

945358-BIGPICTURE
Central Repository for Digital Pathology

WP3 - Data collection & management

D3.01 Report on node network organisation model

Lead contributor	21 - MUW
	renate.kain@meduniwien.ac.at
Other contributors	1 - RUMC 3 – LTHTNHS 4 – UMC-Utrecht, 20 – RÖ 21 – MUW 23 – HUS 24 – UMR 31 – LYGATURE 33 – DECIPHEX 37 – JANSSEN 39 – BII GMBH

Document History

Version	Date	Description
V2.0	30 Jul 2021	Final Draft

Abstract

The first objective of work package 3 (WP3) is to establish and employ a sustainable Node Network Organisation Model (NNOM) responsible for mediating and collecting datasets from beneficiaries and third parties (Slide Contributing Third Parties, SCTP) and to contribute them to the BIGPICTURE data repository.

This document describes deliverable 3.01 in which we will outline the Methods to establish the NNOM and how the progress of this task to date. Next, we describe how the business operates and the consumer's role therein. A reference to the Tasks of WP3 in the Description of Action (DoA) will be given for all items. For convenience, the description of Task 3.1 of WP3 in the DoA is added as appendix 1 for reference.

The Progress to Date is described in four sections:

1 Node Network Organisation Model (NNOM)

Within the BIGPICTURE consortium there are 7 established collector nodes.

The Collector Node Network Organisation is based on the three different types of contribution partners

- EFPIA partners with slides from non-clinical studies in where the tissue comes from animals
- Academic and health organisation partners within the BIGPICTURE consortium submitting slides from human tissues
- Academic, health organisation and AI company potential SCTP partners submitting slides from human tissues

Day to day operation is managed by the Collector Node Coordinators with the task to establishing the data collections that form BIGPICTURE repository and to ensure communication between all partners involved.

2 Business operations

The prospective data collections of Whole Slide Images and mandatory associated metadata will be defined by the Collector Node Network of WP3 and facilitated by to-be-developed semi-automated extraction tools / application programming interfaces (APIs) to link directly to picture archiving and managing systems and pseudonymisation tools to eradicate/remove personal and sensitive information by Sectra and DECI, in concordance with the pseudonymization strategy outlined by WP5.

3 Communication

The efficient collaboration of seventeen beneficiaries and over 20 Slide Contributing Third Parties (SCTP) collectively supplying data to the tasks in WP3 requires an effective communication strategy. Appropriate communication channels to meet the goals of the project have been designed and established.

4 Relation to Legal requirements and Task forces - Prerequisite for Slide Collection

Prerequisite for the collection and transfer of data demands compliance with legal and ethical requirements at local national and international level. To meet this goal, a Task force has been established and, prepared by WP5, guidance documents are prepared.

Methods

Rationale

To ensure smooth operation of data collection, handling, transfer, sharing and use, the central operational body of WP3 is the Node Network. Its architecture and organisation set up by WP3 members (beneficiaries) informed by Task Forces (GDPR, Metadata & Nomenclature, Architecture, Honest Broker) working on the operational foundation of the BIGPICTURE repository is designed to establish and ensure its functionality. Core to the tasks related to its set up are the efficient management of data collection and communication with the partners in BIGPICTURE.

With 45 Beneficiaries and over 20 Slide Contributing Third Parties (SCTP) contributing to the slide repository, it is necessary to manage the collection of Whole Slide Images (WSI) and associated data in such a way that it results in a collection of three million slides that is of value for its purpose. Building a collection takes acquiring, curating and culling. To this end, a **Collector Node Network**^{*} was devised. Each node is led by an expert in their respective field, designated **Collector Node Coordinator**. The coordination of acquiring, curating and culling across nodes takes place in the **Collector Node Network Board**, consisting of all Collector Node Coordinators and the **Collector Node Manager** who chairs the **Collector Node Network Board**.

