



ARCHIVOR

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

PCP Phase 1 - Contractor details and project abstracts

Document type : Information for European Commission

Document version : Final

Document Preparation Date : 30.08.2020



Classification : Public

Author(s) : All project partners

File Name : PCP Phase 1 Contractor details and project abstracts



Complete the following table for every contractor that is awarded a PPI, PCP Phase 1, PCP phase 2 or PCP phase 3 contract.

Contact Details	Type/ size of legal entity	Place of performance of contract activities	Logo
<p><u>Main contractor</u></p> <p>ARKIVUM</p> <p>The Walbrook Building, 25 Walbrook, London EC4N 8AF, UK</p> <p>Yaron Naor yaron.naor@arkivum.com +447912756840</p>	SME	<p>% of contract value allocated to main contractor: 100%¹</p> <p>% of activities for the contract performed by the main contractor in EU Member States or countries associated with Horizon 2020: 100%</p>	
<p><u>Other consortium Member(s) (if applicable)</u></p> <p>Google</p> <p>1600 Amphitheatre Parkway, Mountain View, CA 94043, Etats-Unis.</p> <p>Alex Schroeder alschroeder@google.com +49 40 808179480</p>	Larger company	<p>% of contract value allocated to contractor Google: 0%</p> <p>% of activities for the contract performed by contractor Google in EU Member States or countries associated with Horizon 2020: NA</p>	
<p>Project Abstract (+/- 1000 characters maximum)</p> <p>About Arkivum:</p> <p>Arkivum provides a long-term data management solution that includes the safeguarding, digital preservation and online access of digital content. The solution is delivered as a fully managed service, supports a wide range of content types, and can be deployed onto public or private cloud.</p>			

¹ This includes all other items listed in the Financial Offer.



Arkivum is recognised internationally for its expertise in the archiving and digital preservation of valuable data and digitised assets in large volumes and multiple formats. The long-term security, integrity and accessibility of data is crucial for all of our customers and partners, who share a commitment to good practice in its stewardship and governance.

The solution has been selected by organisations across a range of sectors, including life sciences, corporate archives, research organisations (including higher education) and memory institutions in culture and heritage. With the aid of the Arkivum solution, each customer is able to derive optimum long-term value from their data, collections and intellectual property. In addition, they benefit from 100% data integrity guarantee. Headquartered in the UK, with US offices in Boston and New York City, Arkivum advocates and adheres to best practice in research data management, including the FAIR principles for data: Findable, Accessible, Interoperable, Reusable. In particular, Arkivum helps organisations to build and operate Trusted Digital Repositories for their content.

Proposed solution:

Partnering with Google in the project, Arkivum will develop a solution for long-term data management and online access, to be deployed on the Google Cloud Platform (GCP). While Arkivum's particular expertise lies in reliably keeping large volumes of complex archive data secure and accessible, thus maximising compliance, insight and discovery, the new collaboration with Google Cloud enables the team to manage data on the immense scale required by ARCHIVER. The Arkivum solution and workflows are all about enabling content, which in the case of ARCHIVER includes scientific data from a range of sources, to be captured, ingested, preserved and made accessible to those who need to use that content in the future.

This includes the ability to ingest, validate, organize and manage content as it comes into an archive. The content will then go through appropriate preservation and safeguarding processes, including generating OAIS archiving packages to ensure it is properly protected and remains usable. This will also ensure that the data is searchable, discoverable and accessible for users both today and in the future so people can find and use the content in the archive that they need, when they need it.

This in turn helps research organisations achieve their research data management objectives, for example ensuring that scientific datasets are Findable Accessible Interoperable and Reusable.

