945358-BIGPICTURE
Central Repository for Digital Pathology

WP2 - Infrastructure and database hosting

D2.1 Report on initial requirements

Lead contributor	2 – Linköping University
	joel.hedlund@liu.se
Other contributors	5 – CSC - IT Center for Science Ltd.
	6 – Uppsala University
	38 – Janssen Pharmaceutica NV
	31 – Stichting Lygature

Document History

Version	Date	Description
V0.1	27 Mar 2021	First draft
V0.2	01 Apr 2021	Draft for WP1 and Consortium comment
V0.3	16 Apr 2021	Draft for WP1 and Consortium comment
V1.0	30 Apr 2021	Draft for PMO consideration
V1.1	11 May 2021	Draft for MB consideration
V1.2	28 May 2021	Final Version

Abstract

WP2 concerns development of a cutting-edge, GDPR compliant-by-design repository infrastructure for European digital pathology. Development will be based on existing services, which were originally developed for European genomics, namely ELIXIR AAI, REMS, (Federated) EGA, and Beacon. These services are in production use in European human health data research, at research institutions and at biobanks, for providing and accessing both services and data[1]. There is therefore good reason to believe their operational parameters well serve the needs in this kind of use. However, these services have not been extensively used by large European pharmaceutical companies, such as the BIGPICTURE EFPIA partners. Ascertaining whether new requirements could arise from such use is nontrivial.

BIGPICTURE WP2 therefore conducted a three month investigation at the start of the project, in Task 2.1 - Infrastructure requirements gathering. This report contains the findings of Task 2.1.

Task 2.1 aimed to establish initial requirements on infrastructure as understood by academic and industrial partners, to relate them to the proposed high-level architecture, and to outline suggestions on what developments may be necessary in order for these services to be used or adapted to support digital pathology. This work was carried out in parallel with Task 4.1 and in collaboration with consortium and WP leadership, to get input from the academic and EFPIA partners on requirements for clinical and non-clinical data.

This report includes initial specifications for backend data taking and use, metadata format and private inference execution model. It also includes an analysis of initial EFPIA information security and technical security requirements. These initial requirements are intended to provide a starting point for development, to be refined to final requirements in iterative development and stakeholder dialogue in Task 2.6.

Methods

Intended Data Use for different purposes is a key guiding requirement. For example, clinical ICT systems generally operate under requirements which might not be feasible to implement within the BIGPICTURE budget (like adhering to a portfolio of ISO standards), but could become relevant and feasible over the long-term 5-10 year roadmap.

Data collected in the BIGPICTURE repository could potentially be used for research, innovation, and even clinical processes like decision making. However, the decision how the data in the repository may be legally used is out of scope for the infrastructure work package, which will provide the technical platform.

A high-level analysis of the use cases and usage patterns likely to be supported by the infrastructure has been provided in the Description of Action.

WP2 used the following primary methods for requirement gathering, which are described in further detail in the sections below:

1. Expectation management
2. Invitation to contribute infrastructure requirements
3. Invitation to joint evaluation of security requirements
4. Instating the BIGPICTURE Architecture Task Force
5. Merging identified requirements

[1] The Federated EGA technology is production ready, but the contractual work with EGA is pending at the time of writing this deliverable.

1. Expectation management

During the BIGPICTURE WP2 kickoff on Feb 19, WP2 took the opportunity to identify and fact-check common misapprehensions that could lead to false expectations being placed on WP2, which could jeopardize feasibility/sustainability of the project. These were addressed and clarified in an expectation management document (cf Appendix 1) which was presented to the consortium at the broader BIGPICTURE kickoff on Mar 1-5. The purpose was to identify possible gaps in the proposed solution as understood by different consortium members, which could then be translated into new requirements.

The expectation management document was kept updated throughout Task 2.1. The intention is to keep it up to date throughout the remainder of the project.

2. Invitation to contribute infrastructure requirements

Starting with the Expectation management document, WP2 made every effort to raise and communicate through official channels the urgent need for broad engagement on infrastructure requirement gathering, inviting all parties to reach out to WP2 to contribute relevant concerns or information to the infrastructure requirement gathering effort.