Thus, the Node Network, consisting of seven nodes has been devised, depicted in the figure below:

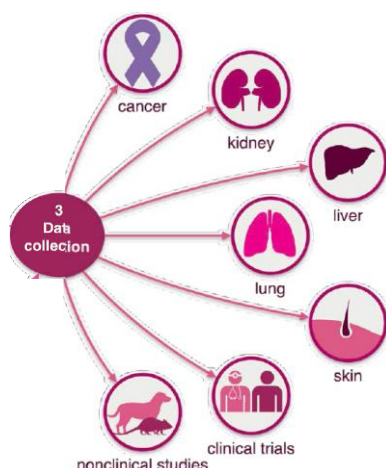


Figure 1: Node network

Note that these nodes represent data collection points and do not refer to or reflect individual WSI collections.

The node network organisation model (NNOM)

The preparation of the **Node Network Organisation Model (NNOM)** was initiated by the **establishment of the network and creating an overview of the parties (Beneficiaries and Slide contributing Third parties, SCTP)**. This was done by administrative work, performing an outline of the different sites and organisations involved in the operations of WP3, and collecting and assorting contacts of individuals from these organisations responsible for data contribution to the individual Nodes. The information was initially compiled from information in the grant agreement (GA), the consortium agreement (CA) and the communication material and Excel files describing datasets and collections from leads of WP3 involved in the BIGPICTURE project since the start. Further development of the NNOM was advanced during regular meetings of WP3 management, node coordination meetings, and direct contact with slide contributors. Management of the Node Network and its advancement is a continuous progress, monitored by WP3.

As part of establishing the NNOM we performed an evaluation and expectation management during the kick-off meeting on Feb 1-5 2021 and a meeting with SCTP candidates held on May 26 2021. During those meetings the lead of WP3 had the opportunity to identify and communicate expectations placed on WP3 in relation to other WPs and those set out by SCTP and to communicate project aims and solidity to maintain interest of the SCTP for the continued collaboration.

The information provided was based on that available from the grant and consortium agreements describing the Node network as well as discussions within WP3, the roles and functions of both organisations and individuals representing them within the Node Network. The information gained, in turn, fed into the development of the NNOM and the tasks associated with it as described below.

Results – Progress to date

3 Node Network Organisation model (NNOM)

The Node structure is organized as individual teams of Beneficiaries and SCTP in a Node led by a **Collector Node coordinator** who coordinates the tasks of the Node. The Node Coordinators communicate with the slide contributors (Beneficiaries and SCTP) in their node and to the lead team of WP3 for which one dedicated person also assumes the role of **Collector Node Manager** who coordinates actions and ensures communication across the nodes. The interaction and collaboration of individual nodes is thus defined as the **Collector Node Network** (see figure 2) and the node coordinators and the node manager as **Collector Node Network Board**.

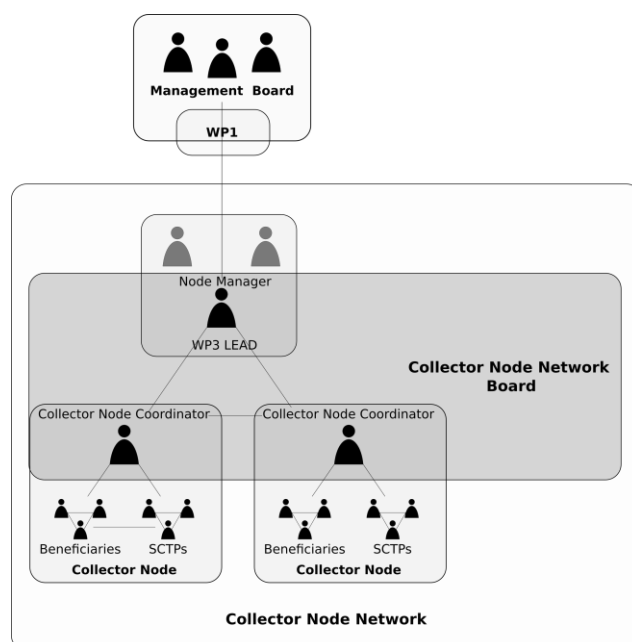


Fig. 2: Collector Node Network organisation

1.1. Node Architecture with reference to Data Collection

According to Task 3.1 (M1-M6) in the Description of Action (DoA) each collector node will be led by a coordinating manager (**Collector Node Coordinator**). The proposed collection of over one million histopathological cases in the six clinical nodes covers a wide range of neoplastic and non-neoplastic human diseases of all organ systems, associated with well documented clinico-pathological and/or molecular data, prognostic data, and information on response to therapy. The repository of Whole Slide

Images (WSI) in the non-clinical node will contain over 2 million slides from non-clinical studies (e.g., systemic toxicity studies evaluating the generalized biological effects to organs and tissues following exposure to xenobiotics) with associated metadata covering different laboratory animal species and a wide distribution of tissue types. Importantly, conceptual models of pathological metadata are being established (WP 2, WP3, WP 4, and TF Metadata/Nomenclature (TF MN)).

