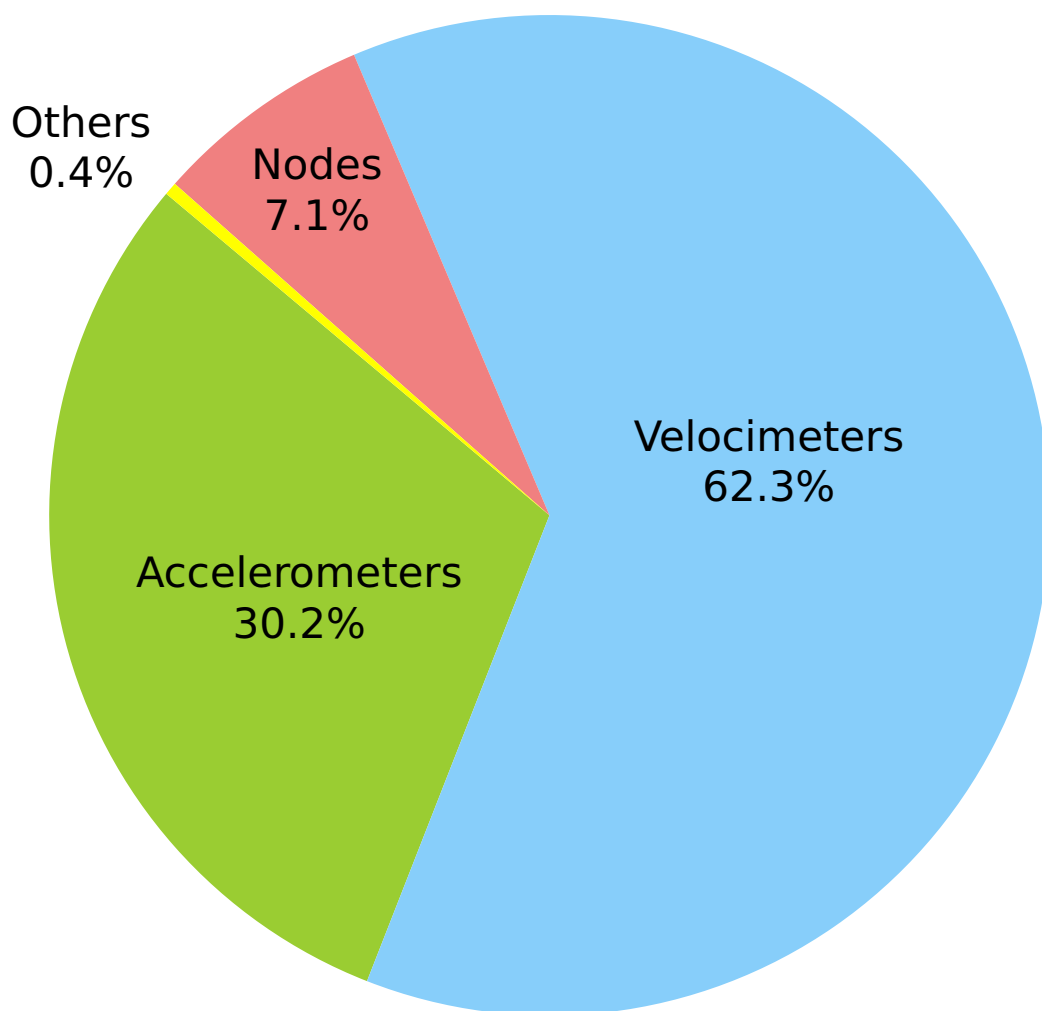


Figure 3: Distribution of data volumes in Résif archive, as of January 1st, 2021. Velocimetric data represents 62.3%. Accelerometric data represents 30.2%, Data from very short period "node" geophones represents 7.1% and other data, like meteorological time series, accounts for 0.4%.