Meeting the ARCHVIER requirements:

The combined Arkivum and GCP solution will meet all layers of the ARCHIVER project requirements;

- In layer 1 of ARCHIVER (storage/basic archiving/secure backup), GCP fulfils the requirements of high-volume data storage with fast ingest and access, all to the petabyte level range.
- In layer 2 (preservation) Arkivum's solution addresses the need for long-term digital preservation following the OAIS model, including obsolescence management, file fixity and authenticity checks, and packaging for preservation and access.
- In layer 3 (baseline user services), the Arkivum solution provides the ability to search, share and index large and complex research datasets
- Then finally in layer 4 (advanced services), GCP provides the basis of hosting and running scientific applications that can be executed against archived and preserved datasets.



In short, we will be combining and extending the solutions from Arkivum and Google to deliver the full ARCHIVER requirement stack.



R&D:

The first phase of the project will focus on the design/architecture of a long-term digital preservation and archiving solution to meet the Buyer requirements. This specifically includes:

- Using GCP to meet the baseline requirements of supporting very large datasets, with high speed ingest and access, but also with the ability to host and run scientific applications against this data.
- Using digital preservation and data archiving services to meet the requirements for OAIS and to provide organisations with a hosted solution for operating their Trusted Digital Repositories.
- Ensuring that research data can be described, organized, sliced/diced, tagged and published in a flexible way that meets FAIR principles.
- Total Cost of Service models and the ability to optimize cost against data volumes, access frequencies, data safety, data processing and retention periods.
- Pervasive use of open standards, open specifications, open source and open APIs to ensure portability, interoperability, exit strategies and removal of vendor and solution lock-in.
- Detailed models and commercialization plan that includes Service Level Agreements,



User Support, Licensing, Service Configuration and Pricing for new commercial services based on ARCHIVER to be provided to the Buyers Group, the European Open Science Cloud (EOSC) and beyond.



Contact Details	Type/ size of legal entity	Place of performance of contract activities	Logo
<p><u>Main contractor</u></p> <p>GMV Soluciones Globales Internet S.A.U</p> <p>P.T. Boecillo, Juan de Herrera, nº 17, 47151 Boecillo, Valladolid, Spain.</p> <p>Nuria Gómez Rojo</p> <p>nngomez@gmv.com</p> <p>+34 600 40 84 68</p>	Large company	<p>% of contract value allocated to main contractor: 96%²</p> <p>% of activities for the contract performed by the main contractor in EU Member States or countries associated with Horizon 2020: 100 %</p>	
<p><u>Other consortium Member(s) (if applicable)</u></p> <p>Piql AS</p> <p>Grønland 56, 3045 Drammen, Norway</p> <p>Roberto González</p> <p>roberto.gonzalez piql.com</p> <p>+34 660 109 686</p>	SME	<p>% of contract value allocated to contractor Piql: 4%</p> <p>% of activities for the contract performed by contractor Piql in EU Member States or countries associated with Horizon 2020: 100%</p>	
<p><u>Subcontractors (if applicable)</u></p>		<p>% of contract value allocated to subcontractor Blue Safespring: 0%</p>	

² This includes all other items listed in the Financial Offer.



<p>Blue Safespring AB Smidesvägen 12 171 41 Solna, Sweden Fredric Wallsten fredric.wallsten@safespring.com +46766292520</p>	<p>SME</p>	<p>% of activities for the contract performed by subcontractor Blue Safespring in EU Member States or countries associated with Horizon 2020: NA</p>	
<p>Amazon Web Services Avenue des Arts 27, 1000 Bruxelles, Belgique Christian Zahorski-Philippe chriszp@amazon.com +3280319408</p>	<p>Large company</p>	<p>% of contract value allocated to contractor AWS: 0%</p> <p>% of activities for the contract performed by contractor AWS in EU Member States or countries associated with Horizon 2020: NA</p>	

Project Abstract (+/- 1000 characters maximum)

The GMV & Piql consortium

The consortium is composed by GVM and Piql as main members and AWS and Safespring as subcontractors providing cloud resources. Piql's main goal, as a company involved in the archiving business, is to define the preservation requirements and the architecture of the service and GMV's, as an expert company in engineering and ICT technologies, is to create a layer of services on top and below the preservation core.

