

Technical Specifications (In-Cash Procurement)

Data Handling Extensions and maintenance framework contracts

this is the technical specifications for the data handling extensions and maintenance framework : it describes the various components of the data handling system.

SERVICE

Table of Contents

1	PREAMBLE.....	3
2	PURPOSE.....	3
3	ACRONYMS & DEFINITIONS	3
3.1	Acronyms.....	3
3.2	Definitions.....	3
4	APPLICABLE DOCUMENTS & CODES AND STANDARDS.....	4
4.1	Applicable Documents.....	4
4.2	Applicable Codes and Standards.....	5
5	CONTRACT DURATION	5
6	SCOPE OF WORK.....	5
6.1	Data Archiving system.....	6
6.1.1	Description.....	6
6.1.2	Service Duration	7
6.2	Data Access.....	7
6.2.1	Description.....	7
6.2.2	Service duration	8
6.3	Data Processors and Mirroring	8
6.3.1	Description.....	8
6.3.2	Service duration	9
6.4	Data visualization.....	9
6.4.1	Description.....	9
6.4.2	Service duration	9
6.5	Data management.....	9
6.5.1	Description.....	9
6.5.2	Service duration	10
7	LOCATION FOR SCOPE OF WORK EXECUTION	10
8	IO DOCUMENTS.....	10
9	LIST OF DELIVERABLES AND DUE DATES	10
10	QUALITY ASSURANCE REQUIREMENTS.....	10
11	SAFETY REQUIREMENTS.....	10
11.1	Nuclear class Safety.....	10
11.2	Seismic class	10
12	SPECIAL MANAGEMENT REQUIREMENTS	10

	SERVICE	
12.1	Contract Gates.....	10
12.2	CAD design requirements.....	10
12.3	Services-specific requirements	11
12.3.1	Common to all.....	11
12.3.2	Data Archiving system.....	11
12.3.3	Data Access.....	11
12.3.4	Data Processors	12
12.3.5	Data visualization.....	12
12.3.6	Data management.....	12
12.4	Skills and qualifications	12
13	APPENDICES	14

SERVICE

1 Preamble

This Technical Specification is to be read in combination with the General Management Specification for Service and Supply (GM3S) – [Ref 1] that constitutes a full part of the technical requirements.

In case of conflict, the content of the Technical Specification supersedes the content of Ref [1].

2 Purpose

The document defines the technical services needed for the maintenance and extensions of the data handling ecosystem and the data visualization tools. The main components of the data handling ecosystem consist of the data archivers, data access and data processors.

This Contract is a framework contract, where each task order is a freestanding engineering activity subject to its own technical specification and statement of work, and with its own budget.

3 Acronyms & Definitions

3.1 Acronyms

The following acronyms are the main one relevant to this document.

Abbreviation	Description
CRO	Contract Responsible Officer
GM3S	General Management Specification for Service and Supply
IO	ITER Organization
PRO	Procurement Responsible Officer
DAN	Data Archiving Network
HDF5	Hierarchical Data Format v 5
HDF5-SWMR	HDF5 Single Writer Multiple Reader
REST	REpresentational State Transfer
PON	Plant Operation Network
SDN	Synchronous Data Network
UDP	User Datagram Protocol
IMAS	Integrated Modelling Analysis Software

3.2 Definitions

For a complete list of CODAC definitions, see ITER CODAC Glossary, Ref [2].

Contractor: shall mean an economic operator who have signed the Contract in which this document is referenced.

SERVICE

4 Applicable Documents & Codes and standards

4.1 Applicable Documents

This is the responsibility of the Contractor to identify and request for any documents that would not have been transmitted by IO, including the below list of reference documents.

This Technical Specification takes precedence over the referenced documents. In case of conflicting information, this is the responsibility of the contractor to seek clarification from IO.

Upon notification of any revision of the applicable document transmitted officially to the contractor, the contractor shall advise within 4 weeks of any impact on the execution of the contract. Without any response after this period, no impact will be considered.

Ref	Title	IDM Doc ID	Version
1	General Management Specification for Service and Supply (GM3S)	82MXQK	1.4
2	ITER CODAC Glossary	34QECT	2.0
3	Plant Control Design Handbook (PCDH)	27LH2V	7.1
4	Software Engineering and Quality Assurance for CODAC	2NRS2K	3.2
5	DAN User manual	Q6GULS	3.12
6	Data archiver	Spring 2021 EPICS Collaboration Meeting · Online (6-July 9, 2021): Archiving and accessing PVA data at ITER · CLS Indico	
7	UDA user manual	TPLTKG	2.3
8	UDA code base	ukaeca/UDA: Universal Data Access library to provide data over the network in a unified data object.	