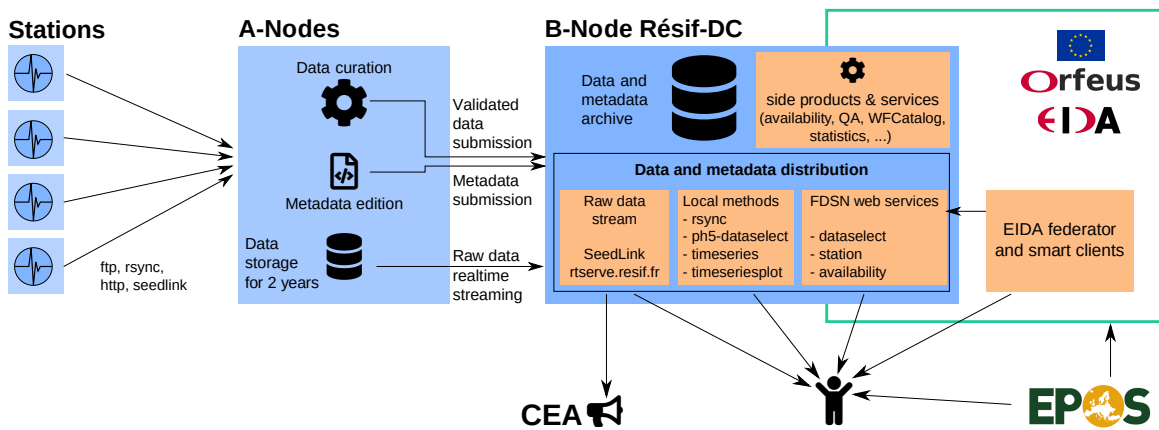


Figure 4: Technical architecture of Résif-SI. A-nodes collect and validate the raw data, manage real-time data flow, edit the metadata and submit them to the B-node Résif-DC. Résif-DC concentrates the real-time data flow which is used by the national seismic alert system operated by the CEA (on top of their own dedicated data flow for the networks they manage), stores data and metadata in the long term, distributes it through standardized and specific media, develops side-products. Federated into EIDA, open data and metadata are accessible by end-users through a federator.

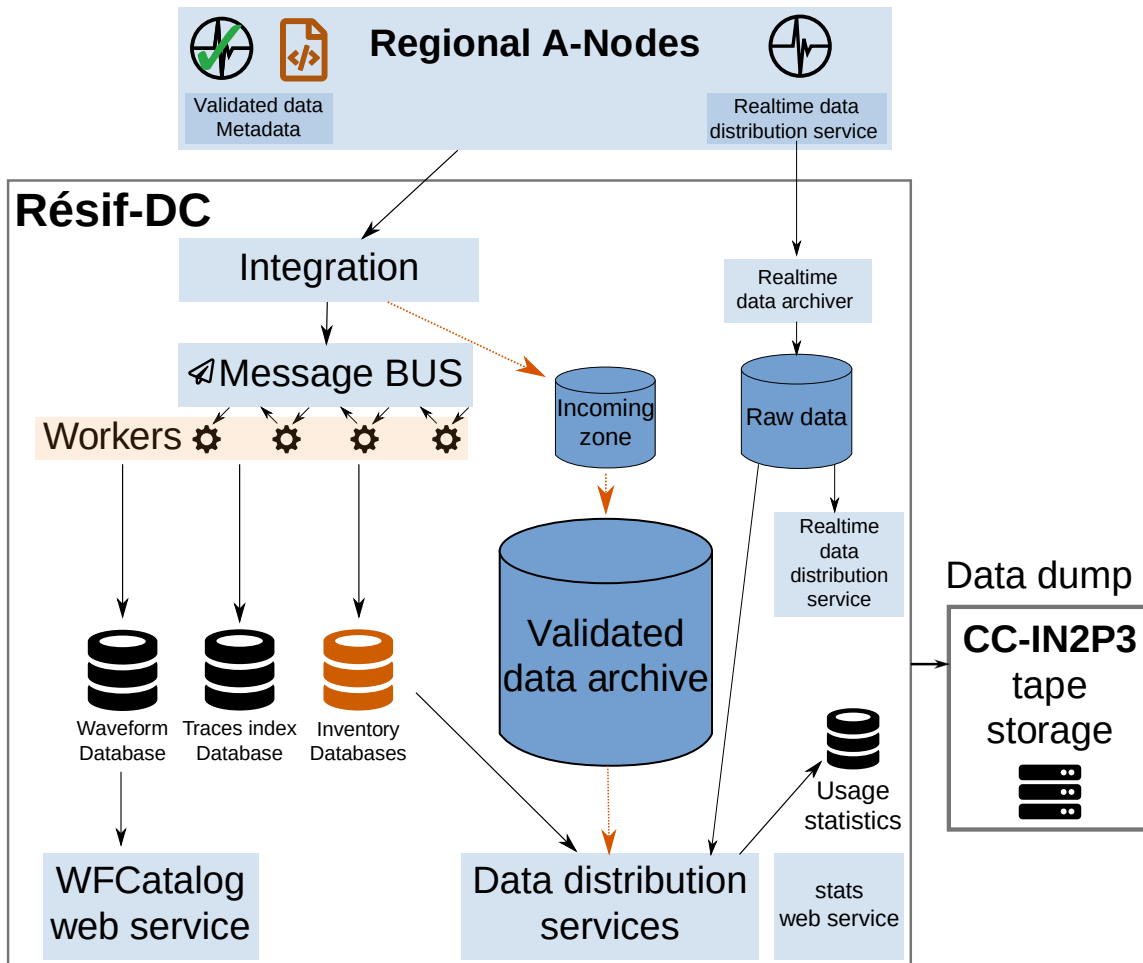


Figure 5: Validated data and metadata are integrated at Résif-DC by passing through a series of workers which ensure that the data and metadata conforms to the specifications, register it in the archive and in the database, index it and create the side-products. Raw data is concentrated, registered and made available through standardized web services with very low latency (around 10 seconds). Data, metadata and other assets are dumped on a distant tape system at the Computing Centre of the National Institute of Nuclear Physics and Particle Physics (CC-IN2P3, Lyon) to secure them from a local disaster.