Proposed solution



Our solution has been divided in 4 layers to address the different levels of design. The key features of the proposed solution fore are described below:

- **Layer 1 (storage, basic archive and secure backup capabilities):** the solution will provide features for multi-cloud deployment (over private, public, hybrid, community and special purpose Clouds) where the buyer will be able to choose which cloud provider wants to use for his storage. Security will be implemented from the scratch and grouped in clearly defined areas that help to achieve requirements and compliance requisites, such as Secure Coding, Secure CD/CI and Secure Infrastructure. Our consortia knowledge and frameworks covers NIST, ISO27000 and European Cybersecurity Act, to protect stored, transmitted or otherwise processed data against accidental or unauthorised storage, processing, access or disclosure during the entire life cycle of the ICT product.
- **Layer 2 (preservation capabilities):** the solution will provide the preservation of files, folders, content management packages (archives) and data types, as well as emails backup in pdf format. Migration and backup mechanisms, APIs, SFTPs, web interfaces, drag&drop ingestions, and metadata and context processing, will be also provided. The preservation services will rely on Archivematica, designed around OAIS requirements, which should be extended due to the high rate, high volume requirements. Piql AMU is used in case of very valuable data need to be kept for 100+ years on PiqlFilm but it should be improved because of the required data sizes. The user could decide which media to be used for the storage of the data among the three available, tape, hard-disk and PiqlFilm, depending on its needs.
- **Layer 3 (user services):** The solution will allow the operation with volumes of hundreds of TBs with support of indexing, elastic search, deduplication, single point access, crawling, cross-checking, vulnerability scanning, and plugins configuration filtering potential datasets rapidly, to access dataset metadata and decide on its relevance (e.g. citation purposes or reusing a dataset). Automated metadata indexing for several tens of PB content will be supported, including also support for dataset filtering and information tagging, aiming at maximum interoperability and easy and broader searches for the research community. Piql Connect and Piql UI will simplify the user interaction with the whole system. Access and permission management against repositories and various collections will be also supported by the Federated Identity and Access Management schemes, as a strategy to promote open data access in the research community.
- **Layer 4 (advanced services capabilities):** the solution will provide features for the retention and integrity of data over a decade ensuring its tamper-proof behaviour whilst allowing easy access and basic re-usability with a collection of artificial intelligence functionalities. The solution will provide the ability for a researcher to replicate a computational experiment that was done by someone else, using the same software and data. Reproducibility of experiments will be based on container-based workflows exposed to the user taking advantage of modern cloud computing paradigms that provide scale on-demand, necessary to run large scientific experiments.



Contact Details	Type/ size of legal entity	Place of performance of contract activities	Logo
<p><u>Main contractor</u></p> <p>LIBNOVA SL</p> <p>Paseo de la Castellana, 153. Madrid (Spain) Teofilo Redondo teo.redondo@libnova.com +34 649 21 17 27</p> <p><u>Other consortium Member(s) (if applicable)</u></p> <p>CSIC</p> <p>Joaquin Costa 22 – 28002. Madrid (Spain) Fernando Aguilar aguilarf@ifca.unican.es Office number: +34 942206779 Mobile: +34 659 642476</p>	<p>SME</p> <p>Public organism</p>	<p>% of contract value allocated to main contractor: 64%³</p> <p>% of activities for the contract performed by the main contractor in EU Member States or countries associated with Horizon 2020: 100%</p> <p>% of contract value allocated to contractor CSIC: 24%</p> <p>% of activities for the contract performed by contractor CSIC in EU Member States or countries associated with Horizon 2020: 100 %</p>	 

³ This includes all other items listed in the Financial Offer.



