HORIZON 2020
The EU Framework Programme for Research and Innovation

CAPABLE

Cancer Patients Better Life Experience

Grant Agreement No. 875052
Start Date: 01/01/2020 (48 Months)

Deliverable No. 1.2
DMP - Data Management Plan

Due Date: [30/06/2020]
Submitted On: [30/06/2020]

Coordinator	University of Pavia (UNIPV)
Deliverable Lead Partner	UNIPV
Contributing Partners	AMC, BIOM, AIMAC, ICSM, NKI
Contact	Prof. Silvana Quaglini
Email	silvana.quaglini@unipv.it
Website	www.capable-project.eu

Deliverable Type		
R	Document, report	
DEM	Demonstrator, pilot, prototype	
DEC	Websites, patent fillings, videos etc.	
OTHER	ORDP	**[X]**
Dissemination Level		
PU	Public	**[X]**
CO	Confidential (Consortium members including the Commission Services)	
CI	Classified Information (Commission Decision 2015/444/EC)	

This deliverable is a part of a project receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875052.

Table of Contents

List of Tables

Versions History

Version	Date	Author	Comments
0.5	18/05/2020	Enea Parimbelli	Defined structure and filled sec 1.Data summary
0.6	27/05/2020	Enea Parimbelli	Added exec summary and quantitative measurements for review
0.7	08/06/2020	Enea Parimbelli, Ronald Cornet	Added to section 2. FAIR
0.9	10/06/2020	Enea Parimbelli, Flora Gilboa, Matteo Gabetta	After inputs from Flora and Matteo. Ready for discussion at CM Pavia and internal review
1.0	11/06/2020	Enea Parimbelli	Submission candidate, ready for internal review
1.1	19/06/2020	Enea Parimbelli	After further feedback from NKI
1.2	24/06/2020	Enea Parimbelli, Flora Gilboa	Version 1.2 submission candidate
2.0	20/02/2023	Enea Parimbelli, Silvia Panzarasa	Version 2.0 updated for reporting period 2

Executive Summary

The CAPABLE project adheres to the Pilot on Open Research Data (ORDP) in Horizon 2020, which aims to improve and maximize access to and re-use of research data generated by actions.

In this context, the present document - Data Management Plan (DMP) - addresses the relevant aspects of making data FAIR – findable, accessible, interoperable and re-usable. The present DMP has been updated at the end of project reporting period 2 (end of 2022), after a first version of the document has been delivered at project month 6, as D1.2. The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

The DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data FAIR, the DMP includes information on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

In order to comply with the expectations for the content of the DMP, the present document adopts the template the European Commission provides with its Horizon 2020 funding guide: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template

1. Data Summary

1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

The overall objective of CAPABLE is to combine the most advanced technologies for data and knowledge management with a sound socio-psychological approach in order to develop a coaching system for improving the quality of life of cancer patients. The system aims at early detecting and managing cancer-related issues and at satisfying the needs of patients and their home caregivers. Ultimately, CAPABLE will exploit several different datasets using AI techniques to effectively monitor individual patients, with the final goal to improve quality of life after cancer treatment.

More specifically, the data collection and analysis activities in the CAPABLE project will help achieve the following objectives:

- Identifying, classifying and ranking new cancer patients' and their home caregivers' needs, mostly leveraging on data provided by the AIMAC patients' association, interviews and questionnaires to be administered in the requirements elicitation work package (WP2).
- Improving patients' compliance to treatment by acquiring Patient Reported Outcomes (PROs) and Patient Reported Experiences (PREs).
- Collecting data for early identification of deterioration in quality of life or emotional issues.
- Improving healthcare professional workflows by promptly identifying priority patients and shortening the duration of control visits due to a better understanding of the patient conditions, thanks to data collected in-between visits at the patient home.
- Identifying adverse events of (relatively) new therapies or unknown long-term effects of cancer treatment.
- Developing new, data-driven AI models for the course of cancer, which could drive more personalized interventions.

1.2 What types and formats of data will the project generate/collect?

The system will rely on both data already available to partners at the beginning of the project and on data that will be collected during the clinical study, which will last the entire fourth year of the project. The clinical study, that will take place at three clinical partner organizations NKI, ICSM and Policlinic of Bari (BARI, in the following) will enroll a set of different cancer patients including: renal, lung, prostate, H&N, breast, thyroid gland (ICSM and BARI);and melanoma patients (NKI). Thus, the data collected by the project pilot will be focused on these cancer patient populations, but many of the findings intend to be generalizable to other cancer domains. Table 1 provides a summary of the data that CAPABLE will collect.

All the data collected by the project during the pilot studies will be stored in a centralized data repository based on the OMOP CDM [1], in order to improve standardization and promote reusability. More detailed information regarding the format of the data collected by CAPABLE will be provided in section 2 - FAIR data.

Table 1 - Data generated and collected by CAPABLE

Data collected during the pilot study				
DATA COLLECTED AT THE HOSPITAL				
Clinical history (comorbidities, allergies, hospitalizations, procedures, etc)	Context (familiar, economic and social)	Follow-up visits (adverse effects*, other signs and symptoms, psychological issues, treatment updates, supportive care, etc)	Laboratory data (routine tests, biomarkers for specific cancers)	
DATA COLLECTED AT HOME				
Self-reported through the CAPABLE patient's app				
Adverse effects of treatment	Diet and Self-prescriptions (herbs, supplements, etc)	Instrumental measures (weight, blood pressure, etc)	Generic Quality of life questionnaires and behaviour-detection questionnaires*	Cancer-specific questionnaires* (EORTC questionnaires**) Caregivers' burden questionnaires*
Out-of-pocket costs	Sick leave days -patient	Sick leave days - caregivers	Behaviour/psychological/social information (e.g. changes in family context)	Other PROs or PREs Additional patient-reported outcomes and experience not foreseen a-priori
Automatically collected through mobile phones/sensors/wireless devices				
Patient's use of mobile phones (indicating that patient is active)	Body Sensors - wearables Physical activity, Temperature, Heart Rate, Sleep duration and quality	Wireless devices Scale, Glucometer, Sphygmomanometer, etc	Environmental sensors Temperature, Humidity, Air quality, etc	
Recommendations delivered by the guideline-based DSS and virtual coach directed at both patients and physicians				

**A comprehensive list of the questionnaire instruments that will be administered during the pilot study at the three clinical institutions is provided in the clinical study protocols of each of the institutions.*

*** The European Organisation for Research and Treatment of Cancer quality of life questionnaire-C30 (EORTC QLQ-C30) is a widely used, cancer-specific, self-administered questionnaire with strong validity* [2,3].

1.3 Will you re-use any existing data and how?

One of the main assets of the CAPABLE consortium at the beginning of the project is original and largely undisclosed retrospective data that, together with literature results, will be exploited to build AI-based models for predicting the disease course, the onset of adverse effects and patient behaviour and engagement. Retrospective data will also aid the requirement elicitation phase in the context of WP2. Table 2 summarizes the existing data that CAPABLE will be able to exploit. Deliverable D5.1 (also due M6) provides additional details regarding the retrospective data sets belonging to the project clinical partners as well as their anonymization procedures.

Table 2 - Retrospective data available to CAPABLE at M0

Data already existing at the beginning of the project	
DATA SETS BELONGING TO AIMAC, THE PATIENTS' ASSOCIATION PARTNER (numbers refer to M0, and are increasing daily)	
Structured data about patients' needs and issues reported (63.437 records available)	Unstructured (textual) data available in the patients' discussion forum (6,040,641 page views; 905,476 visitors; 7863 subscribers; 70395 messages)
DATA SETS BELONGING TO THE PROJECT'S MEDICAL PARTNERS	
ICSM hospital	NKI hospital
Pseudanonymized dataset with: Initial treatments, adverse effects, treatment changes, complete disease course and outcomes for 343 patients (917 treatment lines), who have been followed-up for years.	Anonymized dataset with: Clinical data, adverse effects, treatment and (long-term) follow-up/survival data for 500 patients.
OPEN DATA	
Pollution and other air quality data from open data repository like the ARPA agency of the Lombardia Region	

1.3.1 Available data about kidney cancer patients at the ICSM Hospital in Pavia

Retrospective data on 343 kidney cancer patients has been made available to the project. These data are collected within the international database IMDC (International Metastatic Database Consortium) for Renal Cell Carcinoma (RCC) in which ICSM participates. These data have the specific feature of including a long follow-up (up to a few years). Data include demographic information, the date of surgery and tumour characteristics (TNM classification, size, necrosis, Fuhrman Grade), and the basal renal function (Serum Creatinine). Moreover, data from the first-line therapy up to the fourth-line therapy are reported, namely the drugs used, the response type, the drug dosage modification if any. Weight and all the haematological parameters useful to establish risk (Haemoglobin, Lactate dehydrogenase, Neutrophils, Platelets, Lymphocytes, Serum Creatinine, Sodium and Calcium) are available at each therapy line. Metastases and their characteristics are reported (timing, sites, lymphnodes status), with particular details for brain metastasis, such as date of diagnosis, number of metastases, the size of the largest one, cerebral/cerebellar, symptoms at presentation, stereotactic radiosurgery, neurosurgery, and response data. The follow-up information includes date of the last visit or follow-up termination for other reasons, date of death if it occurred, overall and progression-free survival.

1.3.2 Available data about melanoma cancer patients at the Netherlands Cancer Institute

Retrospective anonymized data on 500 melanoma patients followed up in the NKI-AVL hospital in The Netherlands has been made available to this project. The NKI does not have: any key linking identified subject ID; the NKI is not able to identify individuals from the anonymized dataset and the NKI and no other third parties are able to map or correlate the de-identified list of data that will be available in the consortium. Data is available from 2015 and include patients with a long follow up. Included patients in the dataset have been treated with immunotherapy (anti-PD1 or anti-PD1/anti-CTLA4 combination therapy). The dataset contains over 450 collected variables per patient and is therefore rich in data. Data include patient characteristics (age, sex, comorbidities), tumor characteristics (primary diagnosis date, type, location, size, Breslow, necrosis, staging), lab data specific for melanoma (S100b, LDH), information about metastasis (number, location), primary treatment details (treatment line, start/stop dates, drug type, dose, total administrations, changes in treatment and reasons why), details about surgeries or additional treatments (for example surgical lymph node removal or radiotherapy) and toxicities (grade 3/4) as a result of the therapy and the effects of these (hospitalization/surgery details). Lastly, the dataset has detailed data about the follow-up of the patients with follow up dates and the status of their disease at that moment (for example progressive disease /stable disease/death/complete response) and treatment episode.