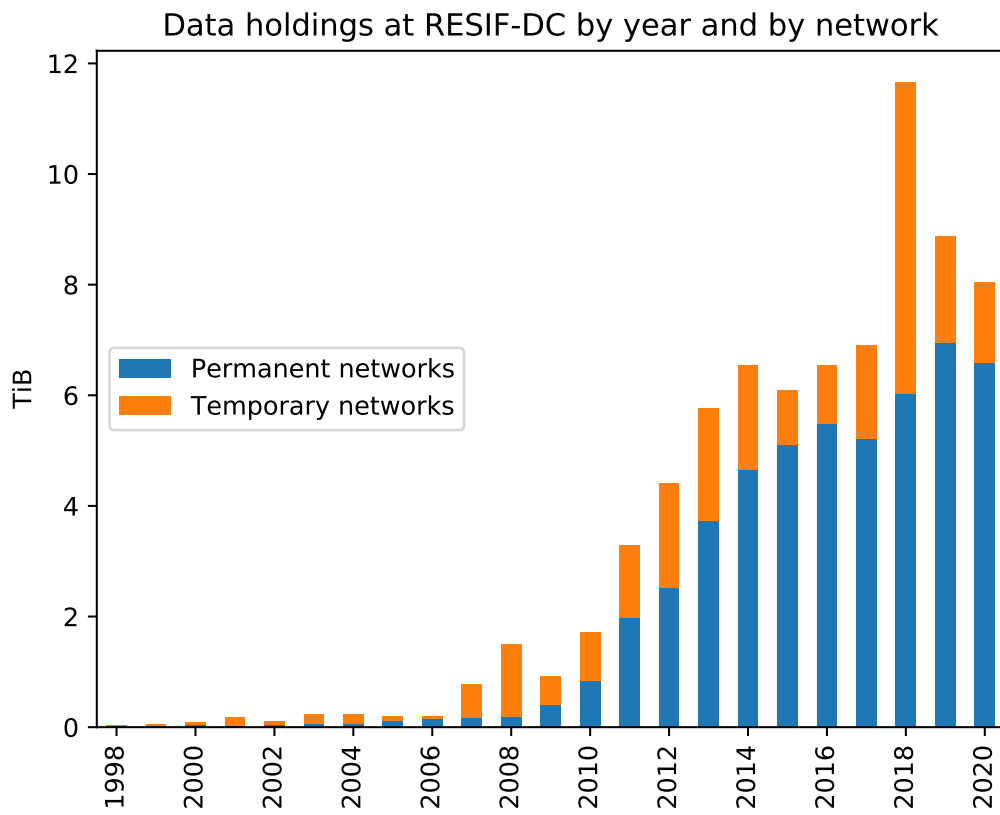


Figure 6: Current validated data holdings at Résif-DC grouped by year. This figures illustrates the growth of the data produced by seismological networks since 1998.