The Cancer (1) node coordinates the collection of neoplastic disease with WSI and tissue microarrays including precursor lesions, and common and rare types of cancer. The Renal node (2) gathers renal biopsy series on glomerulopathies, vascular diseases, and rare non-neoplastic diseases of all other organs. The Clinical trials node (3) will collect clinical trial cases including the relevant metadata. The Liver node (4) collects hepatobiliary neoplasia, native liver biopsies and transplant liver biopsies. The Lung node (5) collects neoplastic and non-neoplastic diseases of the lung and respiratory tract. The Skin node (6) manages skin cancer and inflammatory skin diseases. For the Non-clinical node (7), DECI, having expertise with management of non-clinical images and data, will facilitate the collection of all non-clinical slides and metadata. DECI will collect non-clinical cases including relevant metadata from systemic toxicity studies and submit non-clinical images to the main repository.

1.2. Structure of Node Network with reference to Data Collection

The collector node coordinators of each of the seven collector nodes depicted in figure 3 will propose type and numbers of datasets and associated data to be collected within the node. These proposals will be communicated to the beneficiaries and SCTPs associated within the respective node, to the coordinators of the other nodes, to the collector node manager and to the lead team of WP3 using dedicated communication channels (regular meetings, email, communication platform for sharing and working on documents). While decision on type and number of data to be collected in each node is decided primarily at node level, final decisions on the content of the collections in each node are made by the Collector Node Network Board after discussion to ensure balance of data.



Figure 3: Node coordinators

1.3. Role of Collector Node Coordinators

The role of the node coordinators (coordinating managers) related to the task of establishing the node structure is the preparation of node collection lists, i.e., datasets (diseases/entities and associated metadata) to be collected within the nodes and distribution, for instance information like documents, to slide contributors. Regular meetings of the node coordinators at intervals demanded by the work progress ensure smooth flow of communication within WP3 and all node coordinators.

As is foreseen in the DoA, the node coordinators, in collaboration with WP1 and WP6, will develop value propositions tailored to the needs of the different data contributors.

The node coordinators, based on the guidelines and tools for quality control and quality assurance, evaluates the contributed datasets and determines a prognosis on the upcoming annual needs of BIGPICTURE data sets. Each node will promote the control process for the partners in the node, to contribute to the prospected data sets with appropriate metadata, implemented automated quality control on images and data (developed by WP3 and WP4), adapted to the BIGPICTURE license for data use while the Honest Broker will ensure the safe transit of anonymised or pseudonymised data.

The duty of the **collector node coordinator** is thus to

- a. Identify potential slides for the repository
- b. Approach and maintain contact with beneficiaries and SCTP within the node
- c. Act as first contact point for beneficiaries and SCTP within the node for questions (i.e. Scanning facilities, ethical questions, data sharing)
- d. Facilitate the processes for upload of images and metadata by establishment of metadata fields in collaboration with other related WP (WP2, WP4)
- e. Monitor compliance with workflow of data collections based on statistics from the repository and WP4
- f. Monitor library contents facilitated by and in cooperation with WP2
- g. Approach potential SCTPs and propose prospective inclusion to the lead of WP1 and WP3

The role of the **collector node manager** is to

- a. Coordinate regular meetings with the node coordinators to ensure communication on the update of progress and discuss issues arising
- b. Gather information from each node in relation to compliance of data collections
- c. Report to and coordinate communication to the WP3 lead
- d. Act as first contact point for WP1 within the managing board meetings relating node responsibilities
- e. Initiate meetings between data access committee and node coordinators

According to task 3.5 (M1-M6) the node coordinators, in collaboration with WP1 and WP6, will develop value propositions tailored to the needs of the different data contributors, both beneficiaries and SCTPs. Currently the envisioned value propositions are:

- a. hosting digital slide collection in a curated online library,
- b. access to tools created by the project, including tools to view, annotate slides and AI tools,
- c. the ability to augment own collections or work with data from other collaborators,
- d. priority access to the data collected in the rest of the project,
- e. authorship in publications using the provided data.