1.3.3 Available data from the AIMAC patients' association

AIMAC will provide the conversation texts from its patients Forum (http://forumtumore.aimac.it), a virtual place where those who faced or are facing cancer can meet, share their experiences and discuss. Moreover AIMAC will provide data about needs assessment. As a matter of fact, AIMAC's operators are specially trained for accurate and timely registration of information requests received. All info requests received are detected through a specific survey form, supplied at the helpline and the 45 cancer information desks. In particular, they collect data on type of user, information needs expressed and response methods provided by the operator. On the basis of this form, for an optimal and unified management of the requests collected, an online database has been developed to which each operator has its own credentials for filling-in the online form. The systematic analysis of the data and requests received allows to have indicators and objective data in order to identify the profile of those who contact the information service and measure above all new information needs in order to provide increasingly targeted and personalized answers. For this reason, starting from the systematic analysis of the data, observational surveys have already been carried out. Data have been collected since September 2012 and obviously they increase daily. In the database, as of today, there are ~75000 records (each record corresponds to an inserted questionnaire). All data collected through the forms are stored in a database that can be exported in the most common formats (sql, csv, xml, excel, etc.). Raw data are available year by year, enabling to assess the evolution of needs over time.

NOTE: for a detailed description, including data dictionaries, of the retrospective data already available to CAPABLE (M0-M6) please refer to deliverable D5.1 - Data ready for modelling and reasoning development, including procedures for anonymization /pseudo-anonymisation - and its appendices.

1.4 What is the origin of the data?

Retrospective data are made available by three clinical centers participating in the consortium (ICSM, BARI and NKI), as well as by the AIMAC patients' organization.

Data collected during the pilot study will also be generated in the context of ICSM, BARI and NKI, with their participating clinical staff and enrolled patient cohorts.

1.5 What is the expected size of the data?

1.5.1 Data already available

The data available from ICSM at M0 is currently stored in a REDCap database, which only consists of textual data. The 343 patients available account for < 1MB of storage space (281kb) when exported in csv format.

The retrospective data provided by NKI has a size of around 10MB and will be exported in a Stata file.

About the data provided by AIMAC, the survey results (provided in csv format) have a size < 1MB. The snapshot of the discussion forum at M0 consists of 18.2MB of textual data.

1.5.2 Data generated during the pilot studies in Y4

At the time of writing of the current updated version of the DMP (v2.0) the pilot studies have not started yet, and thus actual data regarding size of information collected by CAPABLE is not available. The estimates performed at the beginning of the project (i.e. DMP v1.2, corresponding to D1.2) are still valid.

Since the number of patients we envision to enroll in the two pilot studies is rather limited (i.e. approx. 35 patients per site in the CAPABLE arm) also the size of their collected data should be comparable to what we declared in the previous section for the retrospective data currently available. Size-critical data types like genomics or raw diagnostic imaging will not be collected in the pilot studies. Thus the nature of the clinical data generated by the project would be mostly textual and stored in an OMOP CMD compliant format, in the CAPABLE Data Platform component.

1.6 To whom might it be useful ('data utility')?

The data generated and collected by CAPABLE will constitute a valuable asset for researchers working in the cancer domain. This is true especially for clinical researchers that would have the opportunity, in addition to the data that is usually collected in clinical practice, to access semantically-integrated, highly detailed data coming directly from patients and their caregivers living in their own home environment. Collecting and seamlessly making available to clinicians patient-generated, patient-reported data items immediately allows shortening the feedback cycle between a patient and his care team. Accumulating well-integrated data in a centralized OMOP-based repository contributes to the creation of a novel, highly reusable, data asset to be exploited for knowledge discovery and new evidence generation as well. Patients and their home caregivers will also have the opportunity to leverage the experience of patients in a similar condition, or who have just gone through a similar journey of cancer care, adverse events, and affected quality of life and wellbeing. Benefits from an educational as well as psychological standpoint of such a possibility should not be overlooked. Finally, clinical institutions and patient associations would be able to access a more detailed analysis of the actual needs and desires of cancer patients, and be able to design more effective supporting services (including e-health and tele-health initiatives) for their patient populations.

2. FAIR data

2.1. Making data findable, accessible, interoperable, and Reusable

CAPABLE WP3 is entirely dedicated to data interoperability through semantic technologies and FAIRification of CAPABLE data. Detailed information on the project-wide standards, methodological and technical solutions are provided in publicly accessible deliverables D3.1[4], D3.2[5] and D3.3[6]. In the present document we provide a compact summary of the same content, highlighting what's more important for CAPABLE Data Management Plan.

Data produced by the project will be made available through the Zenodo community that we set up for the CAPABLE project: https://zenodo.org/communities/capable. Zenodo provides DOIs for all available resources, along with metadata for each resource that is uploaded to the community. At the time of writing, all public project deliverables have been uploaded to the community (including the DMP, with tracking of its versions[7]), and at time of completion of the studies and publication of results, the same Zenodo community has been set up to host data supporting publications and specific analyses (e.g. the ones featured in upcoming deliverables of WP7 such as D7.8, D7.9 and D7.10).

To facilitate data integration and exchange, the CAPABLE Platform needs to adhere to representation, modelling and terminology standards for data as well as metadata, which are accepted by the community. In particular we adopted:

- HL7 FHIR Resources (using JSON format) for representation of patient-level data, as the HL7 FHIR approach is broadly applicable in the healthcare domain and is based on the REST paradigm that is a de facto standard in web-based services;
- Vocabularies to reflect values of patients' attributes, i.e., those from the Observational Health Data Sciences and Informatics (OHDSI) Vocabularies Repository, ATHENA, which consists of about 100 biomedical vocabularies;
- HL7 FHIR Capability Statement as a specific resource to provide metadata, i.e., specify which resources are made available in the CAPABLE Platform, which interactions are allowed (create, read, update, delete), and which search parameters are supported.
- DCAT (Data Catalog) standard to provide generic information about the CAPABLE Platform. With DCAT we have represented relevant metadata using commonly used semantic web standards, represented using the Resource Description Framework (RDF) approach, in Turtle syntax. This representation allows for providing qualified references to other data and metadata;
- persistent identifiers. Using the w3id.org service, any resource can be provided with a globally unique, persistent, and resolvable identifier.

The CAPABLE system encompasses patient data and decision support guidelines, which the implemented information architecture intends to make maximally FAIR , in order to make the data Findable, Accessible, Interoperable and Reusable, while taking into account openness as well as privacy and security concerns.
This was realised by means of the following components:

- A register of globally unique, persistent, resolvable Unique Resource Identifiers (this registry provides identifiers starting with https://w3id.org/CAPABLE/ which redirect to the actual location of the identified resource);
- A FAIR Data Point that complies with the FAIR Data Point Specification (the resources are provided via https://w3id.org/CAPABLE/fdp-catalog both for human processing in a web

browser and for machine-processing via a REST API and SPARQL endpoint. The metadata in this FDP is yet limited, due to the lack of agreement in the community on schemas and ontologies;

- Data Catalog Vocabulary (DCAT) to represent metadata;
- HL7 FHIR server to access data. Access depends on authentication and authorization;
- A licensing policy. This is partially implemented, as desired licensing needs to be determined, agreed on, and aligned with any licences of (meta)data being reused.

The implemented architecture provides a solid and standardized foundation for FAIR data, in which the metadata is separated from the data and will be available independent of the availability of the data, and can be expanded over time to adhere to current and future practices, and in which other data can be included if relevant.

3. Allocation of resources

3.1 What are the costs for making data FAIR in your project?

The costs for making data FAIR will be limited, and are part of the project budget, predominantly that of WP3. It encompasses: representation of data using appropriate standards, specification of relevant metadata, provision of data through a third-party repository.

3.2 How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Costs will be covered by the project budget.

3.3 Who will be responsible for data management in your project?

The clinical partners of the consortium (ICSM, BARI and NKI) will be responsible for data management in the project. Their role of Data Controllers extends to both retrospective, pre-existing data provided to CAPABLE and to the data generated during the pilot studies in Y4. Other partners may have the role of Data Processors, and their obligations regarding data management are regulated by the consortium agreement and supplemented by Data Processing Addenda (DPAs) between the parties involved. An example of DPA used in CAPABLE at the time of consortium agreement signature is provided in Annex I.

AIMAC will retain ownership and responsibility of their own data throughout the project.

Finally, all partners potentially able to access data of patients enrolled in the clinical studies at one of the three institutions (ICSA, BARI, NKI) have signed an appropriate DPIA, as requested by the EU GDPR with the specific institution. Links to the signed versions of the DPIA are provided in Annex II.

3.4 Are the resources for long term preservation discussed (costs and potential value, who decides how and what data will be kept and for how long)?

Given the nature, expected size, and characteristics of data that will be shared we envision no added costs will need to be sustained for long term preservation of the data. A copy of the data will also be retained by the appropriate consortium members (ICSM, BARI, NKI, BIOM) for auditing (2 years) and obligation to keep records (5 years) purposes, as defined in the grant agreement.

4. Data security

4.1 What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Security in CAPABLE is achieved through protocols, policies and standards and all those aspects have been considered from the ground up in developing CAPABLE. The following figure illustrates how we are deploying the different CAPABLE Components. As it transpires, the diagram complies with the classical notion of a three-tier system encompassing zones such as Intranet, DMZ and Internet. More specifically, only adjoining zones may exchange some information while any communication is forbidden between Intranet and Internet zones.

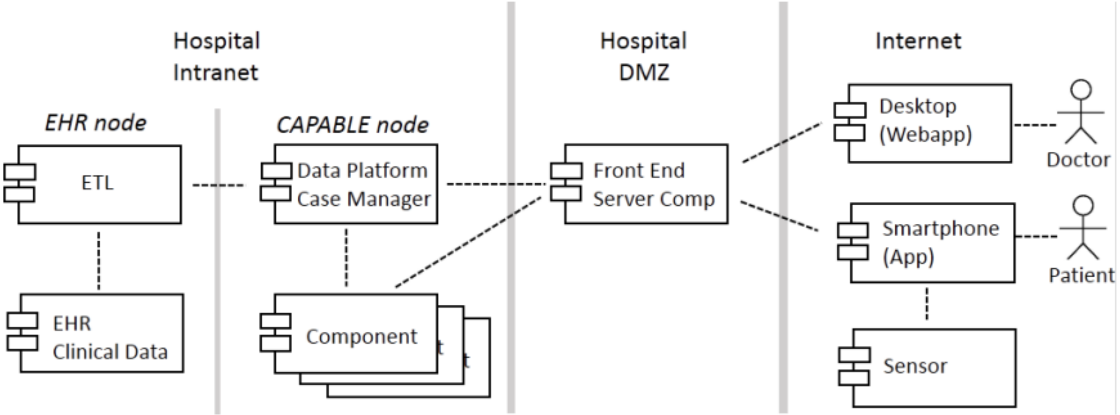


Figure 1: Diagram of CAPABLE components deployment in the hospitals

The Intranet is the innermost security circle provided by institutions where the highest restrictions are enforced preventing any access from the Internet. Sensible data, such as those concerning patient names, social security numbers and the like may only be stored within this zone. For privacy purposes, at hospitals, patient data are often spread on at least two separate repositories located within this zone. One mostly stores demographic data useful for administrative purposes (e.g. patient name and address, social security, caregiver names and addresses, etc..) while the other one is devoted to storing clinical data. This arrangement performs some kind of de-identification and prevents the reconstruction of patient data issuing very simple queries on a single database granting different privileges to applications and system administrators. The various sources of information concerning the same patient may be linked through a system generated code. This procedure is only accomplished dynamically by applications granted the privilege to access the different sources, such as those for the medical staff.