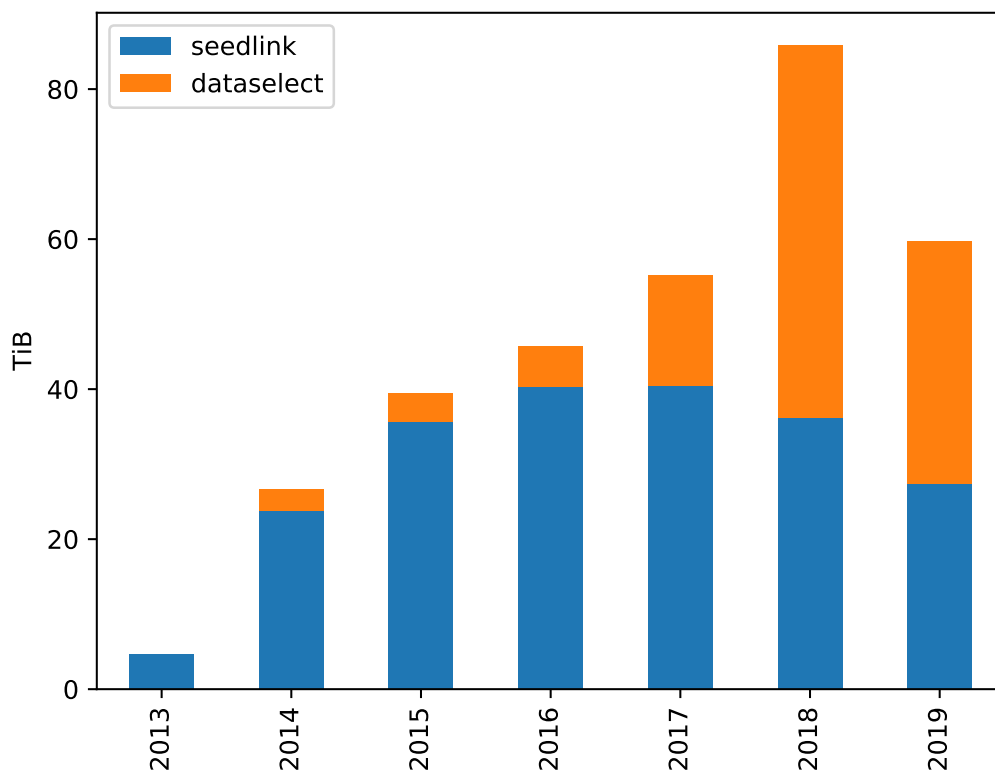


Figure 7: Amount of waveform data distributed yearly through SeedLink (real-time) and FDSN dataselect. Note that statistics from the now retired EIDA Arclink services are not included.

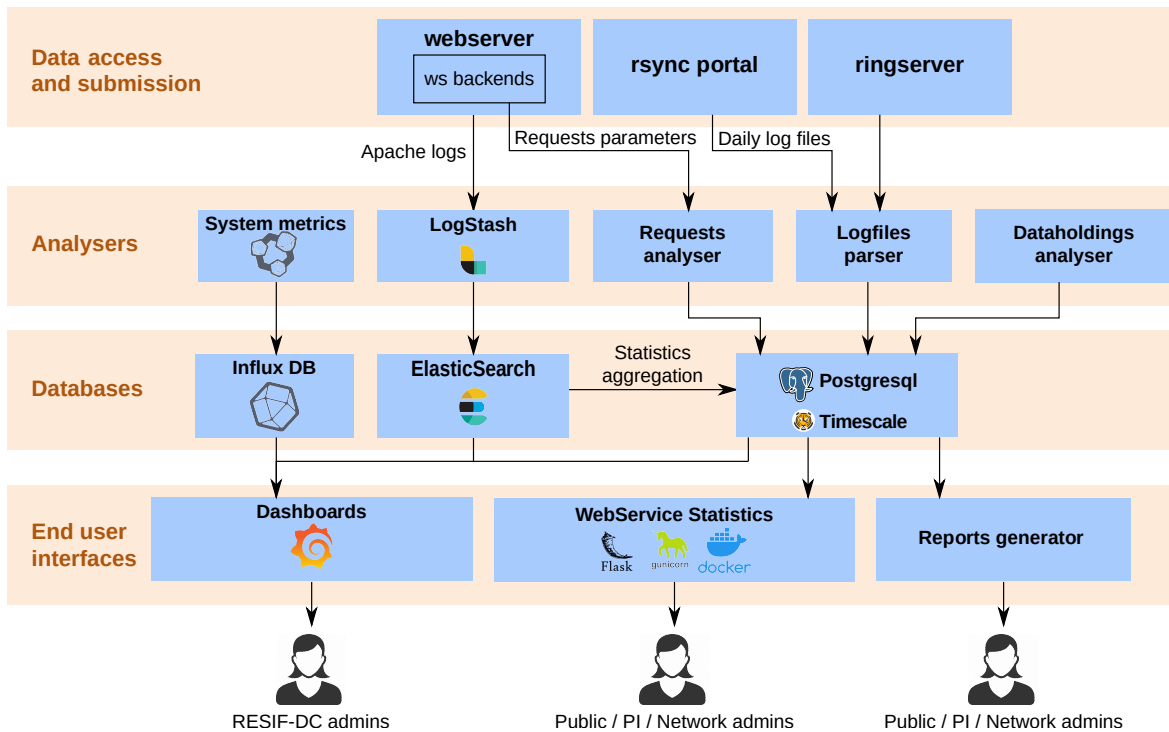


Figure 8: The logging system analyzes events from heterogeneous sources (e.g., web access logs, web services requests, server logs); the statistics are computed by different analyzers and aggregated in one central database (Postgresql). The statistics can be queried by various interfaces intended for different audiences; in particular a public web service `resifws-statistics` is available.