WP2 took care to communicate to the consortium that WP2 would implement the repository infrastructure using Agile methodology, so as to be able to course correct during development in case new and potentially disruptive requirements were to be identified. At the same time, WP2 took care to at all times stress that early sharing of relevant concerns and information would be preferable, as a more cost-effective way of development would be to start in the right direction from the get-go, and that any parties sharing relevant concerns or information in a timely fashion might have a higher chance to find the end result optimal for their purposes.

This took place for example in:

- 2021-02-01 - 2021-04-30 All Managing Board weekly meetings.
- 2021-03-01 BIGPICTURE kickoff Day 1 WP2 flash presentation.
- 2021-03-03 BIGPICTURE kickoff Day 2 WP2 open meeting, was also used to finalize slides for WP2 session on Day 3.
- 2021-03-03 BIGPICTURE kickoff Day 2 GDPR Task Force inaugural meeting.
- 2021-03-03 BIGPICTURE kickoff Day 2 Honest Data Broker Task Force inaugural meeting.
- 2021-03-05 BIGPICTURE kickoff Day 3 WP2 session on plans and discussion critical paths.
- 2021-03-19 Email sendout to the EFPIA partners, requesting infrastructure requirements in all areas, and especially in regards to IT security.

3. Invitation to joint evaluation of security requirements

There is good reason to believe that the BIGPICTURE infrastructure will be a good fit for European digital pathology research, at research institutions and biobanks, given that its development will be based on services that are currently in production use in European human health data research in these types of organization. However, as these services have hitherto not been extensively used by large European pharmaceutical companies, such as the BIGPICTURE EFPIA partners, ascertaining whether new requirements could arise from such use is nontrivial. Also previous installations of the services were more homogeneous in terms of data contributors, data types, and legal basis than the BIGPICTURE project. This may add another layer of complexity to the requirements for the technical infrastructure.

The BIGPICTURE EFPIA partners operate according to in-depth guidelines for information security, which are developed by internal functions for information security in these organizations, according to similar but not identical parameters. These guidelines are typically made holistically, for production-

ready systems. Assessments would start with system components overview before assessing the security of the individual components, the processes of security patching and others. This is one other good reason for instating the BIGPICTURE Architecture Task Force.

This poses a potential problem for BIGPICTURE, for several reasons:

1. The BIGPICTURE platform does not yet exist, and can thus not be evaluated according to standard criteria. A Pilot system will be in place M10, and the production system will be in place M25, however delaying information security evaluation to such dates would introduce significant delay (and significant project risk) to the point of likely being detrimental to project feasibility.
2. Carrying out multiple independent holistic information security evaluations will be very resource intensive and time consuming, potentially adding further delay and project risk to feasibility.
3. Typically assessment is performed with the system vendor, which will be of importance going into the sustainability phase post project.

For these reasons, WP2 invited the BIGPICTURE EFPIA partners to engage in a joint security evaluation (cf Appendix 2), tailored to this project, with the aim to expedite the information security evaluation, thereby reducing risk and delay. The invitation also stressed the need for expediency, and invited connections from Information Security & Risk management departments to provide input on suitable assessment processes for providing clearance to share data on the platform.

4. Instating the BIGPICTURE Architecture Task Force

Early in project execution, WP2 identified that BIGPICTURE could benefit from developing an easily communicable unified view of project components and relations, in order to facilitate decision making and reaching unanimity. One approach to produce such a view is called Enterprise Architecture, which we implement in BIGPICTURE. This approach is usually most effective when carried out in broad and cross-functional engagement with all stakeholders. To this end, WP2 instated the BIGPICTURE Architecture Task Force, headed up by Ari Rouvari, Lead Enterprise Architect at WP2 Partner CSC. One tangible output of Enterprise Architecture work is a set of diagrams of well-defined interconnected components that combine into a functional whole that well describes the activities at every level. Examples are shown in Fig 1 (Overview of Enterprise Architecture Layers) and Fig 2 (example of how a "Data Use" Business process can be broken down into individual application components). The process of producing these diagrams is often productive in highlighting missing connections or knowledge gaps, which could otherwise slow progress or hinder realization of the full potential of the effort.