In CAPABLE the only Component devoted to persistent data storage is the Data Platform, which is therefore located in the Hospital Intranet. The system will only be accessed by the medical staff participating in the study and the patients will be identified by their enrollment number. Thus, no direct link is expected to be set up with demographic information stored in the EHR. On this basis, according to each deployment setting, the CAPABLE node may be stored on a separate zone of the Hospital Intranet.

CAPABLE expects indeed to exploit clinical data available in the hospital EHR concerning patients enrolled in the study that have been acquired by means other than the CAPABLE system (e.g., scheduled visits). For that purpose, an additional Component named ETL will take care of accessing the EHR, translating the relevant data concerning patients enrolled in the study and importing them into CAPABLE. During the pilot study, this component will operate on a regular basis in an offline fashion possibly under the supervision of a staff member. In addition, the Data Platform database will be regularly backed up on a secondary database within the hospital premise.

For what regards data recovery, in case the data at the processors will be deleted or corrupted, the data providers can share again the data they provided as the original copy is also retained on their side. The actual data security management of the CAPABLE project will be active during the fourth project year for the studies data; all the patient data will be stored on virtual hardware within the premises of the respective hospital.

The rest of the CAPABLE Components will also be located on the CAPABLE node accessing the Data Platform and contributing their partial interpretation for the case at hand. The only remarkable exception is the Front-End Server Component.

The DMZ acts as a buffer between Intranet and Internet. It receives connections from the Internet and has access to the Intranet for fetching data or supplying new incoming ones. Thus, suitable policies are set up by the Information System department to enable the exchange of information on either side. Usually in this zone no sensible data must be stored nor accessed from the Intranet in order to prevent any leakage in case of any misconfiguration or exploiting attack.

According to the CAPABLE deployment architecture the only Component located in this area is the Front End Server Component. This is expected to interface with the patient Smartphones and the doctor Desktop where the web application displaying patient information is accessed. No information is planned to be stored on the Front End Server Component except for technical cookies enabling sessions. The connections of the Front End Server Component towards the Intranet are expected to target only the Data Platform and the Case Manager. In the former case the purpose is reading patient data and entering new ones, while in the latter is for receiving notifications about events concerning patients that need to be forwarded to their Smartphones on the internet for triggering suitable actions.

CAPABLE components, whether deployed on Intranet or DMZ, will access the Data Platform (both for read and write operations) through REST APIs which will comply with state-of-the-art network standards (e.g. TLS >= 1.2) and will also use proper authentication methods. All the technical details about communication and authentication methods will be incrementally defined and refined in WP4 during the design and development phase of the project, which will last until the last iteration of requirement collection (M24).

The third zone of the 3-tier architecture is the Internet where no control may be enforced by the Information System Departments of any business. In CAPABLE the Internet will be exploited chiefly for reaching out to patients through their Smartphones. In this way they receive notifications by the CAPABLE system concerning advice to provide data useful for a better follow-up by the clinical staff or coaching actions with suggestions aimed at improving their compliance to the treatment. Through the same link patients may also spontaneously enter Patient Reported Outcomes or Patient Reported Experiences without being solicited by the system. Finally the same Smartphone link is exploited by clinical or ambient Sensors that regularly and autonomously provide information about some clinical parameters (e.g. temperature, heartbeat, blood pressure) or environmental conditions (e.g. humidity, temperature, location, etc...). The link to the Smartphones is provided by the Front-End Server Component located on the DMZ which on its turn has access to the CAPABLE Components on the Intranet for providing the relevant services.

The same Front End Server Component also hosts the Web Application that is accessed by the Desktop stations of the medical staff to oversee the patients enrolled in the study. If the medical staff will only connect from within the institution premises (i.e. the Intranet) accessing this service will be disabled from the Internet zone.

The general architecture (and procedures, which must be considered altogether) of CAPABLE complies with privacy. By design, in fact:

- All patient's data inside the Data Platform (the only place where patient's data can stay) are pseudonymized. The link between CAPABLE id and hospital id is managed (like in every study/trial) separately by the hospital.
- We will collect exclusively data that are relevant for the project: because it is generated by some CAPABLE component or because it is among those that hospitals allowed to be imported from their HIS.
- Also, the data access, thanks to the fact that it is not direct on the database but mediated by a software layer (the FHIR APIs indeed) allow to limit access to particular data for specific

components. Again, this is something that we, as a team, have not decided yet; but if we are arguing about design: using APIs is the best way to possibly do it.

- The loosely coupled design and the fact that components will interact with web services allows to integrate state-of-the-art authentication and encryption technologies: these have not been chosen yet (we plan to do it next year), but design is compliant with them. Furthermore, having http call as the only way of exchanging information between components, allows a fine network route tuning.
- Storage and processing of data are separated.

4.2 Is the data safely stored in certified repositories for long term preservation and curation?

In CAPABLE the only Component devoted to persistent data storage is the Data Platform, which is located in the Hospital Intranet. For long-term preservation of data, also beyond the validation studies completion, overall project duration and 2-5 years audit-required data availability, data will be deposited in the Zenodo CAPABLE community, which is managed in accordance to the open access policies that openAIRE project defined for H2020 projects. More details about safety, access control, data preservation, etc. can be found in Zenodo’s own policies (https://about.zenodo.org/policies/).

5. Ethical aspects

5.1 Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Deliverables from the CAPABLE's WP7 thoroughly address this aspect, in particular D7.2[8], and D7.6[9]. Comprehensive regulatory and ethical documentation is included in those publicly accessible deliverables.

5.2 Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Yes, D7.1 and D7.6 [9,10] are dedicated to the ethical and regulatory requirements for the studies, including a substantial section on informed consent for study participants.

6. Other issues

6.1 Do you make use of other national/ funder/ sectorial/ departmental procedures for data management? If yes, which ones?

ICSM and BARI have their own DMP, according to the Italian and European policies on research and clinical data management, which applies to data generated and managed in the context of the project by ICSM and BARI. The hospital information system is integrated with regional health information systems; data management procedures (digital signature, data transmission, etc.) are issued by Lombardia and Puglia regional authorities. Both institutions have internal procedures that define how backups are performed and verified, the criteria for implementing business continuity and disaster recovery within their IT infrastructure, the standards followed for the implementation of database servers and for the encryption of data storage.

The NKI has set strict conditions for the management of research data. According to The Netherlands Code of Conduct for Academic Practice and in accordance with NKI's policy on research data management, such data will be archived for at least 10 years, together with their accompanying metadata and documentation necessary to understand the data.

The only sensible data that AIMAC owns is for its staff. All the other data collected from patients are anonymously stored within an internal database. In case of sensible data, AIMAC makes use of an external collaborator expert in data protection.

7. Glossary

AIMAC	Associazione Italiana MAlati di Cancro (Italian association of cancer patients)
BARI	Policlinic of Bari
BIOM	Biomeris
CC	Creative Commons
DMP	Data Management Plan
DoA	Description of Action
FAIR	Findable, Accessible, Interoperable and Re-usable
GDPR	EU General Data Protection Regulation
H2020	Horizon 2020 framework programme
ICSM	Istituti Clinici Scientifici Maugeri (Clinical and research institute Maugeri)
NKI	Netherlands Cancer Institute
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
ORDP	Open Research Data Pilot
PRO	Patient Reported Outcome
PRE	Patient Reported Experience

8. References

[1] J.M. Overhage, P.B. Ryan, C.G. Reich, A.G. Hartzema, P.E. Stang, Validation of a common data model for active safety surveillance research, J Am Med Inform Assoc. 19 (2012) 54–60. https://doi.org/10.1136/amiajnl-2011-000376.

[2] M.J. Hjermstad, S.D. Fossa, K. Bjordal, S. Kaasa, Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire, J. Clin. Oncol. 13 (1995) 1249–1254. https://doi.org/10.1200/JCO.1995.13.5.1249.

[3] L.V. van de Poll-Franse, F. Mols, C.M. Gundy, C.L. Creutzberg, R.A. Nout, I.M. Verdonck-de Leeuw, M.J. Taphoorn, N.K. Aaronson, Normative data for the EORTC QLQ-C30 and EORTC-sexuality items in the general Dutch population, Eur. J. Cancer. 47 (2011) 667–675. https://doi.org/10.1016/j.ejca.2010.11.004.

[4] R. Cornet, F. Polce, R. de Groot, CAPABLE D3.1: Information Architecture, (2021). https://doi.org/10.5281/zenodo.5159073.

[5] R. Cornet, F. Polce, S. Quaglini, M. Peleg, S. Glaser, R. de Groot, S. Medlock, CAPABLE D3.2: Data-related Functionality to Realize a FAIR Infrastructure, (2021). https://doi.org/10.5281/zenodo.5005509.

[6] R. Cornet, S. Quaglini, CAPABLE D3.3: Specification of the Information Architecture and Data Modeling Based on FAIR Principles, (2023). https://doi.org/10.5281/zenodo.7671898.

[7] E. Parimbelli, R. Cornet, M. Gabetta, V. Tibollo, B. Bottalico, I. Fraterman, A. Boekhout, F. Gilboa-Solomon, S. Quaglini, CAPABLE D1.2: Data Management Plan, (2020). https://doi.org/10.5281/zenodo.3970580.

