



Data di pubblicazione: 12/10/2022

Nome allegato: 934239543E_Chiarimenti_3^.pdf

CIG: 934239543E;

Nome procedura: *Procedura di dialogo competitivo ai sensi dell'art. 64 del D.Lgs n. 50/2016 per l'Acquisizione di un sistema di "Anonimizzazione dati"*



Istituto Nazionale Previdenza Sociale

ISTITUTO NAZIONALE PREVIDENZA SOCIALE

DIREZIONE CENTRALE RISORSE STRUMENTALI E CENTRALE UNICA ACQUISTI

CHIARIMENTI

**Procedura di Dialogo competitivo per l'affidamento del
"Servizio di Anonimizzazione Dati"**

CIG: 934239543E

Via Ciro il Grande, n. 21 – Roma

C.F. 80078750587 - P.IVA 02121151001

Quesito 1

Oggetto del presente bando è unicamente la fornitura del software di anonimizzazione dei dati, o sono richiesti anche i servizi per la messa in opera della soluzione tecnologica, quali ad esempio, integrazione, configurazione, project management?

Risposta al quesito 1

Oggetto del bando sono tutte le attività indicate nel quesito

*

Quesito 2

In quale momento o fase del procedimento di gara saranno date le informazioni tecniche di perimetro utili a formulare una offerta economica ed un piano di progetto?

Risposta al quesito 2

Durante la Fase 2 verranno forniti ulteriori elementi utili alla quotazione ed alla stesura del piano.

Nella fase 3 il concorrente inserisce la propria offerta economica entro 30 giorni dall'avvio dell'elaborazione del PoC. L'offerta economica terrà conto delle informazioni di contesto acquisite durante la Fase 2.

La Fase 2 è finalizzata a meglio dettagliare il piano di progetto già presentato in Fase 1, che dovrà, quindi, essere consegnato aggiornato. Il piano di delivery, che declinerà il piano di progetto, verrà poi consegnato tra i primi deliverable delle attività in ambito, come descritto nel capitolato tecnico (Par. "Modalità di realizzazione e consegna").

*

Quesito 3

In relazione al requisito di cui al punto 7.2.1 a) del Disciplinare, e in base a quanto indicato nel primo chiarimento trasmesso, si chiede di confermare che il fatturato specifico richiesto debba far riferimento a "servizi di banche dati" - secondo il CPV 72320000.4 indicato anche nel bando di gara - e non necessariamente solo a servizi di anonimizzazione dati.

Risposta al quesito 3

Si conferma.

Si richiama, in proposito, il pronunciamento del CdS, Sez. V, del 3 novembre 2021, n. 7341: "*Nell'ipotesi in cui il bando preveda come requisito di fatturato specifico lo svolgimento pregresso di servizi analoghi, tale nozione non è assimilata a quella di servizi identici, ma piuttosto, di servizi afferenti al medesimo settore imprenditoriale o professionale*".

*

Quesito 4

Si chiede di chiarire se è possibile considerare come "attività analoghe a quelle oggetto di gara" anche la fornitura di sistemi per la protezione e messa in sicurezza dei dati strutturati e semi strutturati.

Risposta al quesito 4

Si conferma.

*

Quesito 5

Si chiede, se possibile, di avere maggiori informazioni circa la modalità di fatturazione prevista.

Risposta al quesito 5

Tenuto conto che il progetto esecutivo dei servizi verrà definito in fase di dialogo competitivo (fase 2) e successivamente sperimentato (fase 3), le modalità di fatturazione verranno concertate tra amministrazione e operatore economico aggiudicatario.

*

Quesito 6

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0.pdf", paragrafo 4 "Oggetto della fornitura" p. 4, in merito alla capacità di eseguire task di anonimizzazione su dati semi-strutturati (e.g., JSON), si chiede di chiarire se i dati in formato JSON sono disponibili come file sul file system o presenti all'interno di database relazionali.

Risposta al quesito 6

I file JSON saranno per lo più da prelevare su DB NO SQL oppure Oracle SODA NoSQL. Potrebbero essere, però, anche già depositati su un HDFS di tipo Cloudera.