<p>Universitat de Barcelona</p> <p>Gran Via de les Corts Catalanes, 585 – 08007. Barcelona (Spain)</p> <p>Miquel Termens termens@ub.edu</p> <p>Office phone: +34 934035773 Mobile: +34 651052174</p>	<p>University</p>	<p>% of contract value allocated to contractor Universitat de Barcelona: 12%</p> <p>% of activities for the contract performed by contractor Universitat de Barcelona in EU Member States or countries associated with Horizon 2020: 100%</p>	
<p><u>Subcontractors (if applicable)</u></p>			
<p>Giaretta Associates</p> <p>24 Cornwall Road, Dorchester, Dorset,</p> <p>David Giaretta david@giaretta.org</p> <p>Land line: +44-1963-230990 Mobile: +44-7770-326304</p>	<p>SME, larger company, natural person, university / research institute, other</p>	<p>% of contract value allocated to subcontractor Giaretta Associates: 0%</p> <p>% of activities for the contract performed by subcontractor Giaretta Associates in EU Member States or countries associated with Horizon 2020: NA</p>	

Project Abstract (+/- 1000 characters maximum)

Short description of the consortium:

The consortium is composed of the following institutions:

LIBNOVA – a company centered around digital preservation technology, with more than 10-years history and customers all over the world. LIBNOVA’s portfolio covers all the needs to curate and preserve both



cultural heritage (libraries, archives, museums) as well as scientific datasets (universities and research institutions). LIBNOVA Research Labs (2017) manages all research initiatives for the company.

CSIC – IFCA – the Spanish National Research Council (in this project through IFCA – Instituto de Física de Cantabria) is the main agent of the Spanish System for Science, Technology and Innovation. CSIC generates knowledge through scientific and technical research, transfers results from research (to boost and create technology-based enterprises), and manages large facilities and unique scientific and technical infrastructures.

University of Barcelona – is the foremost public institution of higher education in Catalonia, and also the principal centre of university research in Spain. A reference European institution for research activity, both in terms of the number of research programmes and the excellence achieved.

Giaretta Associates – David Giaretta has worked in digital preservation since 1990, and was chairman of the OAIS Reference Model (ISO 14721), the “de facto” standard for building digital archives. He leads the group which produced the ISO standard for audit and certification of trustworthy digital repositories (ISO 16363).

Description of the proposed solution:

The solution we are proposing is built on pre-existing digital preservation platforms already in use by many leading organizations across the world. It proposes a solution for the whole organization and for the whole data life-cycle, completely aligned with OAIS, ISO16363, FAIR and TRUST principles, with powerful and really innovative capabilities in all four functionality layers.

The Research, Management and Preservation Platform will combine existing technologies and new components, to solve obstacles for research dataset management (including preservation) identified in the Archiver project.

Five areas comprise the architecture:


- Containers – keep content accessible with several protocols, organized and protected. These containers keep metadata, data and code together to ensure usability (OAIS-aligned).
- Dynamic Insights – help users when dealing with personal information, digital preservation and emissions reduction, with the following components: Data Policies Assistant, GDPR Assistant, Emissions Optimizer, Digital Preservation
- Budget assistant – helps users to plan and follow expenditures
- Content gateway – connects the platform with repositories for discovery solutions, such as Invenio or Dataverse
- Digital Preservation, OAIS and FAIR conformance – as support for the OAIS Information Model and for the Mandatory Responsibilities, and the results will fully support repositories in OAIS conformance. The focus on usability is also critical for the “Interoperability” and “Reusability” required by the FAIR principles.

A brief summary of areas of work follows:

- Not only the type of media to preserve but also the data center, location, etc. and the Quality of Service (QoS).



- A key functionality will be to manage the full Representation Information Network required by the OAIS Information Model. Where possible the import of complete metadata schema automatically will be supported
 - The API will support search, retrieve data and metadata.
 - Connection for data publishing in repositories (i.e. Zenodo - CERN).
 - Provenance, reproducibility or processing are also areas of great importance for scientific data
- The resulting product used within a repository will be ISO 16363 audited.