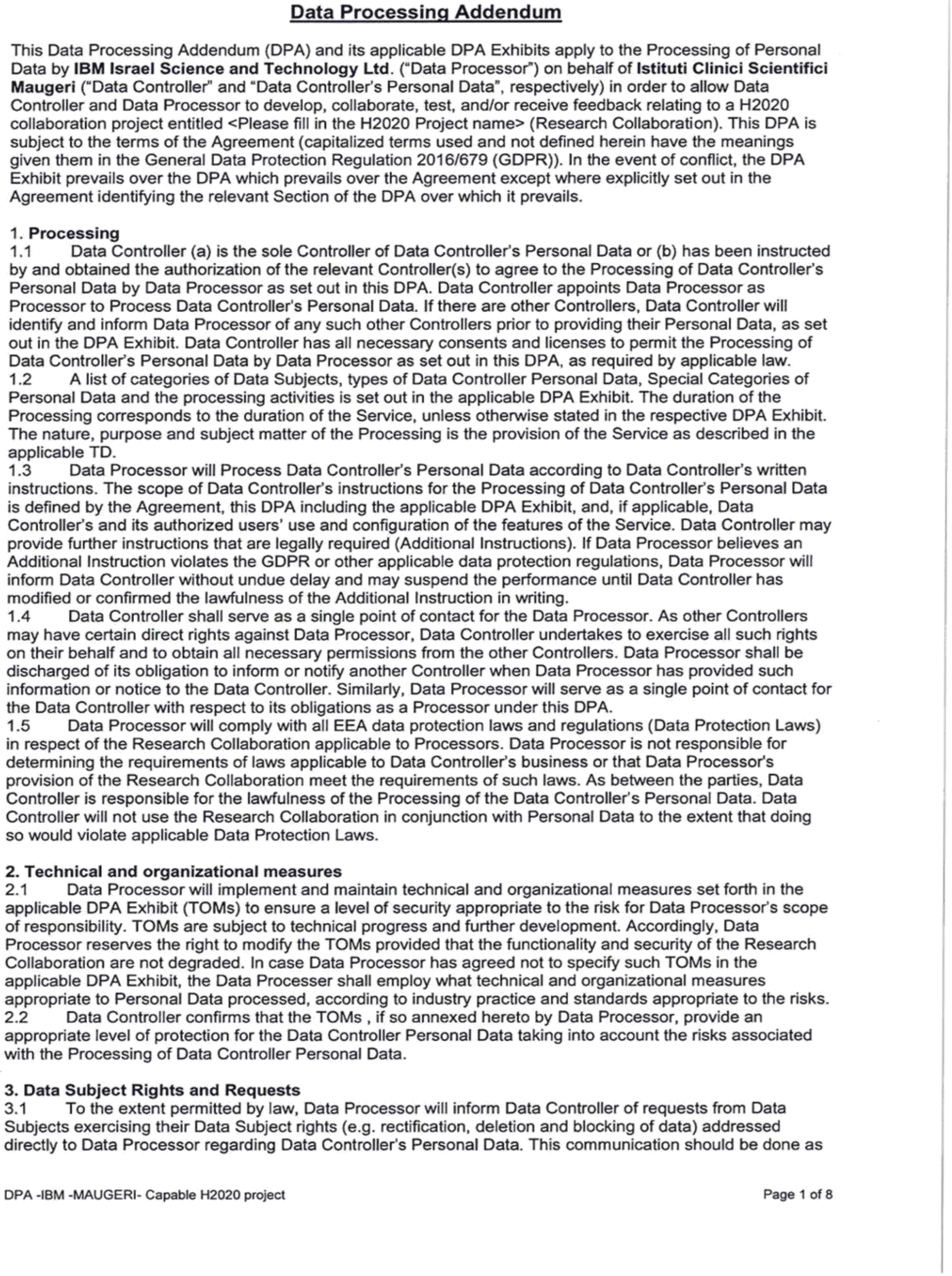
[8] L. Sacchi, B. Bottalico, G. Lanzola, S. Panzarasa, E. Parimbelli, F. Polce, S. Quaglini, A. Kogan, R. Leizer, M. Gabetta, R. Cornet, S. Medlock, E. Barkan, F. Gilboa-Solomon, S. Wilk, V. Ghio, V. Tibollo, I. Fraterman, D. Glasspool, M. Ottaviano, CAPABLE D7.2: AI Ethics and Incidental Findings Policy, (2020). https://doi.org/10.5281/zenodo.4540520.

[9] S. Quaglini, L. Sacchi, M. Peleg, V. Tibollo, CAPABLE D7.6: Informed Patient Consent/Assent Form, Ethical Committee Approval, Training Materials, and Technical Manual for Maintenance, (2022). https://doi.org/10.5281/zenodo.7603369.

[10] L. Sacchi, E. Girani, S. Panzarasa, E. Parimbelli, S. Quaglini, N. Veggiotti, M. Peleg, V. Ghio, M. Rizzo, V. Tibollo, A. Boekhout, I. Fraterman, M. Ottaviano, CAPABLE D7.1: Study Plan, Protocols Definition, and Informed Consent/Assent Drafts, (2020). https://doi.org/10.5281/zenodo.4540503.

9. Annexes

Annex I - DPA



Data Processing Addendum

This Data Processing Addendum (DPA) and its applicable DPA Exhibits apply to the Processing of Personal Data by **IBM Israel Science and Technology Ltd**. ("Data Processor") on behalf of **Istituti Clinici Scientifici Maugeri** ("Data Controller" and "Data Controller's Personal Data", respectively) in order to allow Data Controller and Data Processor to develop, collaborate, test, and/or receive feedback relating to a H2020 collaboration project entitled <Please fill in the H2020 Project name> (Research Collaboration). This DPA is subject to the terms of the Agreement (capitalized terms used and not defined herein have the meanings given them in the General Data Protection Regulation 2016/679 (GDPR)). In the event of conflict, the DPA Exhibit prevails over the DPA which prevails over the Agreement except where explicitly set out in the Agreement identifying the relevant Section of the DPA over which it prevails.

1. **Processing**

1.1 Data Controller (a) is the sole Controller of Data Controller's Personal Data or (b) has been instructed by and obtained the authorization of the relevant Controller(s) to agree to the Processing of Data Controller's Personal Data by Data Processor as set out in this DPA. Data Controller appoints Data Processor as Processor to Process Data Controller's Personal Data. If there are other Controllers, Data Controller will identify and inform Data Processor of any such other Controllers prior to providing their Personal Data, as set out in the DPA Exhibit. Data Controller has all necessary consents and licenses to permit the Processing of Data Controller's Personal Data by Data Processor as set out in this DPA, as required by applicable law.

1.2 A list of categories of Data Subjects, types of Data Controller Personal Data, Special Categories of Personal Data and the processing activities is set out in the applicable DPA Exhibit. The duration of the Processing corresponds to the duration of the Service, unless otherwise stated in the respective DPA Exhibit. The nature, purpose and subject matter of the Processing is the provision of the Service as described in the applicable TD.

1.3 Data Processor will Process Data Controller's Personal Data according to Data Controller's written instructions. The scope of Data Controller's instructions for the Processing of Data Controller's Personal Data is defined by the Agreement, this DPA including the applicable DPA Exhibit, and, if applicable, Data Controller's and its authorized users' use and configuration of the features of the Service. Data Controller may provide further instructions that are legally required (Additional Instructions). If Data Processor believes an Additional Instruction violates the GDPR or other applicable data protection regulations, Data Processor will inform Data Controller without undue delay and may suspend the performance until Data Controller has modified or confirmed the lawfulness of the Additional Instruction in writing.

1.4 Data Controller shall serve as a single point of contact for the Data Processor. As other Controllers may have certain direct rights against Data Processor, Data Controller undertakes to exercise all such rights on their behalf and to obtain all necessary permissions from the other Controllers. Data Processor shall be discharged of its obligation to inform or notify another Controller when Data Processor has provided such information or notice to the Data Controller. Similarly, Data Processor will serve as a single point of contact for the Data Controller with respect to its obligations as a Processor under this DPA.

1.5 Data Processor will comply with all EEA data protection laws and regulations (Data Protection Laws) in respect of the Research Collaboration applicable to Processors. Data Processor is not responsible for determining the requirements of laws applicable to Data Controller's business or that Data Processor's provision of the Research Collaboration meet the requirements of such laws. As between the parties, Data Controller is responsible for the lawfulness of the Processing of the Data Controller's Personal Data. Data Controller will not use the Research Collaboration in conjunction with Personal Data to the extent that doing so would violate applicable Data Protection Laws.

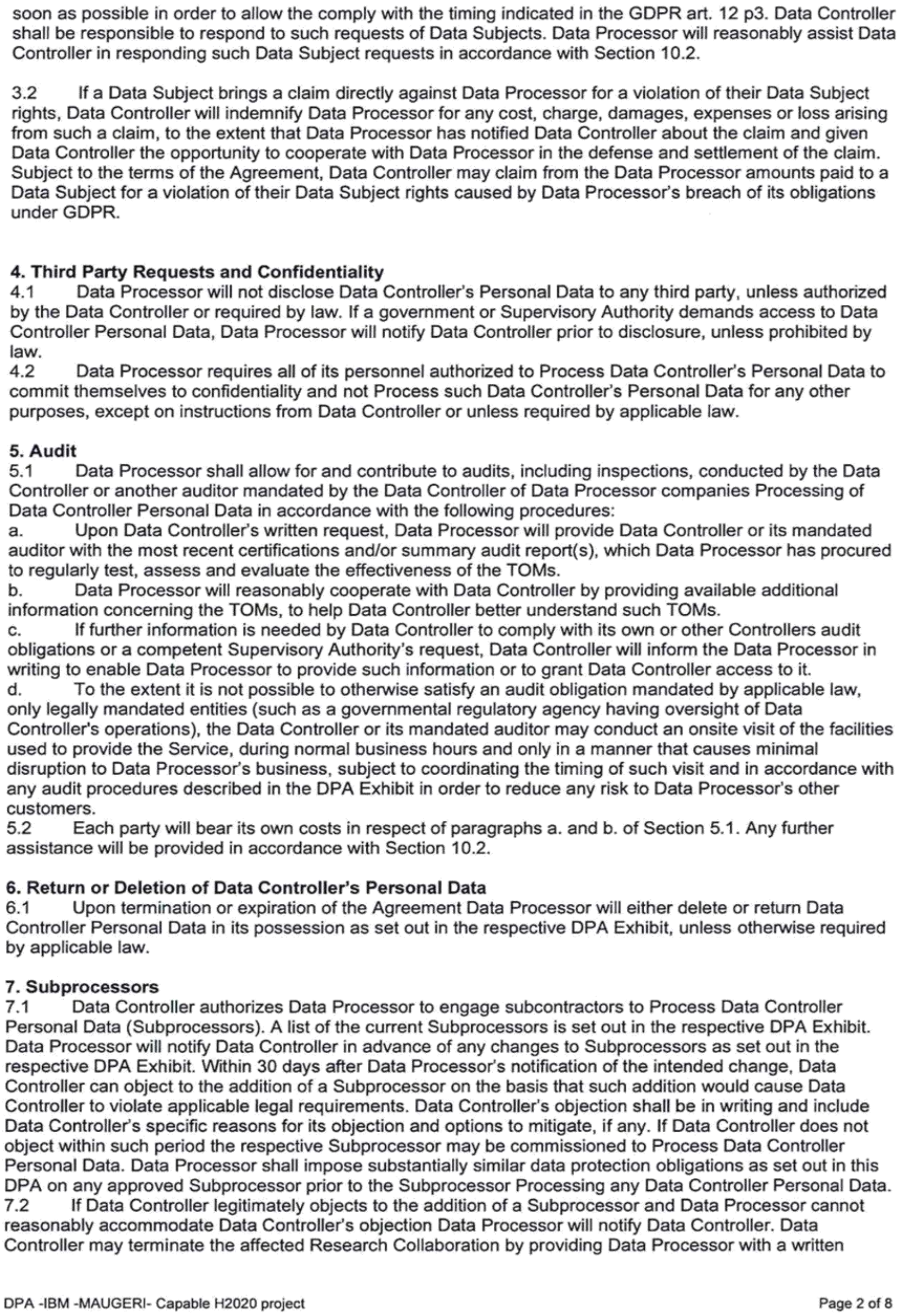
2. Technical and organizational measures

2.1 Data Processor will implement and maintain technical and organizational measures set forth in the applicable DPA Exhibit (TOMs) to ensure a level of security appropriate to the risk for Data Processor's scope of responsibility. TOMs are subject to technical progress and further development. Accordingly, Data Processor reserves the right to modify the TOMs provided that the functionality and security of the Research Collaboration are not degraded. In case Data Processor has agreed not to specify such TOMs in the applicable DPA Exhibit, the Data Processer shall employ what technical and organizational measures appropriate to Personal Data processed, according to industry practice and standards appropriate to the risks.