Enterprise Architecture describes at a targeted abstraction level the services, processes, application and technology layers. Also modelling of data is included. The process of making the architecture diagram is very important since it engages the users and providers of the services working together and thus clarifies various issues by communication. A visual image helps in understanding.

All BIGPICTURE Partners in all capacities are welcome to join the Architecture Task Force, and to indicate what Architecture Layer they envision themselves contributing to. The Architecture Task Force works iteratively to refine the architecture diagrams, in bi-weekly topical workshops with assigned homework between sessions, to address identified needs. The topics for the upcoming workshop are announced to the preceding Managing Board meeting. The topics that are addressed by the Architecture Task Force extend well beyond just infrastructure requirements, and will remain operational beyond Task 2.1.

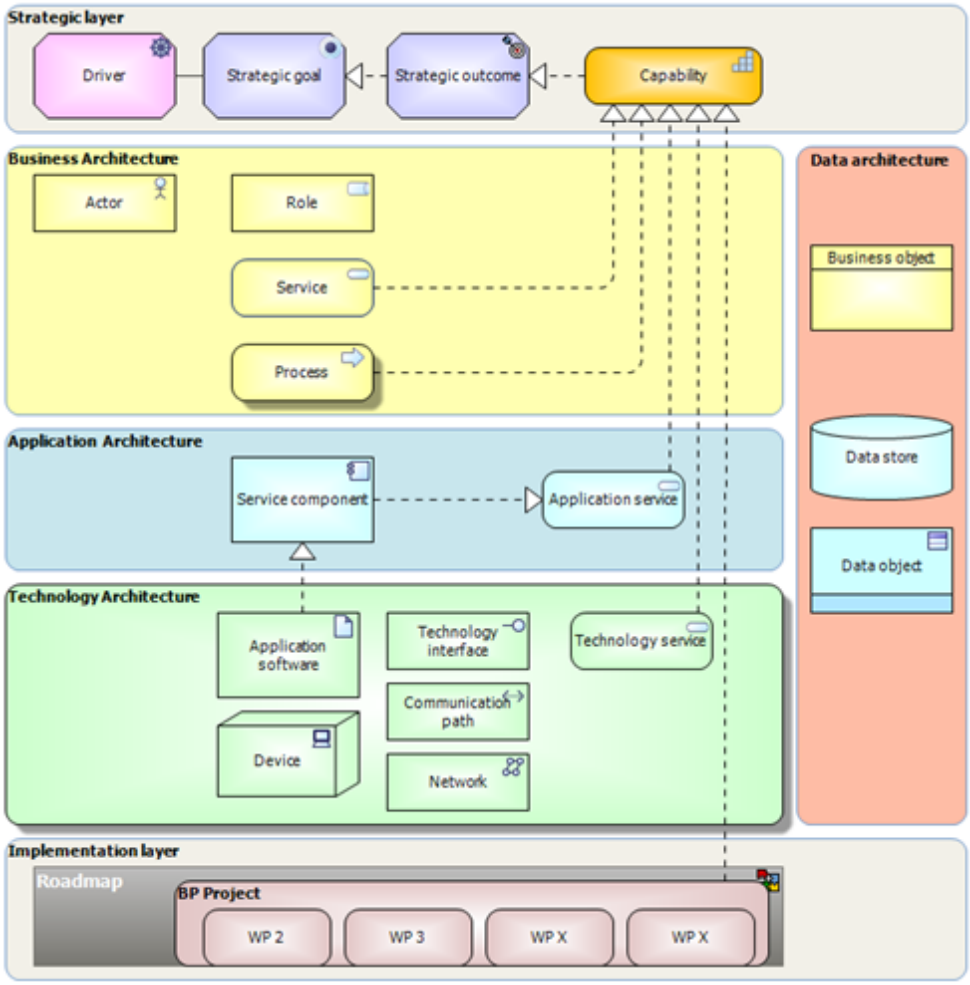

Figure 1. An architecture metamodel describing the different layers.