SERVICE

9	MINT, an ITER tool for interactive vitalization of data	MINT, an ITER Tool for Interactive Visualization of Data	
10	MINT user manual	4N9E8W	1.7
11	CODAC DDD	6M58M9	3.1
12	Data Handling SRS	257MW7	1.4
13	Data Handling SADD	2AWJNL	1.5
14	Latest development in IMAS	CDU5JE	1.0

4.2 Applicable Codes and Standards

See Ref[3] and Ref[4].

5 Contract Duration

The framework contract duration is four (4) years firm and optional period of two (2) years.

6 Scope of Work

This section defines the specific scope of work for the service, in addition to the contract execution requirement as defined in Ref [1].

The scope of work covers the technical services to be provided to IO along the complete lifecycle of the data handling ecosystem.

Some ITER systems have started their commissioning for years now and we have archived so far, a certain amount of data (roughly 20 TB of experimental data). This data needs to be kept for the ITER lifetime.

The service scope covers the data archiving system, the data access system, the data processors and the data visualization tools. It is important to note that most of these components are in production now: *it implies an extremely complete regression test suits when adding new features.*

The code source of any of the below components (when not public) can be requested if it helps.

The global I&C architecture is given Figure 1. At ITER, we are using EPICS to control and monitor our plant systems. Ref [11] gives an overview of the CODAC systems

SERVICE

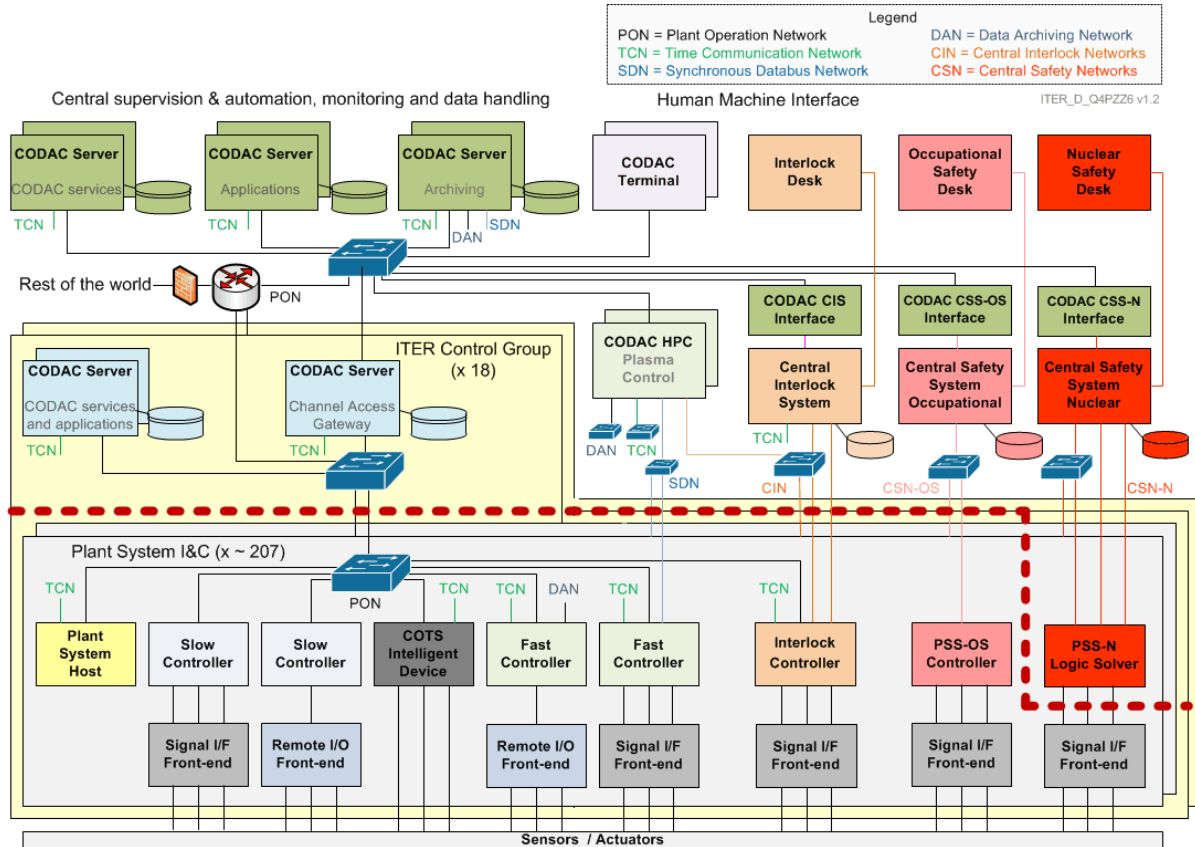


Figure 1. I&C logical architecture

It is important to highlight that there are two types of systems:

- Continuous systems like electrical, cooling water which publish data 24h/7 which need to be archived.
- Pulsed systems like heating systems, diagnostics systems which produce very high throughputs (up to GB/sec/stream) of data to be archived for a period up to 1 hour.

Ref [11] and Ref [12] describes the software requirements and the detailed design of the data handling system.

6.1 Data Archiving system

All data archived at ITER can be considered as time-series data. Some of the data are archived at regular intervals and some are archived based on a value change (typically for slow systems).

6.1.1 Description

This section gives an overview of the ITER data archiving system. Description

The data archiving system includes 3 main modules:

- *DAN* which includes an agent and an archiver to publish and collect high-throughput data. *DAN* is an in-house development product written in C/C++. For more information, refer to Ref[5] and source code can be requested.

SERVICE

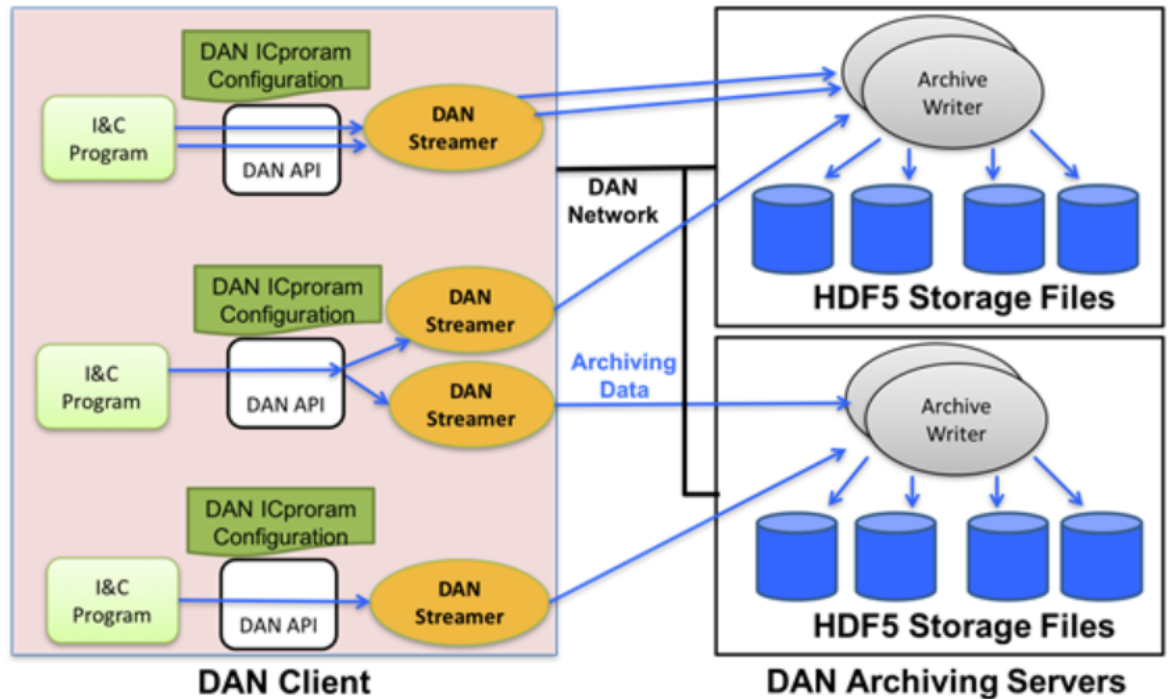


Figure 2. simplify overview of the DAN architecture

- *Data-archivers* for SDN data, PON data (both using Channel Access and PVAAccess protocol). These modules have been written in C++. More information can be found in this presentation Ref [6]

Data is stored in HDF5 format. The data type being archived varies from simple scalar (uint8/16/32/64) to complex structure (mainly the case for SDN and PON/PVAAccess).