2.2 Data Controller confirms that the TOMs , if so annexed hereto by Data Processor, provide an appropriate level of protection for the Data Controller Personal Data taking into account the risks associated with the Processing of Data Controller Personal Data.

3. Data Subject Rights and Requests

3.1 To the extent permitted by law, Data Processor will inform Data Controller of requests from Data Subjects exercising their Data Subject rights (e.g. rectification, deletion and blocking of data) addressed directly to Data Processor regarding Data Controller's Personal Data. This communication should be done as



soon as possible in order to allow the comply with the timing indicated in the GDPR art. 12 p3. Data Controller shall be responsible to respond to such requests of Data Subjects. Data Processor will reasonably assist Data Controller in responding such Data Subject requests in accordance with Section 10.2.

3.2 If a Data Subject brings a claim directly against Data Processor for a violation of their Data Subject rights, Data Controller will indemnify Data Processor for any cost, charge, damages, expenses or loss arising from such a claim, to the extent that Data Processor has notified Data Controller about the claim and given Data Controller the opportunity to cooperate with Data Processor in the defense and settlement of the claim. Subject to the terms of the Agreement, Data Controller may claim from the Data Processor amounts paid to a Data Subject for a violation of their Data Subject rights caused by Data Processor's breach of its obligations under GDPR.

4. Third Party Requests and Confidentiality

4.1 Data Processor will not disclose Data Controller's Personal Data to any third party, unless authorized by the Data Controller or required by law. If a government or Supervisory Authority demands access to Data Controller Personal Data, Data Processor will notify Data Controller prior to disclosure, unless prohibited by law.

4.2 Data Processor requires all of its personnel authorized to Process Data Controller's Personal Data to commit themselves to confidentiality and not Process such Data Controller's Personal Data for any other purposes, except on instructions from Data Controller or unless required by applicable law.

5. Audit

5.1 Data Processor shall allow for and contribute to audits, including inspections, conducted by the Data Controller or another auditor mandated by the Data Controller of Data Processor companies Processing of Data Controller Personal Data in accordance with the following procedures:

a. Upon Data Controller's written request, Data Processor will provide Data Controller or its mandated auditor with the most recent certifications and/or summary audit report(s), which Data Processor has procured to regularly test, assess and evaluate the effectiveness of the TOMs.

b. Data Processor will reasonably cooperate with Data Controller by providing available additional information concerning the TOMs, to help Data Controller better understand such TOMs.

c. If further information is needed by Data Controller to comply with its own or other Controllers audit obligations or a competent Supervisory Authority's request, Data Controller will inform the Data Processor in writing to enable Data Processor to provide such information or to grant Data Controller access to it.

d. To the extent it is not possible to otherwise satisfy an audit obligation mandated by applicable law, only legally mandated entities (such as a governmental regulatory agency having oversight of Data Controller's operations), the Data Controller or its mandated auditor may conduct an onsite visit of the facilities used to provide the Service, during normal business hours and only in a manner that causes minimal disruption to Data Processor's business, subject to coordinating the timing of such visit and in accordance with any audit procedures described in the DPA Exhibit in order to reduce any risk to Data Processor's other customers.

5.2 Each party will bear its own costs in respect of paragraphs a. and b. of Section 5.1. Any further assistance will be provided in accordance with Section 10.2.

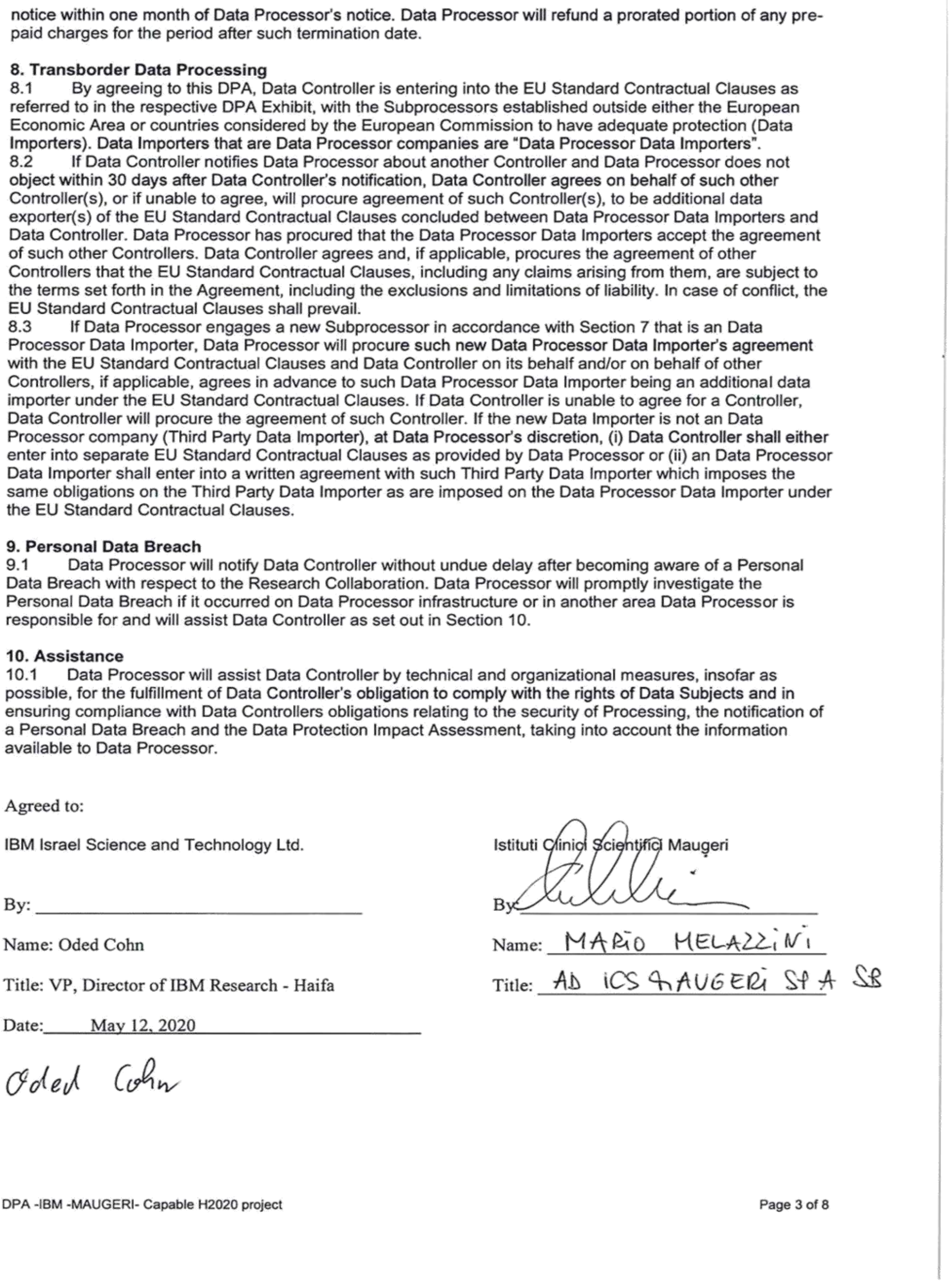
6. Return or Deletion of Data Controller's Personal Data

6.1 Upon termination or expiration of the Agreement Data Processor will either delete or return Data Controller Personal Data in its possession as set out in the respective DPA Exhibit, unless otherwise required by applicable law.

7. Subprocessors

7.1 Data Controller authorizes Data Processor to engage subcontractors to Process Data Controller Personal Data (Subprocessors). A list of the current Subprocessors is set out in the respective DPA Exhibit. Data Processor will notify Data Controller in advance of any changes to Subprocessors as set out in the respective DPA Exhibit. Within 30 days after Data Processor's notification of the intended change, Data Controller can object to the addition of a Subprocessor on the basis that such addition would cause Data Controller to violate applicable legal requirements. Data Controller's objection shall be in writing and include Data Controller's specific reasons for its objection and options to mitigate, if any. If Data Controller does not object within such period the respective Subprocessor may be commissioned to Process Data Controller Personal Data. Data Processor shall impose substantially similar data protection obligations as set out in this DPA on any approved Subprocessor prior to the Subprocessor Processing any Data Controller Personal Data.

7.2 If Data Controller legitimately objects to the addition of a Subprocessor and Data Processor cannot reasonably accommodate Data Controller's objection Data Processor will notify Data Controller. Data Controller may terminate the affected Research Collaboration by providing Data Processor with a written



notice within one month of Data Processor's notice. Data Processor will refund a prorated portion of any pre-paid charges for the period after such termination date.

8. Transborder Data Processing

8.1 By agreeing to this DPA, Data Controller is entering into the EU Standard Contractual Clauses as referred to in the respective DPA Exhibit, with the Subprocessors established outside either the European Economic Area or countries considered by the European Commission to have adequate protection (Data Importers). Data Importers that are Data Processor companies are "Data Processor Data Importers".

8.2 If Data Controller notifies Data Processor about another Controller and Data Processor does not object within 30 days after Data Controller's notification, Data Controller agrees on behalf of such other Controller(s), or if unable to agree, will procure agreement of such Controller(s), to be additional data exporter(s) of the EU Standard Contractual Clauses concluded between Data Processor Data Importers and Data Controller. Data Processor has procured that the Data Processor Data Importers accept the agreement of such other Controllers. Data Controller agrees and, if applicable, procures the agreement of other Controllers that the EU Standard Contractual Clauses, including any claims arising from them, are subject to the terms set forth in the Agreement, including the exclusions and limitations of liability. In case of conflict, the EU Standard Contractual Clauses shall prevail.