In Figure 1 and 2 metamodels the services are presented in the Business architecture layer in the highest level such as “BIGPICTURE Data Repository Service” and in the Application layer split into the service components like “Data Submission Service” consisting of the functions (Data Use in Figure 2). The processes (e.g. Authentication in Figure 2) are described in an appropriate level that there is a common agreement and understanding, how the various workflows go. Technology layer defines the high-level architecture of the system indicating how the applications are deployed.

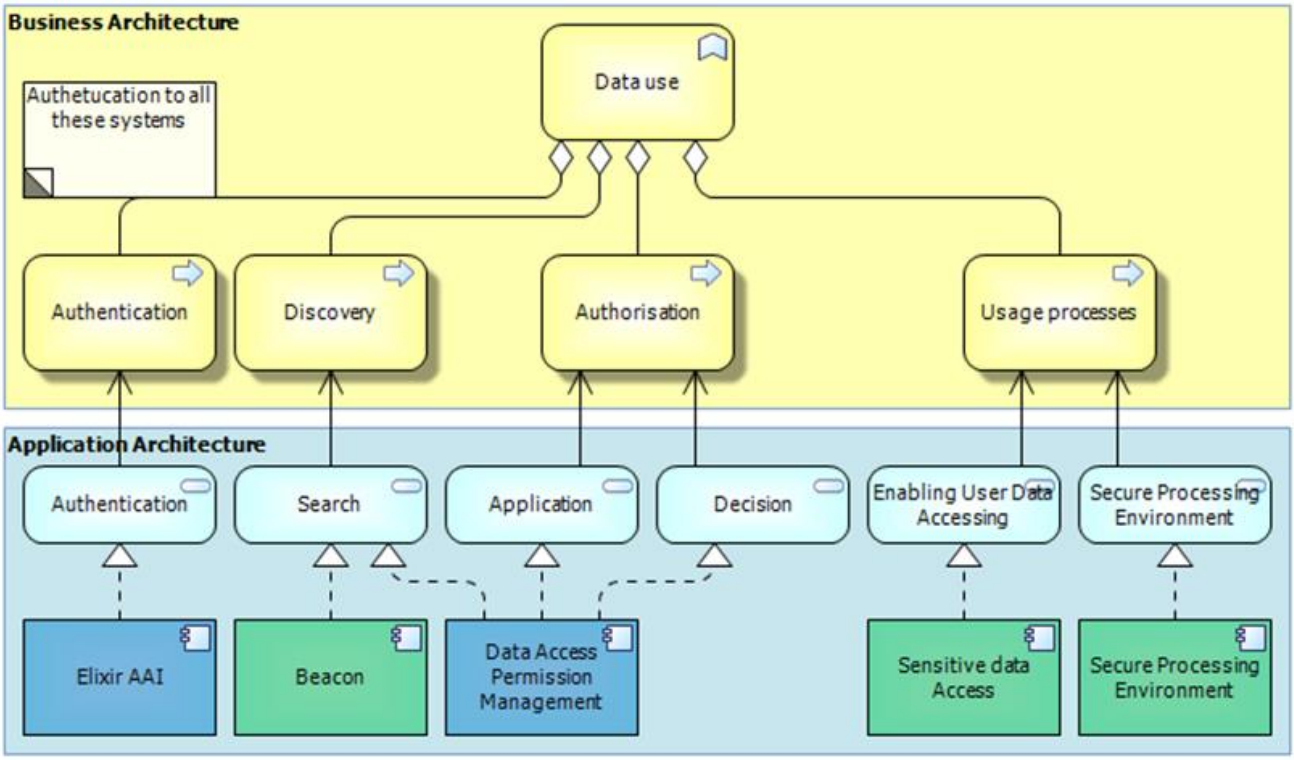

Figure 2: A draft image just to illustrate the business service "data use" based on various processes, services (the second last row) and applications (the bottom layer).