6.1.2 Service Duration

See section 5

6.2 Data Access

Although the data archiving system is domain-oriented, the data access is made uniform, meaning that accessing DAN, SDN, PON/CA or PON/PVA is done via the same API. The end-user does not need to know this distinction when accessing the data.

There are two main ways to access the data, either by giving an absolute time range (used for continuous systems) or by pulse ID (used for pulsed systems).

In terms of data access, we need to provide 3 ways to access the data

- Data streaming or publisher-subscriber API (missing) which allows realization of workflow and pipelines: it is also linked with the data processors
- SQL like API which allows to analyse the data (partially implemented)
- Interoperable API (implemented by proposing a REST API) to interact with web application

6.2.1 Description

UDA Ref[7] developed originally by UKAE Ref[8] is used at ITER to serve the data. It is a server-client model, written in C/C++ and it is based on plugins. At ITER, a plugin has been developed to read our data format. Figure 3 represents an overview of the data access.

SERVICE

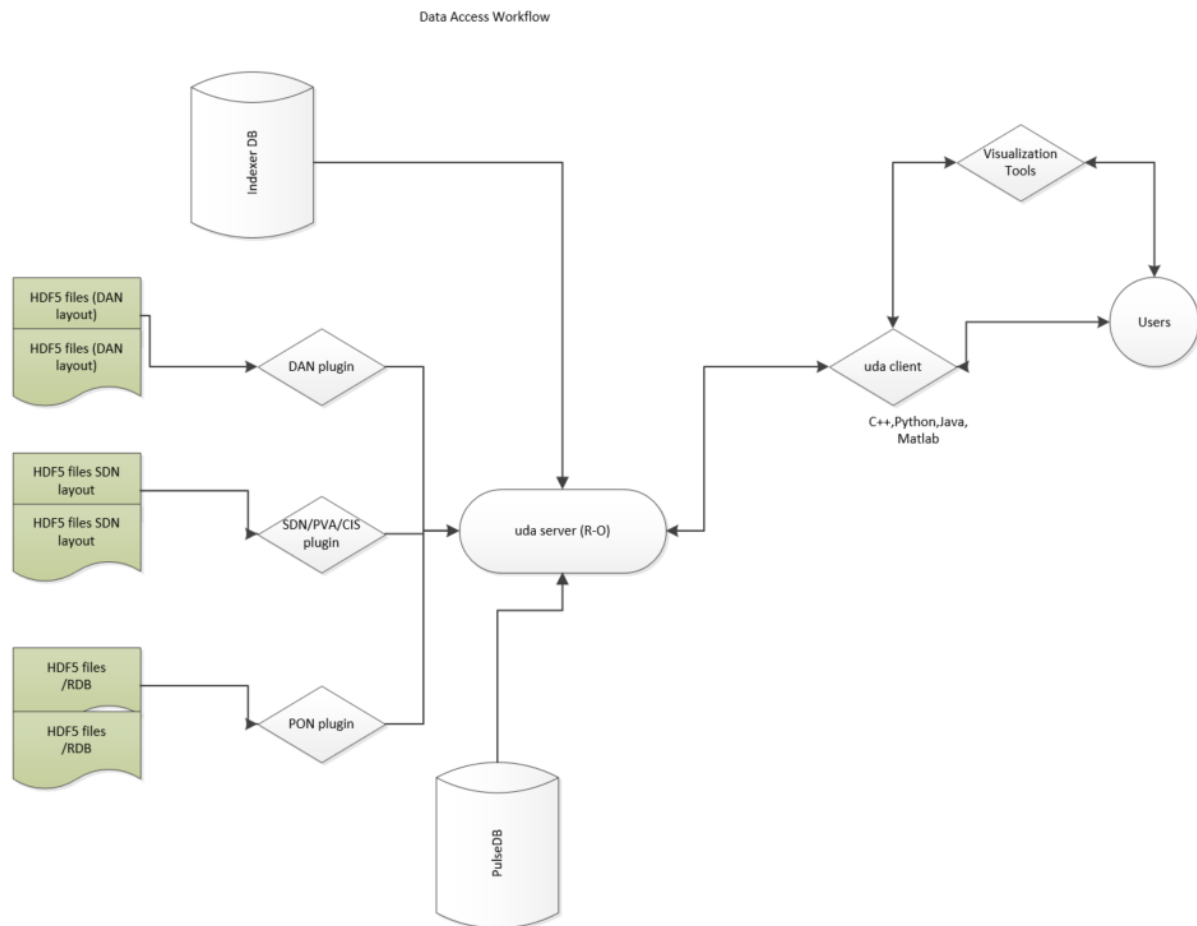


Figure 3. Overview of the Data Access

A REST server (third-party code) has been extended to support communication with a UDA server.