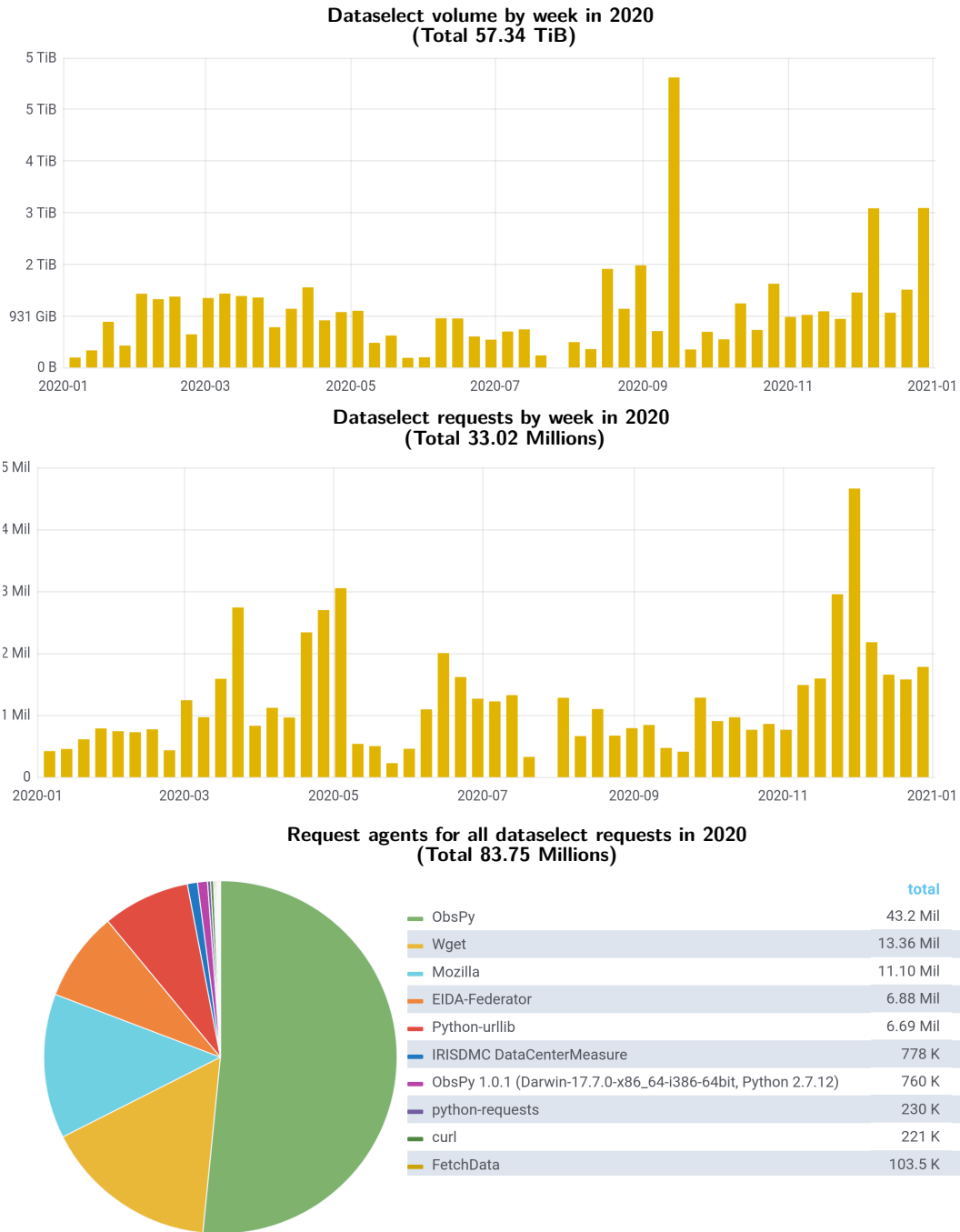


Figure 9: This dashboard view is an example of the web interface presenting some of the Résif-DC distribution metrics. Here, the figures focus on the `fdsnws-dataselect` web service during 2020, showing the volume sent (51.34 TB), the number of successful requests (33.02 millions) and the distribution of user agent requests for all the requests (83.75 millions). This last number accounts for the successful requests to the `/query` and `/queryauth` methods (HTTP 200), the requests returning no data (HTTP 204 and 404), the requests to other methods (documentation, `/auth`, wadl file), the requests producing errors due to improper parameters (HTTP 400 and 401) or insufficient permissions (HTTP 403).