Each collector node is equal in terms of decision making and the collector node coordinators will have an equal vote in the collector node coordinator meetings held at regular intervals, during which tasks and progress are discussed. Notes of the meetings are securely stored and available to the beneficiaries

of the consortium. Results and progress of the node meetings are communicated to the WP3 lead team during dedicated meetings every other week. In addition, all node coordinators and the lead team of WP3 meet on a regular basis to review the progression of data contribution based on QC reports and DAC with relevant persons from WP2 and WP4. These meetings will be led by the collector node manager in collaboration with the collector node coordinators. Decision on next 12 months data collection and AI development will be defined and presented to the managing board for acceptance.

2. Business Operations

The prospective data collections will be facilitated by to-be-developed semi-automated extraction tools / application programming interfaces (APIs) to link directly to picture archiving and managing systems and pseudonymisation tools to eradicate/remove personal and sensitive information by Sectra and DECI, in concordance with the pseudonymization strategy outlined by WP5.

The node coordinators will provide advice on the process of uploading the data to the repository including Quality control (QC) and Quality assurance (QA) measures. A dedicated quality control team (Task force QA/QC) established to carry out the work described in Task 3.5 (II: Data collection related tasks) of WP3 in the DoA will provide guidance on measures to ensure uniform data handling in each of the nodes and employing DICOM (or other appropriate) tools from WP4, in close collaboration with the generic QC strategy developed in WP5. Facilitating uploading mechanisms and protocols (such as templates etc.) for large datasets developed by (WP2) that minimise the overload of hospital and university network will be developed.

Annually, the metrics from contributed dataset will be assessed by the quality control committee. They will generate a report and provide metrics data. The node coordinators' board will review these documents and provide a prognosis for the upcoming year and suggests under-represented data set trends, giving contributors the opportunity to strengthen the diversity of the node.

Several of the beneficiaries within the nodes have scanning facilities at their disposal and it is envisioned that these will be made available to slide contributors if required. The necessity for this service is dependent on priority of data collection within the node and the collector node coordinators will advise on who to approach if slide scanning is required. Terms and conditions for the use of such service, including for instance data transfer agreements, will be prepared by all node coordinators in the framework of WP3, close guidance from WP5 and in coordination with WP1. In close collaboration with Task Force Honest Broker, Task Force Metadata/Nomenclature, and WP5, we will provide guidance on the submission of information required to ensure a fair, effective and trusted mechanism for sharing and analyzing slide collections.

2.2. Datasets within the BIGPICTURE repository

According to Task 3.2 of WP3 described in the DoA (Appendix 1), prioritisation of datasets, (M1-M12) the BIGPICTURE consortium has already identified a number of potential data contributors. The consortium, represented by the node coordinators will continue to work with these data contributors to gain alignment of interests and establish a data access committee (DAC).

As outlined in D1.04 (Initial Data Management Plan) the process of handling and storing over 3 million WSI with mandatory metadata will be governed by the BIGPICTURE Data Management Plan (DMP) to guide the coherent management of data that will be collected, shared and generated.

2.2.1. Whole slide images (WSI)

As outlined above, each node proposes a defined, categorized set of data/disease(s) to be collected within the node and provides a list that is discussed and reviewed by slide contributors in the node, all other node leads, WP3 lead, WP1 and WP6. The disease entities are subject to change during the

course of the project depending on the needs of the different data contributors and upon advice of WP6 and the Scientific Advisory Board.

2.2.2. Datasets proposed to be collected

All nodes have prepared a list of envisioned datasets, which will be briefly mentioned below. However, as this is a continuous process, these lists are not definite and will be iteratively work in progress.

Cancer node (coordinator Paul van Diest)

All organ-specific nodes (lung, liver, kidney, and skin) will collect neoplasia's of the specified organs. The cancer node will include a collection of at least the remaining ten most common solid cancer types in Europe (i.e., breast, prostate, colorectal, corpus- and cervix uteri, bladder, ovary, stomach, thyroid, and pancreas) to gain as many datasets as possible. In a later stage, the collection may also include more rare cancer types, if these datasets/cohorts are of specific scientific interest and appropriate numbers of datasets fit for the specified purpose can be identified. One could for example think of pediatric cancers. We are planning to only include malignant tumors or their precursors, however no benign lesions at this point in time.