*

Quesito 7

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0.pdf", paragrafo 5 "Modalità di realizzazione e consegna" p. 6, in merito all'affermazione "Il sistema è integrato con i sistemi di autenticazione, autorizzazione e sicurezza di INPS", si chiede di indicare i sistemi di autenticazione e sicurezza di INPS .

Risposta al quesito 7

Si tratta di sistemi di vario genere, sistemi sia custom (autenticazione, autorizzazione e tracciabilità) che commerciali (es. network/security). Ulteriori dettagli saranno forniti successivamente.

*

Quesito 8

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0.pdf", paragrafo 2 "Premessa" p. 2, in merito all'affermazione "Messa a disposizione di Open Data nella qualità e quantità utile alla conduzione di ricerche, analisi e hackathon (e.g., con focus su AI e Data Science) basate su dataset non completamente generalizzati", si chiede di specificare meglio cosa s'intende per "*dataset non completamente generalizzati*".

Risposta al quesito 8

Leggasi "dataset anonimizzati che hanno mantenuto un valore informativo sul dato non aggregato".

*

Quesito 9

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0.pdf", paragrafo 2 "Premessa" p. 2, in merito all'affermazione "Messa a disposizione di Open Data nella qualità e quantità utile alla conduzione di ricerche, analisi e hackathon (e.g., con focus su AI e Data Science) basate su dataset non completamente generalizzati", si chiede di specificare meglio cosa s'intende per "*Messa a disposizione di Open Data nella qualità e quantità utile alla conduzione di ricerche, analisi e Hackathon.*" Si intende ad esempio il formato di file da generare?

Risposta al quesito 9

Si intende che i dati anonimizzati possono essere utilizzati come open data a vario titolo.

*

Quesito 10

Con riferimento al documento "Disciplinare_Anonimizzazione2", paragrafo 15.1 "Fase 1 – criteri di valutazione del progetto tecnico" p. 36, in merito alla lett. A. "Grado di aderenza della proposta alle caratteristiche di base della soluzione ed alle "specifiche di realizzazione"", si chiede di specificare come il punteggio relativo al criterio (20 punti) sia distribuito tra il "grado di aderenza della proposta alle caratteristiche di base della soluzione" (§4 del capitolato tecnico "oggetto della fornitura") e il "grado di aderenza della proposta alle specifiche di realizzazione" (§ 5 del capitolato tecnico "modalità di realizzazione e consegna").

Risposta al quesito 10

Fatta la totalità dei punti richiesti, si valuta il grado di aderenza complessivo.

*

Quesito 11

Con riferimento al documento "Disciplinare_Anonimizzazione2", paragrafo 15.1 "Fase 1 – criteri di valutazione del progetto tecnico" p. 36, in merito alla lett. B. "Quantità di "valore aggiunto" della proposta rispetto alle "caratteristiche di base" ed alle "specifiche di realizzazione" ", si chiede di specificare come il punteggio relativo al criterio (20 punti) sia distribuito tra la "quantità di valore aggiunto della proposta rispetto alle caratteristiche di base" (§4 del capitolato tecnico "oggetto della fornitura") e la "quantità di valore aggiunto della proposta rispetto alle "specifiche di realizzazione" (§ 5 del capitolato tecnico "modalità di realizzazione e consegna").

Risposta al quesito 11

Fatta la totalità dei punti proposti, si valuta la quantità di valore aggiunto complessivo.

*

Quesito 12

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0", paragrafo 5 "Modalità di realizzazione e consegna" p. 5, in merito alle specifiche di realizzazione: analisi e sviluppo del PoC, si richiede di chiarire se la durata del POC, per il quale è richiesta una durata massima di 30 giorni solari, sia inclusa nei 6 mesi richiesti per la realizzazione del progetto.

Risposta al quesito 12

Si conferma, i 30 giorni solari sono da intendersi inclusi nel periodo massimo di 6 mesi richiesti per la realizzazione del progetto.

*

Quesito 13

Con riferimento al documento "Disciplinare_Anonimizzazione2", paragrafo 15.1 "Fase 1 – criteri di valutazione del progetto tecnico" p. 37, in merito alla lett. D. "Tempi di fornitura (i punteggi sono normalizzati nel range dei tempi di fornitura dichiarati nelle proposte)", a prescindere dagli use-case proposti, in assenza di elementi dimensionali per la progettazione della soluzione, si richiede di chiarire che cosa si intende con il Criterio "tempo di fornitura".