Data requests to Résif-DC in 2020

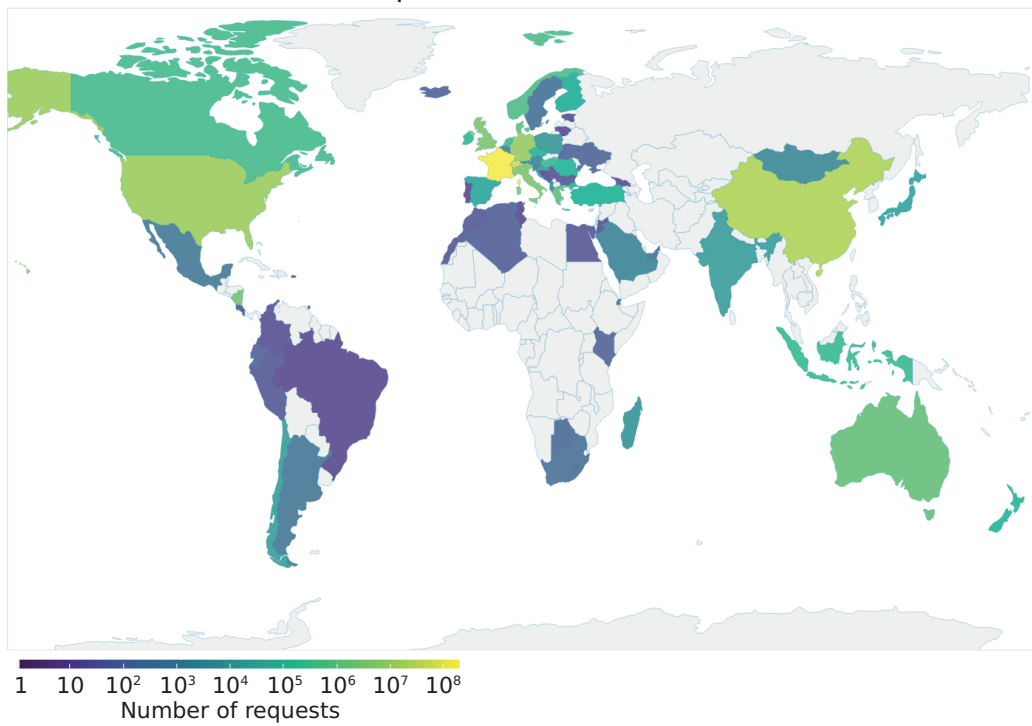


Figure 10: Geographical distribution of data requests to Résif-DC in 2020.

1067 **Appendix: Data access details and examples**

1068 In this appendix we present the different ways for end users to access Résif-SI data and metadata.

1069 **Real-time: SeedLink**

1070 Résif-DC real-time data service (SeedLink server, [GFZ and gempa GmbH, 2008](#)) provides access
1071 to data for the following permanent networks: FR, RD, RA, MT, G, WI, ND, CL, MQ, PF, GL. Data
1072 from temporary campaigns are usually not distributed in real-time unless acquired in the context of a
1073 post-seismic campaign for which Résif does not accept any data embargo time. The lifetime of the
1074 near real-time data in the internal memory buffer is about 3 hours. By convention, Résif-DC does not
1075 distribute any raw data for which the corresponding metadata is not available through the FDSN web
1076 service station.

1077 The SeedLink service is available at: `rtserve.resif.fr` (TCP port 18000).

1078 **Web services**

1079 Web services are the base delivery mechanism for most Résif data and metadata. Résif-DC exposes
1080 the four web services standardized by the FDSN ([FDSN, 2013](#)), an EIDA web service common to all
1081 EIDA DCs, as well as other more specific web services ([Résif, 2019a](#)).

1082 **FDSN web services**

1083 The FDSN web services are characterized by common and standardized interfaces and thus consti-
1084 tute a homogeneous means of access to data, regardless of the data center involved. Those services
1085 are all accessible with the prefix <https://ws.resif.fr/fdsnws/> (Table A1).

1086

1087 Examples:

- 1088 • All Résif stations, XML formatted: <https://ws.resif.fr/fdsnws/station/1/query?level=station>

- 1089 • One hour of data for network ZH (516 Mb): [https://ws.resif.fr/fdsnws/dataselect/1/query?](https://ws.resif.fr/fdsnws/dataselect/1/query?network=ZH&starttime=2003-06-01T00:00:00&endtime=2003-06-01T01:00:00)
1090 [network=ZH&starttime=2003-06-01T00:00:00&endtime=2003-06-01T01:00:00](https://ws.resif.fr/fdsnws/dataselect/1/query?network=ZH&starttime=2003-06-01T00:00:00&endtime=2003-06-01T01:00:00)
- 1091 • Available Data from network G, station SSB2, channel BHZ for a given time interval: [https://ws.](https://ws.resif.fr/fdsnws/availability/1/query?network=G&station=SSB&channel=BHZ&start=2010-02-23T00:00:00&end=2010-03-23T00:00:00)
1092 [resif.fr/fdsnws/availability/1/query?network=G&station=SSB&channel=BHZ&start=2010-02-23T00:](https://ws.resif.fr/fdsnws/availability/1/query?network=G&station=SSB&channel=BHZ&start=2010-02-23T00:00:00&end=2010-03-23T00:00:00)
1093 [00:00&end=2010-03-23T00:00:00](https://ws.resif.fr/fdsnws/availability/1/query?network=G&station=SSB&channel=BHZ&start=2010-02-23T00:00:00&end=2010-03-23T00:00:00)