Liver node (coordinator Darren Treanor, coordinator Clare McGenity)

The liver node will include images from native liver and transplanted liver tissue. It will include the common neoplastic liver diseases (benign and malignant) and non-neoplastic diseases including fatty liver disease, viral hepatitis, biliary disease and autoimmune liver diseases. The transplantation cases will include examples of acute and chronic allograft rejection.

Lung node (coordinator Olli Carpen)

The lung node aims to include pre-neoplastic, neoplastic and nonneoplastic entities from the lung and pleura as well as from the upper respiratory tract and from the salivary glands. These entities comprise mainly epithelial tumors both malignant and benign, a few mesenchymal tumors, as well as metastatic tumors of the lung. At this point, priority will be given to epithelial malignant tumors of the lung.

Kidney node (coordinator Renate Kain)

The kidney or renal node will collect common non-neoplastic entities/diseases from native kidneys as well as from renal transplants. WSIs of kidney transplants and native kidneys from clinical studies with well-curated patient and organ related clinical metadata (organ function, outcome and medication) available will be submitted first (a dataset for testing the pilot system is currently assembled). In addition, WSI of other tissue from the cardio-vascular system, like native and transplant heart biopsies, as well as rare diseases, such as amyloidosis.

Skin node (coordinator Anna Bodén)

The skin node envisions to include at least the common cancer types of skin cancer, melanoma, basal cell carcinoma and squamous cell carcinoma to gain as many datasets as possible. In a later stage we may include inflammatory skin disease and more rare cancer types. Clinical demand for AI tools in current diagnostic routines is one driver for development and further implementation of AI tools. Therefore, the design of some datasets will aim to include precursors as well as malignant lesions.

Clinical trials node (coordinator Carsten Denkert)

The clinical trial node will coordinate the interaction with national and international clinical trial groups, which are either partners or associated parties. The aim of this node is the integration of digital slide sets from clinical trials in breast cancer but also other types of cancer. This includes the upload of digital slides, but also the organisation of trial-specific metadata as well as data access options, which are different in clinical trial cohorts. The first cohorts to be integrated will be breast cancer cohorts from the GBG and the ABCSG, other groups include for example the TEAM trial study group. This communication network of the International TIL working group will be used for these activities. The clinical trial node is open for all clinical trial groups who would like to contribute data.

We have drafted for the purpose of the BIGPICTURE repository a broad list of entities for each topographical site represented in the clinical nodes/collections. These lists are intended to be “living” documents that will evolve with time according to the availability of the samples/metadata from the different contributors.

Non clinical node (coordinator Donal O’Shea)

The non-clinical slide collection covers the whole diversity of tissues (Adrenal gland, Aorta, Bone with bone marrow, Brain, Cecum, Colon, Duodenum, Epididymis, Oesophagus, Eye, Gallbladder, Harderian gland, Heart, Ileum, Jejunum, Kidney, Liver, Lung, Lymph node(s), Mammary gland, Ovary, Pancreas, Parathyroid gland, Peripheral nerve, Pituitary gland, Prostate gland, Salivary gland, Seminal vesicle, Skeletal muscle, Skin, Spinal cord, Spleen, Stomach, Testis, Thymus, Thyroid gland, Trachea, Urinary bladder, Uterus, Vagina, Other organs or tissues with gross lesions, Tissue masses). Moreover, multiple species will be represented in the collection, with high proportion of rodents (rats, mice) and non-rodents (dogs, non-human primates, and to a lesser extent rabbit and mini-pigs). Less than half of this collection will consist of untreated control animals.

The collector node coordinators will work together to harmonize the entities and datasets to be collected. This process will ensure that there is no overlap between nodes and also that all of the most of the relevant entities – as defined by all node coordinators in agreement with the beneficiaries and SCTP in the nodes – are represented. This will be a continuous process that will be evaluated regularly guided by the outcome from discussions with WP6 to ensure sustainability of the project.