8.3 If Data Processor engages a new Subprocessor in accordance with Section 7 that is an Data Processor Data Importer, Data Processor will procure such new Data Processor Data Importer's agreement with the EU Standard Contractual Clauses and Data Controller on its behalf and/or on behalf of other Controllers, if applicable, agrees in advance to such Data Processor Data Importer being an additional data importer under the EU Standard Contractual Clauses. If Data Controller is unable to agree for a Controller, Data Controller will procure the agreement of such Controller. If the new Data Importer is not an Data Processor company (Third Party Data Importer), at Data Processor's discretion, (i) Data Controller shall either enter into separate EU Standard Contractual Clauses as provided by Data Processor or (ii) an Data Processor Data Importer shall enter into a written agreement with such Third Party Data Importer which imposes the same obligations on the Third Party Data Importer as are imposed on the Data Processor Data Importer under the EU Standard Contractual Clauses.

9. Personal Data Breach

9.1 Data Processor will notify Data Controller without undue delay after becoming aware of a Personal Data Breach with respect to the Research Collaboration. Data Processor will promptly investigate the Personal Data Breach if it occurred on Data Processor infrastructure or in another area Data Processor is responsible for and will assist Data Controller as set out in Section 10.

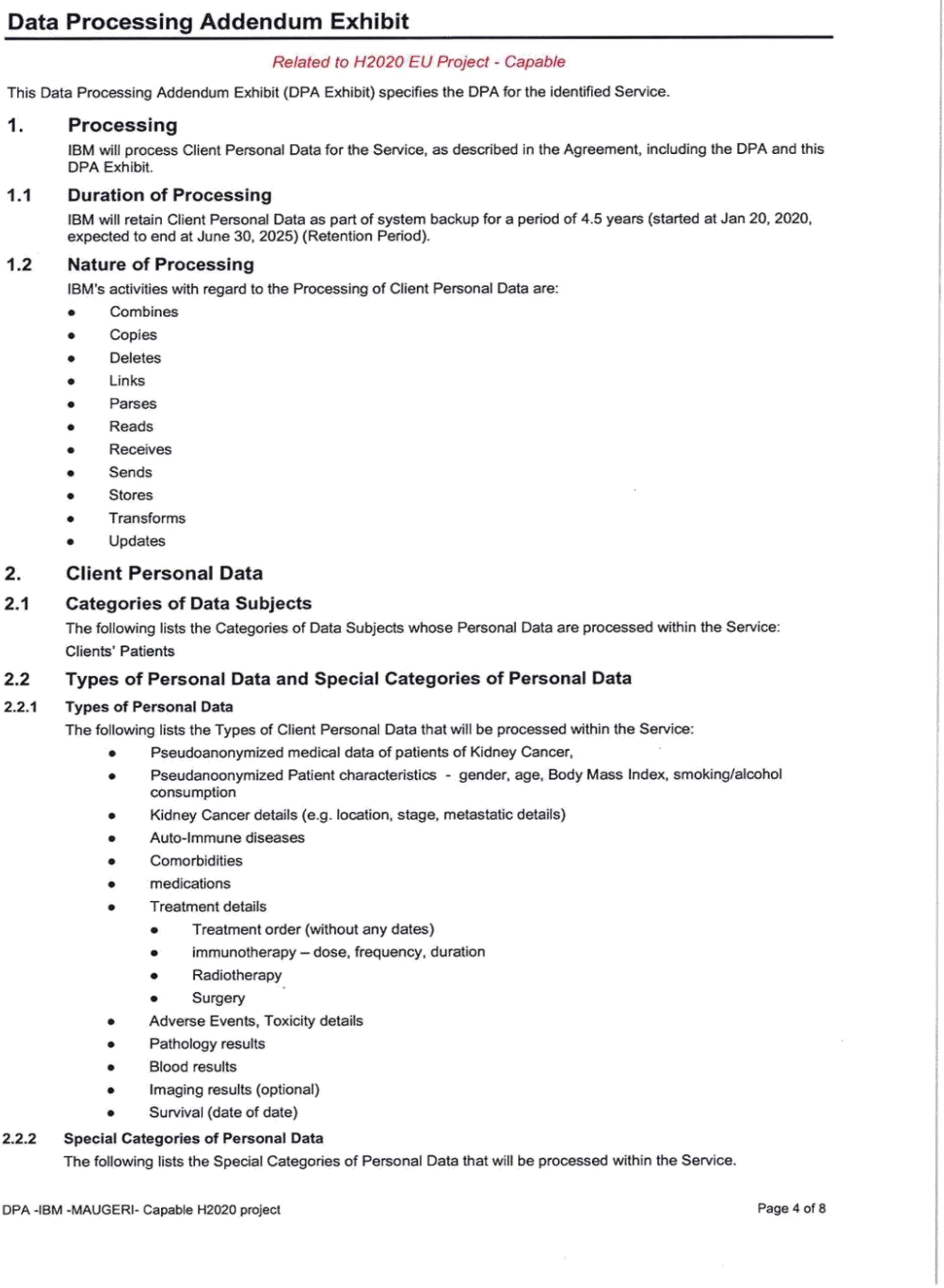
10. Assistance

10.1 Data Processor will assist Data Controller by technical and organizational measures, insofar as possible, for the fulfillment of Data Controller's obligation to comply with the rights of Data Subjects and in ensuring compliance with Data Controllers obligations relating to the security of Processing, the notification of a Personal Data Breach and the Data Protection Impact Assessment, taking into account the information available to Data Processor.

Agreed to:

IBM Israel Science and Technology Ltd.	Istituti Clinici Scientifici Maugeri
By: ______________________	By: ______________________
Name: Oded Cohn	Name: MARIO MELAZZINI
Title: VP, Director of IBM Research - Haifa	Title: AD ICS MAUGERI SPA SB
Date: May 12, 2020	

Oded Cohn



Data Processing Addendum Exhibit

Related to H2020 EU Project - Capable

This Data Processing Addendum Exhibit (DPA Exhibit) specifies the DPA for the identified Service.

1. Processing

IBM will process Client Personal Data for the Service, as described in the Agreement, including the DPA and this DPA Exhibit.

1.1 Duration of Processing

IBM will retain Client Personal Data as part of system backup for a period of 4.5 years (started at Jan 20, 2020, expected to end at June 30, 2025) (Retention Period).

1.2 Nature of Processing

IBM's activities with regard to the Processing of Client Personal Data are:

- Combines
- Copies
- Deletes
- Links
- Parses
- Reads
- Receives
- Sends
- Stores
- Transforms
- Updates

2. Client Personal Data

2.1 Categories of Data Subjects

The following lists the Categories of Data Subjects whose Personal Data are processed within the Service:

Clients' Patients

2.2 Types of Personal Data and Special Categories of Personal Data

2.2.1 Types of Personal Data

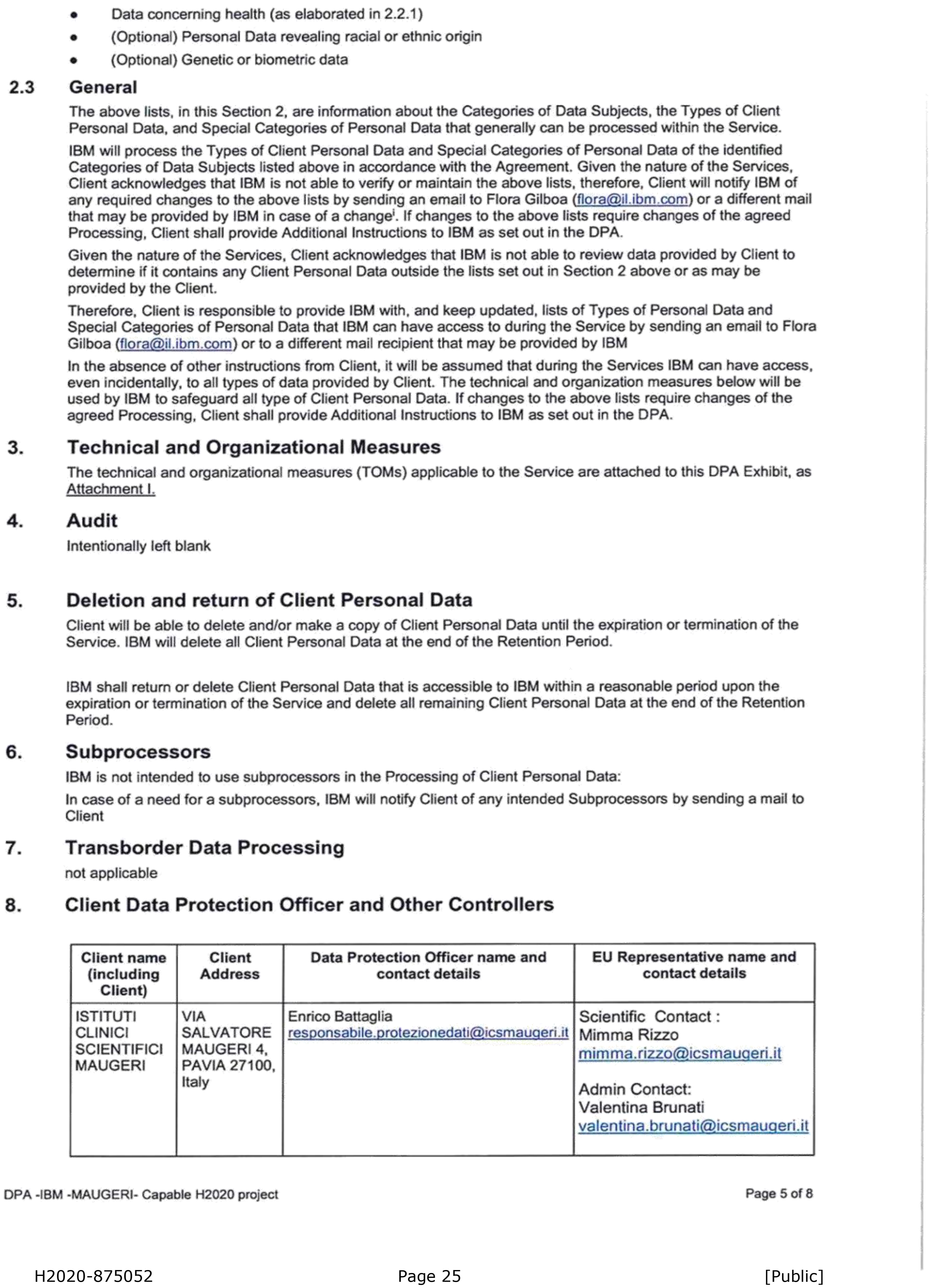
The following lists the Types of Client Personal Data that will be processed within the Service:

- Pseudoanonymized medical data of patients of Kidney Cancer,
- Pseudanoonymized Patient characteristics - gender, age, Body Mass Index, smoking/alcohol consumption
- Kidney Cancer details (e.g. location, stage, metastatic details)
- Auto-Immune diseases
- Comorbidities
- medications
- Treatment details
  - Treatment order (without any dates)
  - immunotherapy – dose, frequency, duration
  - Radiotherapy
  - Surgery
- Adverse Events, Toxicity details
- Pathology results
- Blood results
- Imaging results (optional)
- Survival (date of date)

2.2.2 Special Categories of Personal Data

The following lists the Special Categories of Personal Data that will be processed within the Service.