5. Merging identified requirements

There is a very real possibility that any broad requirement collection activity may yield potentially contradictory requirements, which nonetheless must still be merged into something unified and implementable. The WP2 approach to merging requirements was to seek unanimity, through compromise and clarification, with the goal to find solutions that can be feasibly addressed within the ethical and legal frameworks (elaborated within WP5) and available resources.

In this Task, however, no merging was necessary, since no conflicting requirements were identified. WP2 will however remain attentive to continually identified requirements in interaction with project stakeholders throughout the iterative development of the infrastructure (cf descriptions for Task 2.3 Pilot system development and Task 2.6 Iterative development). Should any conflicting requirements be identified in the course of iterative development, the above approach will be taken.

Results

Initial EFPIA information security and technical security requirements

The primary finding in the responses to the infrastructure requirements gathering activity in Task 2.1 is that no immediate concerns or obstacles have been expressed or identified that would hinder the immediate start of Task 2.3 - Pilot system development. Indeed, feedback has rather indicated that proceeding with implementation of the pilot system as soon as possible would be preferable, for several reasons:

1. A prototype system would provide experts in Partner organizations with something concrete to comment on, which is closer to the preferred way of working for example at EFPIA information security functions.

2. This would also enable WP2 developers to iteratively address dynamically identified requirements, which is closer to the preferred way of working in WP2 Agile development teams.

Parameters for joint information security evaluation by EFPIA Partners

We understand from the responses that the EFPIA evaluation will likely include the overall platform, and be much broader than just the infrastructure. The BIGPICTURE EFPIA Partners have begun the work to define common requirements for the information security evaluation, with an initial focus on IT security. Initial input has been provided by Janssen, Bayer and UCB.

The most critical interfaces for security testing are the user facing ones (essentially: data in/out, and application level). Special considerations exist, such as only designated encryption algorithms are approved for use. BIGPICTURE also needs to take into account that these allowlists are subject to constant change, as technology progresses and as new vulnerabilities are discovered.

One piece of feedback to the call for joint information security evaluation was that one of the necessary inputs will likely be an architecture diagram, which is an expected output from the Architecture Task Force.

A question also arose whether there is any standard for software testing / penetration testing for this platform, however to the best of our knowledge, no such standard exists. Some similar standards exist and are gaining in utility, such as ISO27k certification of IT Security Management Systems of organisational units operating production services, however such certification can only be obtained for existing production services in an organisation with a focused scope. The list of services that will be relevant for certification is an expected output from the Architecture Task Force.

In response to the WP2 call for infrastructure requirements, the IT security team at BIGPICTURE EFPIA Partner UCB contributed a list of aspects that are generally taken into account in their information security evaluations. UCB has also indicated willingness to engage in a joint sec evaluation.

BIGPICTURE EFPIA Partner Janssen will be required to perform an application assessment on the platform that will be developed and used. This is to evaluate the developed solution against the internally mandated minimum viable security controls. During the WP2 kickoff Janssen also provided their perspective and shared some details with the WP2 members, they also provided a Supplier Information Security Requirements document for use within the BIGPICTURE project.

BIGPICTURE EFPIA Partner Pfizer have responded that they are looking into whether they will be able to participate in joint information security evaluation, and have also contributed a security questionnaire that is routinely used in Pfizer service validation, and have requested architecture diagrams be made available as a starting point. An initial review of security questionnaires indicates that there may indeed be large overlaps between (at least) Janssen and Pfizer security questions, providing credence to the belief that a potential future joint information security evaluation may be an effective strategy.

BIGPICTURE EFPIA Partners Novo Nordisk and Roche have provided contact information for future joint information security evaluation work.

Authentication and Authentication Infrastructure (AAI) requirement gathering

Many of the initially contributed considerations for requirement gathering addressed whether existing solutions for user identification for example in EFPIA organizations would be interoperable with the mechanisms intended for use in BIGPICTURE (either ELIXIR AAI or the upcoming Life-Science AAI, cf Description of Action), and whether security policies would align.