Risposta al quesito 13

Elementi dimensionali. Il perimetro potenziale di lavoro è costituito da circa 10.000 istanze DB/schema in produzione con 1.300 TB di storage utilizzato, distribuiti tra piattaforme SQL Server, Oracle, DB2 e Mongo DB. Per tale perimetro ci si aspetta dalla soluzione di anonimizzazione un impatto medio annuo, nei primi 3 anni, stimato tra il 10% ed il 20% dello storage, tra anonimizzazione spot e continuativa (flussi).

*

Quesito 14

Premesso che la concorrente è in possesso di tutti i requisiti richiesti, compreso quello di capacità tecnica e professionale, siamo a chiedere il seguente chiarimento:

1. È possibile Avvalersi di una azienda che pur non possedendo la certificazione di qualità UNI EN ISO 9001/2015 permetta alla concorrente di presentarsi al meglio?

Ci riferiamo in particolare alla possibilità di esprimere capacità tecnica e professionali, specifiche su soluzioni o servizi analoghi a quelli oggetto della presente procedura, che soddisfino al meglio i Requisiti espressi al capitolo 7.3 del DISCIPLINARE DI GARA.

Poniamo il quesito anche a causa di quello che riteniamo essere un refuso che appare nel DISCIPLINARE DI GARA dove a pagina 18, capitolo 7.4, al punto 6, ci si riferisce a "elenco dei servizi analoghi di cui al paragrafo 7.3, lettera a)" che in realtà è relativo alla certificazione di qualità UNI EN ISO 9001/2015. Mentre a pagina 19, capitolo 8. AVVALIMENTO, al punto 4, si fa riferimento a "Il ricorso all'avvalimento per la certificazione di cui al paragrafo 7.3, lettera b)", cui non troviamo riscontro nel documento. Questo ci ha fatto ipotizzare che sia possibile, se non opportuno, esprimere requisiti tecnici e professionali relativi a servizi analoghi a quelli oggetto di gara anche attraverso l'avvalimento di una azienda Ausiliaria non in possesso della certificazione di qualità UNI EN ISO 9001/2015, comunque posseduta dalla Concorrente ausiliata.

Risposta al quesito 14

Come già risposto in altri quesiti già pubblicati si precisa che:

Relativamente ai requisiti di capacità tecnico-professionale si conferma che:

1. l'unico requisito tecnico-professionale richiesto è il possesso della certificazione di qualità UNI EN ISO 9001/2015, (art. 7.3 lett. a) del disciplinare di gara) che in caso di RTI deve essere posseduto da tutti gli operatori del raggruppamento;
2. il requisito citato come 7.3 lett. b) (elenco dei servizi analoghi) è un refuso;
3. al capitolo 8 – Avvalimento, al punto 4 deve intendersi "7.3 lett. a)";
4. Tutti i riferimenti alla lettera b) del paragrafo 7.3 devono intendersi come refuso.

Pertanto, non si conferma che possa esserci l'istituto dell'Avvalimento per la certificazione di qualità UNI EN ISO 9001/2015 da una società ausiliaria che non ne sia in possesso.

*

Quesito 15

Nell'allegato "Disciplinare_Anonimizzazione2" al par 18 "Stipula del contratto" ai pt. 11 e 12 sono previste le spese di pubblicazione a carico dell'aggiudicatario, ma non viene specificato l'importo dovuto; si richiede pertanto una quantificazione di queste.

Risposta al quesito 15

Le spese di pubblicazione (indizione ed esito della procedura su Gazzetta Ufficiale Italiana e quotidiani) ammontano a circa € 8.000,00, IVA inclusa.

*

Quesito 16

Qual è lo spazio stimato in Gigabyte o Terabyte o righe di database associato ai dati da mascherare rispettivamente per DB2 su z/OS, per Oracle su piattaforma Unix/Windows e per MS SQL Server su piattaforma Windows?

Risposta al quesito 16

Il data masking è solo una delle funzionalità. Di seguito, si riporta il perimetro dettagliato di anonimizzazione.