DPA -IBM -MAUGERI- Capable H2020 project Page 4 of 8



- Data concerning health (as elaborated in 2.2.1)
- (Optional) Personal Data revealing racial or ethnic origin
- (Optional) Genetic or biometric data

2.3 General

The above lists, in this Section 2, are information about the Categories of Data Subjects, the Types of Client Personal Data, and Special Categories of Personal Data that generally can be processed within the Service.

IBM will process the Types of Client Personal Data and Special Categories of Personal Data of the identified Categories of Data Subjects listed above in accordance with the Agreement. Given the nature of the Services, Client acknowledges that IBM is not able to verify or maintain the above lists, therefore, Client will notify IBM of any required changes to the above lists by sending an email to Flora Gilboa (flora@il.ibm.com) or a different mail that may be provided by IBM in case of a change[i]. If changes to the above lists require changes of the agreed Processing, Client shall provide Additional Instructions to IBM as set out in the DPA.

Given the nature of the Services, Client acknowledges that IBM is not able to review data provided by Client to determine if it contains any Client Personal Data outside the lists set out in Section 2 above or as may be provided by the Client.

Therefore, Client is responsible to provide IBM with, and keep updated, lists of Types of Personal Data and Special Categories of Personal Data that IBM can have access to during the Service by sending an email to Flora Gilboa (flora@il.ibm.com) or to a different mail recipient that may be provided by IBM

In the absence of other instructions from Client, it will be assumed that during the Services IBM can have access, even incidentally, to all types of data provided by Client. The technical and organization measures below will be used by IBM to safeguard all type of Client Personal Data. If changes to the above lists require changes of the agreed Processing, Client shall provide Additional Instructions to IBM as set out in the DPA.

3. Technical and Organizational Measures

The technical and organizational measures (TOMs) applicable to the Service are attached to this DPA Exhibit, as Attachment I.

4. Audit

Intentionally left blank

5. Deletion and return of Client Personal Data

Client will be able to delete and/or make a copy of Client Personal Data until the expiration or termination of the Service. IBM will delete all Client Personal Data at the end of the Retention Period.

IBM shall return or delete Client Personal Data that is accessible to IBM within a reasonable period upon the expiration or termination of the Service and delete all remaining Client Personal Data at the end of the Retention Period.

6. Subprocessors

IBM is not intended to use subprocessors in the Processing of Client Personal Data:

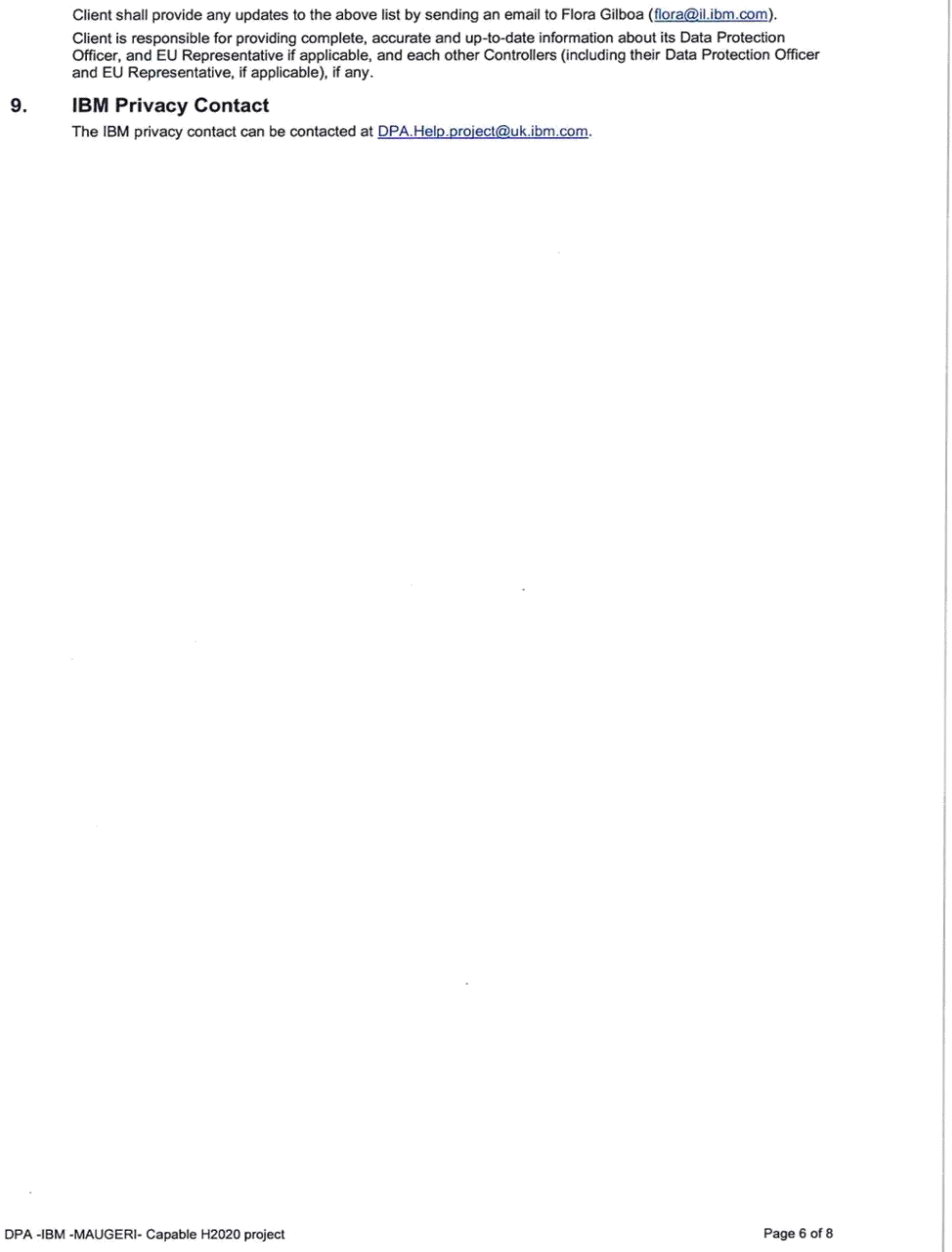
In case of a need for a subprocessors, IBM will notify Client of any intended Subprocessors by sending a mail to Client

7. Transborder Data Processing

not applicable

8. Client Data Protection Officer and Other Controllers

Client name (including Client)	Client Address	Data Protection Officer name and contact details	EU Representative name and contact details
ISTITUTI CLINICI SCIENTIFICI MAUGERI	VIA SALVATORE MAUGERI 4, PAVIA 27100, Italy	Enrico Battaglia responsabile.protezionedati@icsmaugeri.it	Scientific Contact : Mimma Rizzo mimma.rizzo@icsmaugeri.it Admin Contact: Valentina Brunati valentina.brunati@icsmaugeri.it

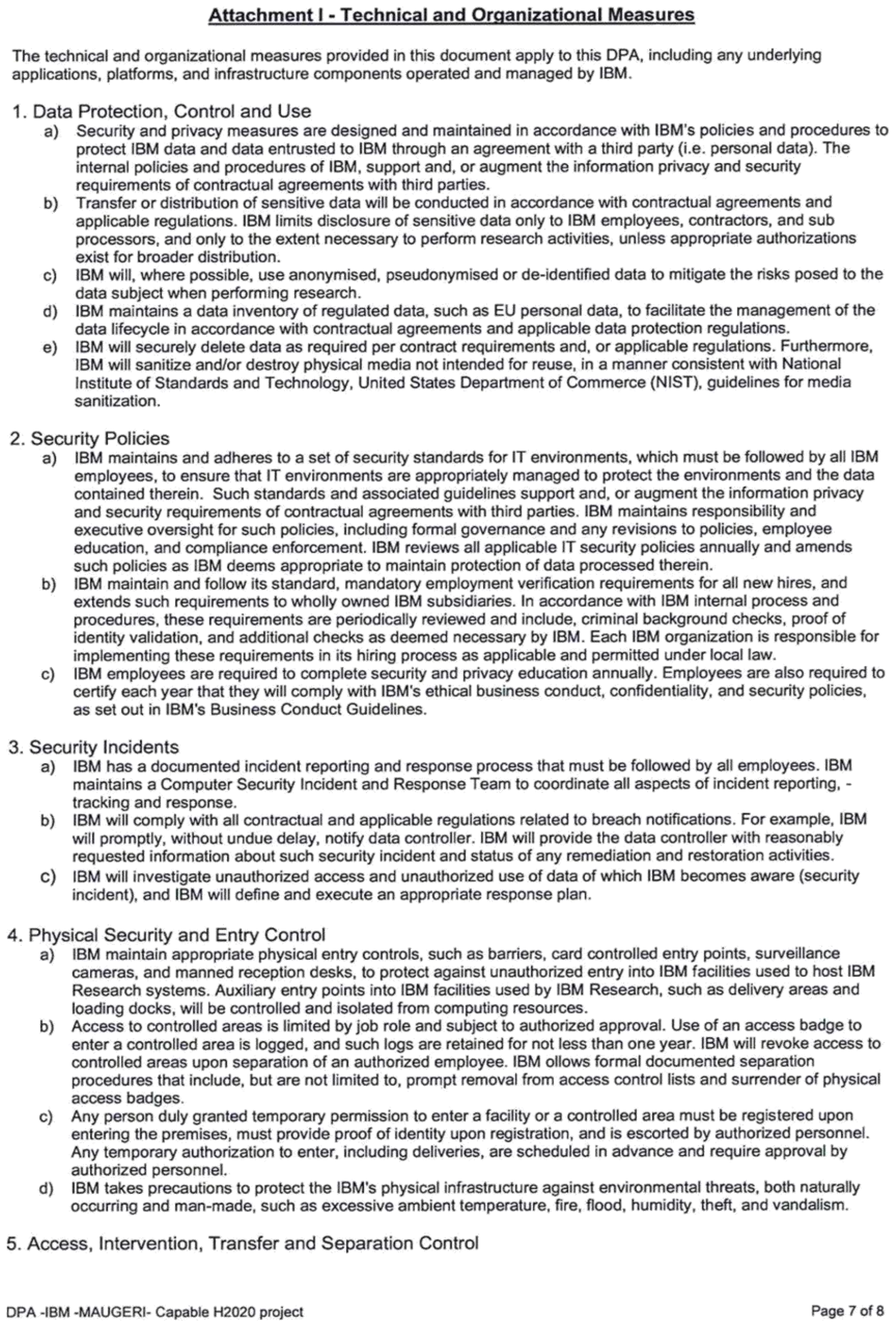


Client shall provide any updates to the above list by sending an email to Flora Gilboa (flora@il.ibm.com).