At the time of writing, it is expected that these either will be interoperable and compliant, or can be made interoperable and compliant within the course of the project. This will however be followed up as part of iterative development of the pilot system in Task 2.3. Continued security requirement dialogue will involve user identification as part of the registration and user management process, as well as authentication methods including 2-factor authentication for subsystems where such will be deemed necessary.
The authorisation process is built in close collaboration with the Honest Broker implementation.

Initial specifications for backend data taking and use, metadata format and private inference execution model

As described in the BIGPICTURE Description of Action, WP2 will implement a pilot system in Task 2.3 - Pilot system development, based on the production services ELIXIR AAI, REMS, (Federated) EGA, and Beacon. These services provide de-facto initial specifications for backend data taking and use, and metadata format. The context that they are currently operating in provides de-facto initial specifications for a private inference execution model.

The feedback obtained in the infrastructure requirement gathering activity has strongly indicated that there is no immediate need to redesign these specifications before proceeding to implement the pilot system, and that there rather is a strong preference to get a prototype system set up as soon as possible, to act as a concrete example and a starting point for development of more refined specifications.

According to the BIGPICTURE Description of Action section on Task 2.1 - Infrastructure requirements gathering, this development of the final requirements will be carried out iteratively and through stakeholder dialogue in Task 2.6 - Iterative development. The work collecting and refining specifications will be carried out in BIGPICTURE Task Forces (e.g. Architecture, GDPR, Honest Broker, Metadata and nomenclature) and other cross-work-package initiatives.

As a background material there exist a set of the deliverables regarding the Federated EGA model produced by ELIXIR EXCELERATE Program, WP9[2] . For example ELIXIR EXCELERATE D9.2 Report on implementation of submissions interface and API.

General information on initial services

Current documentation of the Nordic implementation of the Federated EGA model is here. In broad terms, it is a microservice architecture that is using an AMQP broker for communication between services. It's designed so that decryption keys are never unlocked on services that are reachable from the Internet. All files are encrypted with Crypt4GH, a GA4GH encryption standard utilising public key cryptography. The difference from PGP encryption is the ability to share only portions of a file without re-encryption, by using the index file of the genome file.

There is a demonstrator video[3] that shows how EGA, ELIXIR AAI and REMS work.

[2] EXCELERATE WP9: Secure archiving, dissemination and analysis of human access-controlled data https://elixir-europe.org/about-us/how-funded/eu-projects/excelerate/wp9
[3] Demo on transferring data access permissions from REMS to EGA. CINECA AMG meeting, March, 2021. https://www.youtube.com/watch?v=FpC6nncjJ1s

Initial specifications for backend data taking and use

Federated EGA interface

The Federated EGA, which we are planning to build the WP2 services on top of, exposes services for data in and data out. All data submitted to the archive has to be encrypted with the correct public key and all data exported from the archive is encrypted with the public key of the requestor. The design is modular so that the outward facing inbox and outbox implementations can be easily replaced while the internal services for QA, encrypt, decryption and authorization can stay the same.

REMS interface

(Not relevant for data taking right now.)

ELIXIR AAI interface

Authentication for the S3 proxy is done with GA4GH Passports, a JWT standard. To get a visa there's a login portal that implements OpenID Connect that uses the ELIXIR AAI service to authenticate the submitter and supplies the JWT to the user to be able to upload.

Based on Authentication and Authorization Infrastructure (AAI) services, the GA4GH Passport standard defines a machine-readable digital identity that conveys data access rights (granted by a Data Access Committee or other authority) and increases data user interoperability. These permissions, or 'visas' are used by technical infrastructure services to manage data access permissions for researchers by maintaining a record of the signing authority granting the permissions. The visa provides the human and technical governance structures with a durable high-grade assurance that the person(s) accessing the data has been authorized to do so.

Initial specifications for metadata format

The BIGPICTURE repository infrastructure will be developed by WP2, based on services that are currently in production use in other communities. This section summarises the de-facto baseline definitions of the metadata formats used by these services, with links to further reading. The intention is to provide a baseline understanding that may be useful for further development work to better suit the needs of BIGPICTURE. However, WP2 will not be the primary driver of BIGPICTURE metadata format development, as this is primarily worked on by WP3, WP4 and the Metadata Task Force, with the Architecture Task Force also addressing data modelling. Instead, WP2 intends to actively follow these developments, and strive to be as forthcoming as possible in supporting the needs arising from such development.