1094 FDSN web services can be accessed by standard tools and libraries, like ObsPy (Beyreuther
1095 et al., 2010; Megies et al., 2011; Krischer et al., 2015). As an example, we provide in the Appendix
1096 an example Python script to retrieve data and metadata through the FDSN web services.

1097 **EIDA web services**

1098 Résif-DC exposes the EIDAWS *wfcatalog* web service (Trani et al., 2017). This web service provides
1099 detailed metrics on waveform data (e.g., gaps, overlaps, RMS, SEED data quality flag). The WFCata-
1100 log interface can be used for data discovery as it supports range filtering on all available metrics. The
1101 EIDAWS web service is accessible with the prefix <https://ws.resif.fr/eidaws/> (Table A2).

1102

1103 Examples:

- 1104 • waveform metadata documents with a daily granularity from network CL between 2018-01-
1105 01T00:00:00 and 2018-01-10T00:00:00, in JSON format: [https://ws.resif.fr/eidaws/wfcatalog/1/](https://ws.resif.fr/eidaws/wfcatalog/1/query?network=CL&include=sample&start=2018-01-01&end=2018-01-10)
1106 [query?network=CL&include=sample&start=2018-01-01&end=2018-01-10](https://ws.resif.fr/eidaws/wfcatalog/1/query?network=CL&include=sample&start=2018-01-01&end=2018-01-10)

1107 **Other web services**

1108 Résif-DC has implemented additional web services, whose functionalities have been inspired by the
1109 services provided by IRIS Data Services (Weertman, 2010) or needed by the French users of Résif-SI.
1110 The former are accessible with the prefix <https://ws.resif.fr/resifws/> while the latter are accessible
1111 with the prefix <https://ws.resif.fr/resifsi/>. Tables A3 and A4 show complete lists of these two groups of

1112 services.

1113

1114 Examples:

- 1115 • five minutes of data for station PYTO, channel HN2, with demean and response deconvolution,
1116 ASCII format: <https://ws.resif.fr/resifws/timeseries/1/query?net=RA&station=PYTO&cha=HN2&loc=02&demean&correct&start=2017-11-02T13:35:00&end=2017-11-02T13:40:00&format=ascii>
1117
- 1118 • interactive plot of five minutes of data for station PYTO, channel HN2, location code 02, with de-
1119 mean and response deconvolution: <https://ws.resif.fr/resifws/timeseriesplot/1/query?net=RA&station=PYTO&cha=HN2&loc=02&demean&correct&start=2017-11-02T13:35:00&end=2017-11-02T13:40:00&iplot>
1120
1121
- 1122 • plot of instrument response for station CIEL, network FR, channel HHE, location code 00: <https://ws.resif.fr/resifws/evalresp/1/query?net=FR&sta=CIEL&loc=00&cha=HHE&time=2017-01-01T00:00:00&format=plot>
1123
1124
- 1125 • response (RESP format) for station RUSF, channel ??N, location code 06: <https://ws.resif.fr/resifws/resp/1/query?net=FR&sta=RUSF&loc=06&cha=??N&time=2017-01-01T00:00:00&nodata=404>
1126
1127
- 1128 • data availability for an experiment stored as PH5 archive: <http://ws.resif.fr/resifws/ph5-availability/1/extent?net=3C>
1129
- 1130 • miniSEED data from a network stored as PH5 archive: <http://ws.resif.fr/resifws/ph5-dataselect/1/query?net=3C&starttime=2019-12-09T00:00:00&endtime=2019-12-09T00:00:10>
1131
- 1132 • quantity of data stored for year 2020 by network FR: <http://ws.resif.fr/resifws/statistics/1/query?request=storage&net=FR&type=validated&year=2020>
1133
- 1134 • quantity of data shipped via dataselect for network WI: <http://ws.resif.fr/resifws/statistics/1/query?request=send&media=dataselect&net=WI>
1135