6.2.2 Service duration

See section 5

6.3 Data Processors and Mirroring

Due to the requirement to make the data available to outside POZ, there is a need to have a process which replicates the data from POZ to XPOZ (location will be SDCC).

With the huge amount of data collected at ITER, there is a need to precompute information to visualize the data in a fast way.

6.3.1 Description

Data mirroring has been implemented for DAN as duplicating the flows would have been too expensive. For SDN and PVA the data archivers are duplicated one in POZ and in XPOZ. But there could be a need to also develop data mirroring for other networks too.

Data processors are two-folds

- Components which downsample the data to speed-up data visualization: they can use different decimation algorithms

SERVICE

- Components which transform the data, typical use case is generated calibrated data. Data resampling to put signals on a common time base or to resample PON data to a regular time interval falls into this category.

Data Streamers subscribe to PON or SDN data and stream them as UDP streams. In case of PON data, the streamers are essentially transforming the data, whereas in case of SDN they downsample the data and then transform it. The output messages of the streamers are the same in case of PON or SDN. These streams are ingested by a web server (Node.js) which turns them into server-side events.

These components need to be highly-performant as for instance the first one shall run while data is being acquired.

6.3.2 *Service duration*

[See section 5](#)

6.4 Data visualization

Visualizing data is an important aspect: data retrieval and plotting need to be fast. It also needs quite advanced and domain-specific features. Two sets of tools have been implemented:

- Web applications
- Desktop utilities

6.4.1 *Description*

We have developed a set of Python modules based on PySide6 to allow scientists and diagnostics to build their own tools Ref[9]. It is also used to build the common tools one can find in fusion control rooms. See Ref[10]

A web application has been developed to allow users to make their own dashboards mixing archived data and real-time data. It makes use of the web server mentioned in 6.3.1

6.4.2 *Service duration*

[See section 5](#)

6.5 Data management

With the ITER lifetime, utilities and new or extensions to existing databases could be needed to curate the data. Raw data is written once, read many times and never deleted. Issues could happen such as calibration settings are wrong; units have changed or mistyped. Time information could also be wrong.

6.5.1 *Description*

Utilities to ensure data integrity between POZ and XPOZ are required. In case of mismatch, data need to be carefully analyzed and retrofit (with preserving the original dataset). This part will need to be integrated in an automatic workflow.

Utilities to correct data or metadata will be needed. Data access need to consider the different revisions of the data keeping performance. Utilities whenever possible shall be generic to cope with multiple use cases to avoid profusion of them.

SERVICE

New databases or extension of existing could be needed to annotate data to make the data retrieval more efficient. A typical use case is to overlay the data with component failures which come from another database for instance (logbook or alarm database).

6.5.2 *Service duration*

See section 5.

7 Location for Scope of Work Execution

Contractor can perform the work at their own location.

8 IO Documents

No input is expected from IO.

9 List of deliverables and due dates

The deliverables and due dates are specified in each task order drawn under this framework contract.

10 Quality Assurance requirements

The Quality class under this contract is [\[QC2\]](#), [Ref 1] GM3S section 8 applies in line with the defined Quality Class.

11 Safety requirements

No specific safety requirement related to PIC and/or PIA and/or PE/NPE components apply.

11.1 Nuclear class Safety

N.A

11.2 Seismic class

No specific safety requirement related to PIC and/or PIA and/or PE/NPE components apply

12 Special Management requirements

Requirement from Ref [1] GM3S section 6 applies completed/amended with the below specific requirements.

12.1 Contract Gates

The contract gates are defined in [Ref 1] section 6.1.5, this scope of service call for the following technical gates: [[Contract Gates for Service](#)]

12.2 CAD design requirements

This contract does not imply CAD activities

SERVICE

12.3 Services-specific requirements

12.3.1 *Common to all*

REQ1-1: the contractor shall document any architecture change in the current design or any introduction of a new component

REQ1-2: the contractor shall commit all the codes in ITER standard repositories

REQ1-3: the contractor shall demonstrate capability to understand, debug and extend a third-party code like UDA or HDF5

REQ1-4: the contractor shall develop and deliver improvements to the current ecosystem

REQ1-5 : the contractor shall help identifying weaknesses and propose solutions in the various components

REQ1-6: the contractor shall demonstrate a solid experience in building and maintaining long lasting complex architecture.