Many technical infrastructure services are dependent on the outcome of the metadata task force. The key services that are affected are data submission, discovery (e.g. Beacon, data catalogue), access (e.g. Honest Broker mechanism) services.

REMS metadata format

(Not relevant right now)

Federated EGA metadata format

Datasets in EGA are described with the ENA metadata model. XML Schemas for the metadata are published on github. To summarize, there are metadata objects for:

- Study - describes a sequencing project
- Sample - describes the source material
- Experiment - information about how sequencing was done
- Run - contains the mapping between files and experiment
- Analysis - secondary files generated from the runs, for example a genome assembly

The ENA metadata model is highly tuned to genome sequencing, but is nonetheless a valuable reference and starting point for developing Federated EGA compatible metadata formats suitable for digital pathology.

Beacon metadata format

The ELIXIR Beacon network portal[4] has links to Beacon related pages. The data model for genomics is illustrated on GA4GH Schemablock page[5]. Since the domain is different, this project has to define strategies, how to implement the discovery.

Initial specifications for private inference execution model

The purpose of BIGPICTURE private inference execution is to enable morphological data-mining, algorithm validation, where arbitrary algorithms can be used to make inference against arbitrary data, in a way such that no sensitive details of neither data nor algorithms become accessible to either rights-holder or third parties.

Such inference can be used to answer questions such as:

- "How well does this algorithm perform against this given reference dataset?" which is a useful question to ask both for algorithm developers looking to improve their product, and for algorithm users evaluating whether it is worth their while to procure a given product.
- "Which of the repository datasets contain a significant amount of images relevant to my research interests, as identified as positive matches to this algorithm?" which is a useful question to ask for researchers looking to discover datasets relevant for their research, and for evaluating which could be most worth their while to formally apply for direct access to.

This section describes a minimal model for private inference execution, designed to be extendable and adaptable to fit, as the BIGPICTURE metadata model becomes more well-defined. The intention for this initial specification is not to commit early to any design choice (a.k.a. "big design upfront"), but to provide a starting point for discussion and iterative refinement.

The role of WP2 in this activity is to provide GPU capacity and a technology platform to safely execute inference using models and data that are both provided outside of WP2. WP2 will implement the data transfer interfaces, where the related data and metadata as well as the algorithms will flow. The Architecture Task Force will define the functions and processes in collaboration with the other WPs.

The BIGPICTURE initial specification for private inference execution is:

- The general approach will be secure sandboxing, with defined and controlled inputs and outputs, which will be parseable and validatable to ensure that no confidential detail is disclosed to unauthorized parties.
- The algorithm is supplied as a container image (such as a Docker or Singularity image).
- The container contributor is responsible for the correct operation of the container.
- On execution, the container will not have network access.
- On execution, the container is supplied with
    - Readable directory with input data.
    - Writable directory for output data.
- If the algorithm produces output compatible with the BIGPICTURE Beacon metadata format (TBD) then it will be possible to use the algorithm to make datasets discoverable through the BIGPICTURE Beacon, and to use the algorithm for morphological data-mining.

[4] https://beacon-network.elixir-europe.org/

[5] https://schemablocks.org/standards/ga4gh-data-model.html

- Inference can be used for validation or annotation.
    - In validation, the task is for the algorithm to produce classifications, which are then compared to known correct answers, to assess algorithm performance. Datasets intended for use in validation will not provide access to the known correct answers to the container.

Discussion

This report does not include the BBMRI Directory, since the focus of this report is on requirements on the infrastructure that is being developed by WP2. The BBMRI Directory will not be further developed in WP2. However BIGPICTURE plans to use the BBMRI Directory service as-is, by exporting compatible metrics for advertising available datasets, to the extent possible, as described for example in Task 2.3.

Conclusion

By performing the steps described above, we have been able to establish the initial requirements on the repository infrastructure that will be developed by WP2. These requirements will be iteratively improved during the course of the project, in Task 2.3 Pilot system development and Task 2.6 Iterative development.