1136 **Specificity of OBS data and metadata**

1137 Ocean Bottom Seismometer (OBS) data is archived both in time-corrected format (quality label
1138 "Q") and in original, unmodified format (quality label "D"). The following request (with no quality
1139 parameter specified), delivers both corrected (Q label) and raw data (D label):

```
1140 _____  
1141 wget --http-user=USER --http-password=PASSWORD "https://ws.  
1142     resif.fr/fdsnws/dataselect/1/queryauth?network=Z3&station=  
1143     A4*&starttime=2017-08-01&endtime=2017-08-01T01:00:00" -O  
1144     mydata.miniseed  
1145 _____
```

1146 By specifying the quality=Q parameter, only corrected data (Q label) is retrieved:

```
1147 _____  
1148 wget --http-user=USER --http-password=PASSWORD "https://ws.  
1149     resif.fr/fdsnws/dataselect/1/queryauth?network=Z3&station=  
1150     A4*&starttime=2017-08-01&endtime=2017-08-01T01:00:00&  
1151     quality=Q" -O mydata.miniseed  
1152 _____
```

1153 With quality=D, only raw data (uncorrected) is delivered.

1154

1155 The StationXML metadata associated with the OBS are progressively refined with information on
1156 time corrections, instrument positioning uncertainties, water column depth and precise instrument
1157 description (sensor, data logger). A typical example of OBS metadata at Résif-DC at the station level
1158 is the following:

```
1159 _____  
1160 <Station code="A401A" startDate="2017-06-22T01:59:00" endDate  
1161     ="2018-02-16T00:01:00" restrictedStatus="closed" resif:  
1162     alternateNetworkCodes="Z32015">
```

```

1163     <Comment><Value>{"clock_correction": {
1164         "linear_drift": {
1165             "time_base": "Seascan MCX0, ~1e-8 nominal drift",
1166             "reference": "GPS", "start_sync_instrument": 0,
1167             "start_sync_reference": "2017-06-21T06:36:00Z",
1168             "end_sync_reference": "2018-02-16T01:44:00.146Z",
1169             "end_sync_instrument": "2018-02-16T01:44:00.00Z"
1170         }
1171     }}</Value></Comment>
1172     <Latitude minusError="4.5e-05" plusError="4.5e-05" unit="
1173         DEGREES">42.647</Latitude>
1174     <Longitude minusError="6.12e-05" plusError="6.12e-05"
1175         unit="DEGREES">5.0198</Longitude>
1176     <Elevation minusError="10" plusError="10" unit="METERS">
1177         -1619.0</Elevation>
1178     <Site><Name>N/W Mediterranean Sea</Name></Site>
1179     <Vault><Name>Sea floor</Name></Vault>
1180     <WaterLevel minusError="10" plusError="10" unit="METERS">
1181         1619.</WaterLevel>
1182 </Station>
1183

```

Table A1: The FDSN web services provided by Résif-DC, with base URL <https://ws.resif.fr/fdsnws/>. Note that the event web service is operated by RéNaSS (Réseau National de Surveillance Sismique – National Seismic Monitoring Network).

<i>web service/version</i>	<i>comment</i>	<i>format</i>
station/1	metadata inventory	StationXML or text
dataselect/1	time series data	miniSEED
availability/1	data inventory	text, json, geocsv
event/1	event parameters	quakeML

Table A2: The EIDAWS web services provided by Résif-DC, with base URL <https://ws.resif.fr/eidaws/>.

<i>web service/version</i>	<i>comment</i>	<i>format</i>
wfcatalog/1	data quality and metrics	json

Table A3: The RESIFWS web services provided by Résif-DC, with base URL <https://ws.resif.fr/resifws/>.

Note: ph5-dataselect and ph5-availability comply to the FDSN Web Services specification (FDSN, 2013) and might therefore be migrated in the future to a standard URL. For up-to-date information on Résif web service URLs, please refer to <https://ws.resif.fr>.

<i>web service/version</i>	<i>comment</i>	<i>format</i>
timeseries/1	time series data	miniSEED, sac, text, png
timeseriesplot/1	time series data	jpeg or png
ph5-dataselect/1	time series data	miniSEED, sac, geocsv (converted from PH5)
ph5-availability/1	data inventory	text, json, geocsv
evalresp/1	response information evaluated from Résif metadata	text or png
resp/1	channel response information	resp (text)
sacpz/1	channel response information	sacpz (text)
statistics/1	data download statistics from Résif-DC	text

Table A4: The web services provided by Résif-DC for data providers, with base URL <https://ws.resif.fr/resifsi/>.

<i>web service/version</i>	<i>comment</i>	<i>format</i>
transaction/1	status of an data or metadata integration transaction; this service is for Résif data providers	text, json
orphanfile/1	Data (SDS files) without metadata in Résif archive; this service is for Résif data providers	text
assembleddata/1	historical and validated collections of the permanent accelerometric network (1995-2013)	tar.gz