<i>Piattaforma</i>	SQL Server	DB 2	Oracle
<i>Num. di DB/schema</i>	5500	862	5000 (DB schema)
<i>Storage utilizzato</i>	1150 TB	50 TB	100 TB
<i>Sizing medio per DB</i>	148 GB	59 GB	2 TB
<i>% Anonimizzaz. (stima)</i>	10-20%	5-10%	10-20%

*

Quesito 17

Con che frequenza temporale devono essere prodotti i dati anonimizzati, rispettivamente per ambienti DB2 su z/OS, Oracle su piattaforma Unix/Windows e SQL Server su piattaforma Windows?

Risposta al quesito 17

Ci saranno scenari di anonimizzazione di tipo spot ed altri di tipo flusso con frequenza variabile.

*

Quesito 18

Qual è la dimensione totale in MIPS di tutte le partizioni (LPAR) dove sono presenti dati DB2 di input del processo di anonimizzazione? Qual è invece il totale dei MIPS delle CPU fisiche con sistema operativo z/OS su cui è presente il DB2 oppure i loro modelli?

Risposta al quesito 18

Sono in uso due elaboratori Z14 con un totale, per entrambi gli elaboratori, di 14.000 MIPS.

*

Quesito 19

Con riferimento al documento "All_4_Capitolato_tecnico_anonimizzazione_draft_final_v4.0", capitolo 4 "Oggetto della fornitura", in relazione alla caratteristica di base:

"Disponibilità di connettori per l'acquisizione semplificata e continuativa di dati da molteplici tipi di database, in particolare: DB2, Oracle e Microsoft SQL Server".

Si chiede di conoscere le versioni in uso dei database sopra citati (DB2, Oracle e Microsoft SQL Server).

Risposta al quesito 19

- Sql Server 2016 e Sql Server 2019
- Piattaforma Db2 su z, versione 12M500
- Oracle 18.14.0.0.0 su "Oracle Linux Server 6.10"
- Oracle 19.15.0.0.0 su "Red Hat Enterprise Linux Server 7.9"

*

Quesito 20

Al fine di comprendere al meglio il contesto INPS, per poter valutare adeguatamente il sizing della soluzione di anonimizzazione che si intende proporre, si chiede quale è il numero di basi di dati in ambiente di produzione INPS?

Risposta al quesito 20

Di seguito, si riporta il perimetro dettagliato di anonimizzazione.

Piattaforma	SQL Server	DB 2	Oracle
Num. di DB/schema	5500	862	5000 (DB schema)
Storage utilizzato	1150 TB	50 TB	100 TB
Sizing medio per DB	148 GB	59 GB	2 TB
% Anonimizzaz. (stima)	10-20%	5-10%	10-20%

*

Quesito 21

Quale è la distribuzione delle differenti tecnologie DB di produzione tra SQL, noSQL, altro?

Risposta al quesito 21

Vedi risposta 20.

*

Quesito 22

In linea di massima quante e quali sono le tipologie di fonti esterne di dati utilizzate in produzione?

Risposta al quesito 22

Vedi risposta 20.

*

Quesito 23

Quale è la dimensione in termini di spazio disco utilizzata in INPS del DB più grande presente in produzione?

Risposta al quesito 23

Max size è quella di un DB Oracle, di 22 Tb. La max size di un DB SQL Server relazionale (no blob) è di 8 Tb.

*

Quesito 24

Quante potrebbero essere le utenze amministrative che avranno possibilità di gestire/customizzare la soluzione di masking?

Risposta al quesito 24

Massimo 10 utente amministrativo/operative. Sarà un servizio centralizzato con team dedicato.

*

Quesito 25

Quante sono le utenze che dovranno fruire la componente di monitoraggio del masking?

Risposta al quesito 25

Massimo 10 utente amministrativo/operative. Sarà un servizio centralizzato con team dedicato.

*

Quesito 26

Quali sono le finestre temporali nelle quali poter effettuare il masking dei DB di produzione?

Risposta al quesito 26

Si ritiene non corretto applicare tale vincolo a prescindere. Il data masking sarà effettuato su flussi pipeline, batch ed export in base allo scenario. Le finestre temporali potrebbero, quindi, coincidere con quelle del flusso.