Our analysis at this point is that there are no immediate concerns regarding security requirements or other requirements.

The following next steps and definition of WP2 will contribute to further refinement of the requirements:

WP2 should move to pilot implementation and iterative refinement as soon as possible, as this mode of work is preferable to both information security personnel and the WP2 Agile development teams.

WP2 should remain open to feedback, and proactively seek input on potential new identified requirements, by engaging in Task Forces (eg Architecture, GDPR, Honest Broker, Metadata and nomenclature.) and other cross-work-package initiatives.

WP2 should use the Architecture Task Force to carry out iterative refinement of architecture, design, and implementation choices as we go.

WP2 should continue to stress the importance of the Architecture Task Force, and make its outputs available to all as soon as possible.

WP2 will make sure the technical security measures taken to protect the data privacy of any dataset regulated by the GDPR can be clearly audited, and aims for a security specification suitable for processing sensitive data.

WP2 will not make decisions regarding data quality and (semantic) metadata requirements.

WP2 will not decide on Data access authorisation for a user, but WP2 will technically implement the decisions made by the Honest data broker process.

Appendix 1
Expectation management

This document was originally drawn up for clarification and fact-checking of commonly occurring misconceptions that could lead to false expectations being placed on WP2, which could jeopardize feasibility/sustainability of the project. This document was updated with findings in the BIGPICTURE WP2 kickoff, and further refined at the BIGPICTURE kickoff Day 2 WP2 open meeting, and presented in this form at the BIGPICTURE kickoff Day 3 WP2 session on plans and discussion critical paths.

An up-to-date version of this document is maintained by WP2 and kept accessible to the rest of the project.

The submitter must handle access requests

Data rights will not be transferred to the repository infrastructure built by WP2, but will remain with the data submitters. WP2 will not handle access requests on behalf of data submitters; data submitters must do this themselves. WP2 will however provide services to facilitate these processes, and WP(1/3/5?) will develop standard licenses that can (must?) be used. Also, the ambition is that WP3 Honest Data Broker will be instated as an instance to which such handling can be delegated.

No AI training on WP2 budget

WP2 will offer compute capacities for inference. WP2 will not offer compute capacities for AI training. This has been clarified multiple times in the past. This is by design. AI training requires orders of magnitude greater capacities than does inference. WP2 will build repository infrastructure in scalable compute centers that also offer commercial services, procurable by parties interested in having a high performance storage/compute system set up close to the repository.

No Federated Nodes operated on WP2 budget

WP2 will only fund operation of the central repository. WP2 will demonstrate ease of deployment of Federated Nodes, but will not fund their operation. For project feasibility, BIGPICTURE is only able to fund the central repository (economy of scale). For project sustainability BIGPICTURE must demonstrate that it is easy for external parties to deploy their own interoperable Federated Nodes *at own cost* to enable them to share their own data, while keeping the sensitive information inside the original jurisdiction.

No training allowed on non-directly-accessible data

WP2 will not support AI training on data that is not directly accessible to the user. Assessing the quality of an AI model (taking performance and consequences of failure into account) is strongly

dependent on the ability to inspect outliers and surprising failure modes. In the field of pathology, failure to correctly assess quality could lead to **death, harm and liability**.

WP2 will still strive to develop methods to enable federated learning, but only with directly-accessible data. People must always have the ability to oversee the training, and inspect any (potential surprising catastrophic) failure modes.

WP2 will support inference on indirectly accessible data (for example to do validation against clinical trial data that is approved for validation but not other types of sharing), but will not support training on indirectly accessible data (or risk causing death, harm and liability).

WP2 will develop secure APIs, not polished end-user products

WP2 will provide robust secure APIs, suitable for use by technological expert organizations. Data contributors and users will interact with the BIGPICTURE platform mainly through user interfaces developed by WP4 (such as Cytomine). WP2 aims to have a functional Cytomine instance running on top of the WP2 infrastructure at all times, as an integration target, so the cytomine APIs will also be available to end users.