Client is responsible for providing complete, accurate and up-to-date information about its Data Protection Officer, and EU Representative if applicable, and each other Controllers (including their Data Protection Officer and EU Representative, if applicable), if any.

9. IBM Privacy Contact

The IBM privacy contact can be contacted at DPA.Help.project@uk.ibm.com.



Attachment I - Technical and Organizational Measures

The technical and organizational measures provided in this document apply to this DPA, including any underlying applications, platforms, and infrastructure components operated and managed by IBM.

1. Data Protection, Control and Use

a) Security and privacy measures are designed and maintained in accordance with IBM's policies and procedures to protect IBM data and data entrusted to IBM through an agreement with a third party (i.e. personal data). The internal policies and procedures of IBM, support and, or augment the information privacy and security requirements of contractual agreements with third parties.

b) Transfer or distribution of sensitive data will be conducted in accordance with contractual agreements and applicable regulations. IBM limits disclosure of sensitive data only to IBM employees, contractors, and sub processors, and only to the extent necessary to perform research activities, unless appropriate authorizations exist for broader distribution.

c) IBM will, where possible, use anonymised, pseudonymised or de-identified data to mitigate the risks posed to the data subject when performing research.

d) IBM maintains a data inventory of regulated data, such as EU personal data, to facilitate the management of the data lifecycle in accordance with contractual agreements and applicable data protection regulations.

e) IBM will securely delete data as required per contract requirements and, or applicable regulations. Furthermore, IBM will sanitize and/or destroy physical media not intended for reuse, in a manner consistent with National Institute of Standards and Technology, United States Department of Commerce (NIST), guidelines for media sanitization.

2. Security Policies

a) IBM maintains and adheres to a set of security standards for IT environments, which must be followed by all IBM employees, to ensure that IT environments are appropriately managed to protect the environments and the data contained therein. Such standards and associated guidelines support and, or augment the information privacy and security requirements of contractual agreements with third parties. IBM maintains responsibility and executive oversight for such policies, including formal governance and any revisions to policies, employee education, and compliance enforcement. IBM reviews all applicable IT security policies annually and amends such policies as IBM deems appropriate to maintain protection of data processed therein.

b) IBM maintain and follow its standard, mandatory employment verification requirements for all new hires, and extends such requirements to wholly owned IBM subsidiaries. In accordance with IBM internal process and procedures, these requirements are periodically reviewed and include, criminal background checks, proof of identity validation, and additional checks as deemed necessary by IBM. Each IBM organization is responsible for implementing these requirements in its hiring process as applicable and permitted under local law.

c) IBM employees are required to complete security and privacy education annually. Employees are also required to certify each year that they will comply with IBM's ethical business conduct, confidentiality, and security policies, as set out in IBM's Business Conduct Guidelines.

3. Security Incidents

a) IBM has a documented incident reporting and response process that must be followed by all employees. IBM maintains a Computer Security Incident and Response Team to coordinate all aspects of incident reporting, -tracking and response.

b) IBM will comply with all contractual and applicable regulations related to breach notifications. For example, IBM will promptly, without undue delay, notify data controller. IBM will provide the data controller with reasonably requested information about such security incident and status of any remediation and restoration activities.

c) IBM will investigate unauthorized access and unauthorized use of data of which IBM becomes aware (security incident), and IBM will define and execute an appropriate response plan.

4. Physical Security and Entry Control

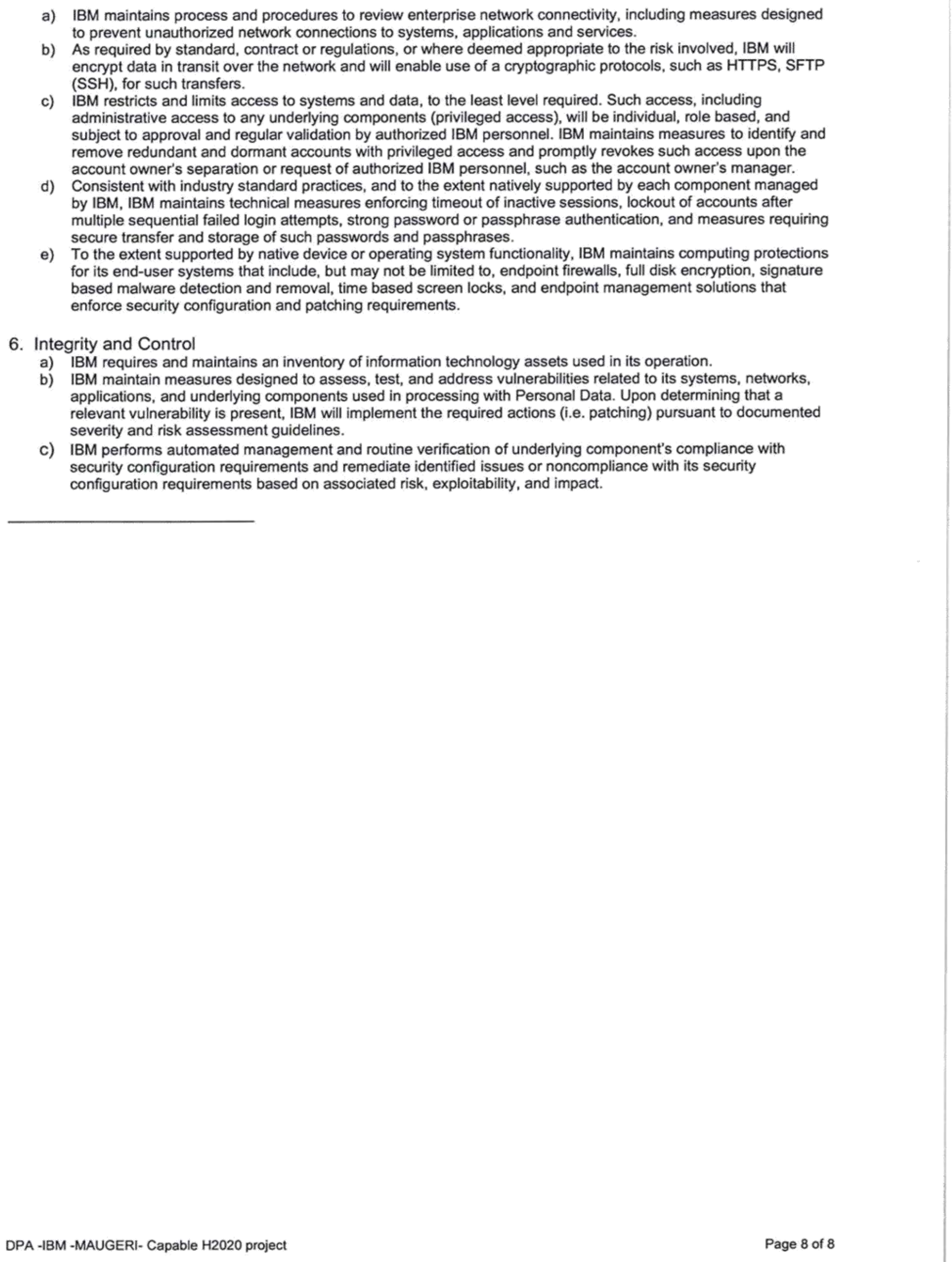
a) IBM maintain appropriate physical entry controls, such as barriers, card controlled entry points, surveillance cameras, and manned reception desks, to protect against unauthorized entry into IBM facilities used to host IBM Research systems. Auxiliary entry points into IBM facilities used by IBM Research, such as delivery areas and loading docks, will be controlled and isolated from computing resources.

b) Access to controlled areas is limited by job role and subject to authorized approval. Use of an access badge to enter a controlled area is logged, and such logs are retained for not less than one year. IBM will revoke access to controlled areas upon separation of an authorized employee. IBM ollows formal documented separation procedures that include, but are not limited to, prompt removal from access control lists and surrender of physical access badges.

c) Any person duly granted temporary permission to enter a facility or a controlled area must be registered upon entering the premises, must provide proof of identity upon registration, and is escorted by authorized personnel. Any temporary authorization to enter, including deliveries, are scheduled in advance and require approval by authorized personnel.

d) IBM takes precautions to protect the IBM's physical infrastructure against environmental threats, both naturally occurring and man-made, such as excessive ambient temperature, fire, flood, humidity, theft, and vandalism.

5. Access, Intervention, Transfer and Separation Control



a) IBM maintains process and procedures to review enterprise network connectivity, including measures designed to prevent unauthorized network connections to systems, applications and services.

b) As required by standard, contract or regulations, or where deemed appropriate to the risk involved, IBM will encrypt data in transit over the network and will enable use of a cryptographic protocols, such as HTTPS, SFTP (SSH), for such transfers.

c) IBM restricts and limits access to systems and data, to the least level required. Such access, including administrative access to any underlying components (privileged access), will be individual, role based, and subject to approval and regular validation by authorized IBM personnel. IBM maintains measures to identify and remove redundant and dormant accounts with privileged access and promptly revokes such access upon the account owner's separation or request of authorized IBM personnel, such as the account owner's manager.

d) Consistent with industry standard practices, and to the extent natively supported by each component managed by IBM, IBM maintains technical measures enforcing timeout of inactive sessions, lockout of accounts after multiple sequential failed login attempts, strong password or passphrase authentication, and measures requiring secure transfer and storage of such passwords and passphrases.

e) To the extent supported by native device or operating system functionality, IBM maintains computing protections for its end-user systems that include, but may not be limited to, endpoint firewalls, full disk encryption, signature based malware detection and removal, time based screen locks, and endpoint management solutions that enforce security configuration and patching requirements.

6. Integrity and Control

a) IBM requires and maintains an inventory of information technology assets used in its operation.

b) IBM maintain measures designed to assess, test, and address vulnerabilities related to its systems, networks, applications, and underlying components used in processing with Personal Data. Upon determining that a relevant vulnerability is present, IBM will implement the required actions (i.e. patching) pursuant to documented severity and risk assessment guidelines.

c) IBM performs automated management and routine verification of underlying component's compliance with security configuration requirements and remediate identified issues or noncompliance with its security configuration requirements based on associated risk, exploitability, and impact.

Annex II – DPIA

https://drive.google.com/file/d/1OySHQz47NfqrObZjiWsn0sIooyLwE8nc/view?usp=share_link
https://drive.google.com/file/d/1PKTLb9vRW_ifCWn5UTxqAqjYe9MXqAlN/view?usp=share_link