Contact Details	Type/ size of legal entity	Place of performance of contract activities	Logo
<p><u>Main contractor</u></p> <p>RHEA SYSTEM S.p.A.</p> <p>La Piramide, Via di Grotte Portella 6/8, 00044 Frascati, Italy</p> <p>Marco Marigliano m.marigliano@rheagroup.com +39 3488622524</p>	<p>Large company</p>	<p>% of contract value allocated to main contractor: 54%⁴</p> <p>% of activities for the contract performed by the main contractor in EU Member States or countries associated with Horizon 2020: 100%</p>	

⁴ This includes all other items listed in the Financial Offer.



<p><u>Other consortium Member(s)</u> <u>(if applicable)</u></p> <p>DEDAGROUP S.p.A. Via di Spini 50, 38121 Trento, Italy</p> <p>Fabrizio Della Bruna Fabrizio- dellabruna@dedagroup.it +39335233487</p>	<p>Large company</p>	<p>% of contract value allocated to contractor DEDAGROUP: 36 %</p> <p>% of activities for the contract performed by contractor DEDAGROUP in EU Member States or countries associated with Horizon 2020: 100%</p>	
<p><u>Subcontractors (if applicable)</u></p> <p>GTT Switzerland SARL Chemin de L'Epinglier 2, 1217 Meyrin, Switzerland</p> <p>Luigi Spinillo Luigi.Spinillo@gtt.net +39 335 7271341</p> <p>Rahim Bengrine Rahim.Bengrine@gtt.net +41.79.267.29.54</p>	<p>Large company</p>	<p>% of contract value allocated to subcontractor GTT: 10%</p> <p>% of activities for the contract performed by subcontractor GTT in EU Member States or countries associated with Horizon 2020: 100%</p>	



Project Abstract (+/- 1000 characters maximum)

The Consortium is composed of RHEA as Lead Tenderer, DEDAGROUP as Consortium member, and GTT as RHEA's subcontractor.

Excellence of the proposed solution:

Ambition

The Consortium ambition is to contribute to the R&D activities for designing, implementation and deploying of a complete and secure service model to fully cover the stewardship lifecycle of multiple and heterogeneous research data exceeding the occurring lacks in these concepts, involving organisation, processes and technologies. The Consortium will consolidate the achieved outcomes creating relevant Best Practices.

The Consortium has long and strong experience and competence in Data archiving and Long Term Preservation, specifically in:

- Space (Science and Earth Observation);
- Life Science (Pharmaceutical, Biotech, Clinical Research Organization);
- Libraries and Historical Archives.

The Consortium expertise is focused on:

- Scientific and Space data;
- Clinical data and Records;
- Confidential, Reserved and Unstructured data;
- Historical and Unique data.

The solution is therefore composed of open source applications and a dedicated hybrid cloud in order to grant the whole integrated and fully protected data stewardship and curation lifecycle. The solution assures full adherence to the EOSC guidelines and specifications and readiness for any kind of future Early Adopters.

Solution description

The proposed solution is in line with the vision and scientific scope of the ARCHIVER project and addresses the challenges set forth by the ARCHIVER project applying a methodology gained through extensive and proven experiences in R&D processes, long-term secure data archiving and preservation, cloud infrastructures and services management in several industries and European agencies with heterogeneous datasets, needs and requirements.

The Consortium (RHEA, DEDAGROUP and GTT) is fully confident of its complementary key competences and experiences allowing to develop innovative and robust services covering all ARCHIVER requirements for Organizational Change Management and Governance, Service Business and Cost Model, Service Processes, Technical Solutions Design and Implementation. The solution addresses business requirements raised by the Buyer Group under strong regulations




(Energy and Life Sciences domain) and supports testing and validations of IT infrastructures, applications, services and change control processes.

The high level architecture components of the solution are the following:



- Secure Service Portal (Identity Access Management, Access Layer Interface, Validation and Pre-ingestion services), implemented by RHEA;
- Existing and mature Open Source platforms for data archiving, preservation, reporting and access/discovery (Archivematica/AToM/JasperSoft), integrated and configured by DEDAGROUP;
- Readiness XaaS services, provided by all partners;
- Cloud connect product for integration with proposed robust and scalable managed Hybrid Cloud, provided by GTT.

