

---

# DMP du projet SCALER (Smart SCALing for MicrosEvice ARchitectures)

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

## Plan Details

<b>Plan title</b>	DMP du projet SCALER (Smart SCALing for MicrosEvice ARchitectures)
<b>Version</b>	First version
<b>Plan purpose/scope</b>	This plan defines the origin, management and evolution of the data to be used in the SCALER project. This data is either to be generated withing the SCALER's experimentations, or to be provided by the industrial partners. This is to be done with respect to the consortium agreement.
<b>Fields of science and technology (from OECD classification)</b>	Computer and information sciences
<b>Language</b>	eng
<b>Creation date</b>	2023-04-04
<b>Last modification date</b>	2023-06-30
<b>Identifier type</b>	DOI
<b>License</b>	Creative Commons Attribution Non Commercial 4.0 International

## Project Details

<b>Project title</b>	Smart SCALing for MicrosEvice ARchitectures
<b>Acronym</b>	SCALER

## Abstract

The last decade has acknowledged the emergence of a new paradigm for building large-scale distributed systems: the microservice architecture. As opposed to the classical tightly coupled monolith architectures, this paradigm promotes small loosely coupled autonomous services, so called microservices. Such software services are to be small to encapsulate a well-defined functionality and thus facilitate bug isolation and maintenance. They should be loosely coupled to support independent software updates. Finally, they need to be autonomous to enable independent delivery to production. Adopting microservices becomes a general trend and experts predict that "by 2022, 90% of all new apps will feature microservices architectures". Among companies working around microservices are infrastructure providers, including GAFAM and EOLAS, major IT players, including Netflix and Uber, but also telco operators, including Orange. They all build on the container technology to facilitate functionality encapsulation, dependency isolation and standalone deployment, as required by microservices. However, the current used orchestration solutions (e.g., Kubernetes) struggle with the scale and the heterogeneity of microservices and fail to fix the performance problems fast enough. The challenge addressed by the SCALER project is to optimize the scaling of microservice-based networked services while satisfying their stringent IT and telco requirements. The proposal targets three research aspects, namely automatic microservice characterization, identification of microservice interaction patterns and microservice smart-scaling. SCALER brings together two academical and two industrial partners, namely LIG ERODS, INRIA Spirals, Orange Innovation and EOLAS B&D.

## Funding

- French National Research Agency : ANR-22-CE25-0009

## Start date

2023-01-01

## End date

2027-06-30

## Partners

- Orange Business Services SA, EOLAS ()
- Orange SA, Innovation ()
- Institut national de la recherche en informatique et automatique (200818245B)
- UGA,LIG (200711886U)

## Research outputs :

1. Default research output

## Contributors

Name	Affiliation	Roles
MARANGOZOVA Vania	LIG	<ul style="list-style-type: none"><li>• DMP manager</li><li>• Data contact</li><li>• Project coordinator</li></ul>

# DMP du projet SCALER (Smart SCALing for MicroService ARchitectures)

---

## 1. Data description and collection or re-use of existing data

### 1a. How will new data be collected or produced and/or how will existing data be re-used?

- **New Data**

New data will be produced by the people involved in the SCALER project during their software experimental work. Researchers and students will work with different software, typically involving microservice-based applications deployed on Kubernetes and will generate execution traces (through monitoring, logging or profiling). These traces will be analyzed and the results will be either analytical, or visualisations. New SCALER data will as a consequence include the newly produced execution traces, as well as the results from the analysis of available data.

- **Existing Data**

As industrial partners working with clients, ORANGE Innovation and ORANGE Eolas have access to production data i.e. coming from real world applications. These data will be valuable for validating SCALER's developments. The access to this data is to be submitted to the internal industrial privacy regulations, will be anonymized and will be done with respect the CONSORTIUM AGREEMENT.

---

### 1b. What data (for example the kind, formats, and volumes), will be collected or produced?

- **Types of Data**

- Execution traces
- Execution logs
- Execution profiling data
- Existing databases with usage data

- **Formats**

- Textual data, possibly formatted using XML/CSV/JSON or similar.
  - Database data
  - Graphics representing analysis results and/or scripts/programs for creating graphics from analysis data
  - Structured data represented in a suitable format (e.g. graph representations)
- 

## 2. Documentation and data quality

### 2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

The data that will be produced will come with

- an explicit description of the configuration i.e. hardware and software information about the environment in which it has been produced
  - a description of the format and the volume of the data
  - a description of the experience the data corresponds to
- 

### 2b. What data quality control measures will be used?

- Peer review
  - Verify data integrity when values are transformed using scripted or automated processes
  - Rerun experiments when data presents inconsistent values
  - Reproduce experiments multiple times
- 

### 3. Storage and backup during the research process

#### 3a. How will data and metadata be stored and backed up during the research?

SCALER's newly produced data will be stored using the GRICAD platform (<https://gricad-gitlab.univ-grenoble-alpes.fr/>). This will ensure availability, perenity and version management.

---

#### 3b. How will data security and protection of sensitive data be taken care during the research

The only sensitive data that could be used in SCALER could come from usage related data collected by ORANGE (related either to phone usage, or to the cloud usage of EOLAS). If such cases present themselves, data will respect the requirements for data protection and anonymization, as defined by the industrial partner detaining the data.

---

### 4. Legal and ethical requirements, code of conduct

#### 4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

see question 3b

---

#### 4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

There is a consortium agreement under writing, respecting French law.

---

#### 4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

Open science

---

### 5. Data sharing and long-term preservation

#### 5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

SCALER's datasets will be published via the PerSCiDO platform (<https://perscido.univ-grenoble-alpes.fr/>) which is hosted on GRICAD. PerSCiDO ensures easy access to multiple datasets through advanced research criteria and provides unique identifiers (DOI).

---

**5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**

GRICAD & PERSCIDO infrastructure (question 3a)

---

**5c. What methods or software tools are needed to access and use data?**

Perscido is a data repository, so data will be directly available.  
No particular software tools are planned for now but this aspects is subject to evolution.

---

**5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

By the perscido platform.

---

## **6. Data management responsibilities and resources**

**6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

All the partners will share the responsibility for data management. The data steward is the scientific leader of the project i.e. UGA.

---

**6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

The financial resources are provided under the form of the data management platform available through UGA.  
The time resources are part of the research time of the people involved in the SCALER project.