REQ1-7: the contractor shall demonstrate a solid experience in developing and investigating in Linux environments

REQ1-8: The contractor shall comply with the ITER software quality standards and requirements

REQ1-9: the contractor shall demonstrate a solid experience in network protocols

REQ1-10: the contractor shall demonstrate a solid experience in tuning parallel filesystems such Spectrum Scale or similar.

REQ1-11: the contractor

12.3.2 *Data Archiving system*

REQ2-1: the contractor shall have the capability to develop robust, reliable and performant C/C++ code

REQ2-2: the contractor shall keep the backward compatibility of the DAN API

REQ2-3: the contractor shall demonstrate a solid experience of HDF5-SWMR

REQ2-4: the contractor shall have the technical capabilities to extend the current code base with necessary features for image acquisition.

REQ2-5: the contractor shall have the technical capabilities to investigate performance bottleneck of the current code base and enhance it.

REQ2-6: the contractor shall have the technical capabilities to investigate and propose suggestion to improve the plant system I&C developer's code which integrate the DAN library or to fix the DAN library if needed.

REQ2-7: the contractor shall have the technical capabilities to extend the current code base with necessary features for data acquisition

REQ2-8: the contractor shall demonstrate the technical capabilities to work and tune very high data rates acquisition systems, that is machine equipped with 100Gb NIC. It also involves Linux kernel settings.

12.3.3 *Data Access*

REQ3-1: the contractor shall demonstrate the capability to develop performant C/C++ code

REQ3-2: the contractor shall keep as much as possible the backward compatibility of the UDA API

REQ3-3: the contractor shall be able to extend the current APIs with new SQL-features

REQ3-4: the contractor shall be able to extend the current REST API if needed

SERVICE

REQ3-5: the contractor shall be able to investigate and fix performance bottlenecks in the UDA current code base (including the plugins)

REQ3-6: the contractor shall be able to extend the current code base (including the plugins) with new needs such as cancelling a request

REQ3-7: the contractor shall be able to propose new solutions and be able to implement them to improve the performance of the data access.

REQ3-7: the contractor shall be able to propose new solutions and be able to implement them to improve the performance of the data access for images if needed.

REQ3-8: the contractor shall be able to propose and implement new signal decimation mechanisms if needed

12.3.4 Data Processors

REQ4-1: the contractor shall demonstrate the capability to develop performant C/C++ code

REQ4-2 the contractor shall demonstrate a solid experience in building publisher-subscribers component

REQ4-3 the contractor shall demonstrate a solid experience in tuning high-throughput systems (up to several GBs/sec)

REQ4-4 the contractor shall demonstrate a solid experience in realizing complex ETL and streaming processing using open-source technologies.

REQ4-5 the contractor shall be able to develop a reliable data mirroring application if needed

12.3.5 Data visualization

REQ5-1: the contractor shall demonstrate a solid experience in building efficient and interactive desktop tools using Pyside, PyQtGraph

REQ5-2: the contractor shall demonstrate a solid experience building web applications using state of the art technologies

12.3.6 Data management

REQ6-1: the contractor shall demonstrate a solid experience in data curation

REQ6-2: the contractor shall develop performant tools which can be either running synchronously or offline tools.

REQ6-3: the contractor shall demonstrate a solid experience in managing large datasets and workflows to indicate missing metadata to make the data access more useful

REQ6-4: the contractor shall be able to develop and enhance the data integrity data checks

REQ6-5: the contractor shall be able to contribute to signal mapping between CODAC and IMAS

12.4 Skills and qualifications

The Contractor shall provide a team with qualified and experienced persons to perform efficient planning, management, supervision and inspection of all services. The contractor shall be organised in terms of systems and processes.

SERVICE

The Contractor shall establish a management and records system to demonstrate staff member level of experience, including educational qualifications, work experience and training. Records shall be kept by the contractor and provided to IO upon request.

The Contractor shall present to the IO the CV of one named individual for each level of expertise requested and identify for each a substitute capable of replacing them in case of absence. If qualified, one person may be identified for up to two roles. These individuals shall not be changed during the period of execution of the services, unless agreed with IO as specified in the Contract.

The strategic functions per level of expertise are:

- Junior Software Engineer (< 5 years)
- Confirmed Software Engineer (5-10 years)
- Senior Software Engineer (> 10 years)

13 Appendices