*

Quesito 27

Per la PoC richiesta in FASE 2 dal capitolato tecnico (Allegato 4, capitolo 9):

Il Dataset fornito sarà distribuito tra differenti DB oppure appartiene ad un unico DB?

Risposta al quesito 27

Sarà chiarito in FASE 2 nella documentazione che dettaglierà il caso d'uso per la PoC.

*

Quesito 28

Quante tabelle del dataset saranno oggetto di mascheramento?

Risposta al quesito 28

Sarà chiarito in FASE 2 nella documentazione che dettaglierà il caso d'uso per la PoC.

*

Quesito 29

Quanti record fanno parte di questo dataset?

Risposta al quesito 29

Sarà chiarito in FASE 2 nella documentazione che dettaglierà il caso d'uso per la PoC.

*

Quesito 30

Si chiede pertanto di chiarire quale sia il criterio della valutazione finale/criterio di aggiudicazione della FASE 3 (competizione) che nell'allegato 4 capitolo 3 sembra essere 80 tecnico 20 economico e nel disciplinare di gara capitolo 15.2 sembrerebbe 70 tecnico e 30 economico.

Risposta al quesito 30

Vedasi il primo chiarimento pubblicato.

*

Quesito 31

In riferimento al Capitolato tecnico, Allegato 4, Capitolo 4, Oggetto della fornitura, in relazione alle specifiche del pannello di amministrazione richiesto, "Analisi dell'accuratezza (output)" si chiede se sarà fornito un algoritmo da INPS per:

- a. misurazione del rischio dell'output
- b. misurazione del valore dell'output

Risposta al quesito 31

Si conferma. Le relative specifiche saranno chiarite in FASE 2.

*

Quesito 32

In riferimento al Capitolato tecnico, Allegato 4, Capitolo 10, Criteri di valutazione – FASE 3 (Sviluppo del PoC), in merito al "gioco di ruolo" si chiede quanto si prevede durerà la finestra a disposizione di ciascun fornitore per la de-anonimizzazione di quanto fatto dagli altri?

Risposta al quesito 32

Sarà chiarito in FASE 2.

*

Quesito 33

In riferimento al Capitolato tecnico, Allegato 4, Capitolo 4, Oggetto della fornitura, in relazione alle specifiche del pannello di amministrazione richiesto, è possibile avere una breve descrizione di un caso d'uso specifico INPS in merito alla "Possibilità di applicare dinamicamente più algoritmi di anonimizzazione"?

Risposta al quesito 33

Viene scelto dinamicamente un algoritmo sullo specifico dataset affinché sia massimizzata l'accuratezza.

*

Quesito 34

In riferimento al Capitolato tecnico, Allegato 4, Capitolo 5, in relazione all'elenco fornito di caratteristiche di valore aggiunto è riportato "Sviluppo di prototipi innovativi a supporto delle attività di analisi (e.g., ML per mimetizzare i dati)", è possibile avere una breve descrizione di un caso d'uso?

Risposta al quesito 34

Si fornisce il seguente esempio: invece di condividere dati aggregati sull'andamento dell'occupazione per dimensioni quali le categorie professionali (codici ateco) e le aree territoriali (regioni/provincie/città), sarà possibile fornire dati sulle singole posizioni in parte simulando il valore degli attributi, mantenendo la valenza statistica del campione, in parte fornendo attributi puntuali in chiaro, pur restando non re-identificabile i soggetti.

*

Quesito 35

In riferimento al Capitolato tecnico, Allegato 4, Capitolo 6, Strategia di anonimizzazione, in relazione al dettaglio fornito come esempi di maggior sofisticazione della strategia di anonimizzazione viene citata la possibilità di una scelta dinamica (condition / content-based) della migliore tecnica da utilizzare sul data set, è possibile avere una breve descrizione di un caso d'uso?

Risposta al quesito 35

Viene scelto dinamicamente un algoritmo sullo specifico dataset affinché sia massimizzata l'accuratezza.

*

Quesito 36

In riferimento al Disciplinare di gara al paragrafo 7.2 punto 1.a) si chiede di specificare se il fatturato minimo annuo richiesto è da intendersi quale fatturato specifico non essendo specificato.

Risposta al quesito 36

Si conferma

f.to Gianfranco Ruberto