The complete and detailed solution design and architecture will be provided as deliverable of Phase-1 as requested.

The proposed foundation will accelerate the R&D prototyping phases in order to incrementally provide the Buyers with a robust and scalable service model covering all ARCHIVER's objectives and Use Cases.

Contact Details	Type/ size of legal entity	Place of performance of contract activities	Logo
<p><u>Main contractor</u></p> <p>T-Systems International GmbH</p> <p>Heinrich-Hertz-Str. 1 D-64295 Darmstadt Germany</p> <p>Jurry de la Mar</p> <p>Jurry.delaMar@t-systems.com</p> <p>+49 6151 5833347 (T)</p> <p>+49 170 9236058 (M)</p>	<p>Large company</p>	<p>% of contract value allocated to main contractor: 49%</p> <p>% of activities for the contract performed by the main contractor in EU Member States or countries associated with Horizon 2020: 100%</p>	



<p><u>Subcontractors (if applicable)</u></p> <p>GWDG – (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen)</p> <p>Am Faßberg 11 D-37077 Göttingen Germany</p> <p>Ramin Yahyapour ramin.yahyapour@gwdg.de Tel.: +49 551 201-1545 Mobile: +49 179 5261973</p>	<p>SME</p>	<p>% of contract value allocated to subcontractor GWDG: 26%</p> <p>% of activities for the contract performed by subcontractor GWDG in EU Member States or countries associated with Horizon 2020: 100%</p>	
<p><i>Onedata.io S.A.</i></p> <p><i>ul. Lea 114 30-133 Krakow Poland</i></p> <p><i>Lukasz Dutka lukasz.dutka@onedata.org +48 608 62 77 12</i></p>	<p>SME</p>	<p>% of contract value allocated to subcontractor Onedata: 25%</p> <p>% of activities for the contract performed by subcontractor Onedata in EU Member States or countries associated with Horizon 2020: 100%</p>	



Project Abstract (+/- 1000 characters maximum)

To meet the challenge of the Archiver PCP to create cost-effective preservation services for the European science community, T-Systems' approach is to provide an OAIS-compliant solution that is open, easy-to-use, extendable, cost- and energy-efficient.

The solution follows a full open-source and cloud-agnostic approach, building on pre-existing and proven components for data preservation and workflow management. The core components include Archivematica, Onedata and Flowable. The modular approach is supported by a large set of APIs that will enable users to extend and integrate the components with other preferred services.

The R&D will focus on new innovative functionality for baseline and advanced data preservation services, including Petabyte-scale storage options, compliance with OAIS, PREMIS, METS and BagIT standards and new innovate functions for distributed data and workflow management, search and discovery, data representation and scientific analysis.

The objective is an integrated service offer for end-users, integrators and cloud providers for deployments on local and public cloud infrastructures. T-Systems will operate the services as part of its Open Telekom Cloud portfolio, a leading European public cloud service based on OpenStack. GWDG will extend its portfolio with the service offer for its established public and academic community. Both will support other candidates that will adapt the solution and join the community.

New business models will be evaluated for public sector organizations, enabling competitive procurements of preservation services while taking full advantage of the flexibility, scalability and competitive pricing that commercial and academic cloud providers can offer. The project will support the European Open Science Cloud initiative, to eventually provide Europe's 1.7 million researchers and 70 million science and technology professionals with preservation services and re-use of research data across borders and scientific disciplines.

T-Systems is the lead contractor from Germany and will be supported by subcontractor GWDG, an academic service provider from Germany and Onedata, an SME from Poland.

Is the project based on / a continuation of R&D activities that were previously funded by the EU?: Yes
If yes, identify this EU funding: H2020-ICT-2015, HNSciCloud, 687614