Yet, in the (current) early phase of the project, demonstrator cases will be prioritized, as well as case series that allow testing and validation (see also D3.02 for metadata). In due course, priority will shift to data sets that fit the needs of the ever-changing field of medical science, and digital pathology in particular. For example, the focus of the nodes may be expanded to other areas of non-neoplastic disease in, e.g., gastro-intestinal, urological, gynaecological disease, to infectious disease and to therapy-related pathologies.

2.3 Metadata within the BIGPICTURE repository

In addition to the previously mentioned WSIs within a given datasets, the repository also aims on equipping those sets with standardized, structured, curated and most notably useful metadata (data about the data (i.e. WSIs) within a given dataset). This is reflected in dedicated tasks of WP3 according to the DoA: Task 3.4 - Defining and implementing datasets and metadata, as well as, Task 3.3 – Development of ontologies and standardized metadata. As decisions regarding the structure of the metadata and ontologies used, have direct implications on the node level (e.g. kind of data retrieval, exclusion or inclusion of WSIs), care was taken to incorporate all the nodes into these tasks. For this, the current NNOM as proven to be eminently helpful, as it allowed tight interaction between the Task Forces, the WP3 lead and the Node Coordinators. With the upcoming deliverable D3.02 ‘Minimal Dataset Deliverable’ in eye, the following paragraphs briefly delineate the current progress in regard to metadata and ontologies.

2.3.1 Ontologies and Coding

As coding is of utmost importance to facilitate an efficient, structured and useful data discovery within the repository, the node network in cooperation with the Task Force Metadata/Nomenclature currently works on establishing a Common Mandatory Metadata Structure that fulfills these needs.

It has been decided that on a node level, disease entities as well as other metadata relating to biological concepts, must be coded according to reporting and diagnostic standards as well as international nomenclatures (e.g. metadata relating to tumor entities, in a clinical context, shall be based on the WHO classification and ICCR standards).

However, due to the complexity and diversity of the pathological sub-speciality of each node, it became apparent that it seems not feasible to allow only one specific nomenclature/ontology or controlled

vocabulary but rather a whole set of such to be linked to a given metadata field. Following this approach will not only reflect the complex nature but also allow more flexibility for uploading parties at the same time granting a structured data retrieval.

In order to improve metadata consistency, mappings between these nomenclatures/ontologies and controlled vocabularies, if possible, will be established and algorithms to automatically translate between them will be developed in cooperation with WP4.

2.3.2 Mandatory Submission Metadata

In order to provide high quality, standardized, structured and curated metadata. It has been decided that a dataset intended for direct access can only be uploaded to the repository if a minimum set of metadata is provided, which makes a given WSI useful. This set of metadata is currently coined as the Mandatory Submission Metadata for Directly Accessible Datasets (MSMdad).

Due to the already mentioned complexity as well as diversity of the pathological sub-specialties the MSM might not suffice for the need of a given node. Hence, it was decided that the MSM can be expanded by a given node such that the WSIs are of use given its context. These expanded sets of metadata are currently termed Specific Mandatory Submission Metadata.

In line with this decision, a first version of the Mandatory Submission Metadata for Directly Accessible Datasets has been established by WP3 in cooperation with WP2, Task Force Metadata/Nomenclature, Task Force Architecture, Task Force Honest Broker and Task Force GDPR. The fields are currently reviewed and checked for feasibility by the Node Coordinators and will iteratively change until they are feasible for all nodes. After the MSM has been established each Node Coordinator has the possibility to expand the MSM to a Specific Mandatory Submission Metadata if the MSM does not suffice the needs of the given node. This process will take place internally on a node level and can be supported by the Task Metadata/Nomenclature.

3. Communication

The collector nodes currently consist of a total of 17 beneficiaries and over 20 SCTP collectively contributing data to the tasks in WP3. These tasks and the definition of pre-requisites for data collection and sharing as well as use of data are supported by beneficiaries in all WPs and Task forces. Consequently, establishing an effective communication strategy requires the establishment of appropriate communication channels to meet the goals of the project.

3.1. Communication and dialogue on meetings at WP3 level

Virtual meetings are established at regular intervals and on demand and organised by WP3 lead since 10 March 2021 with focus on dissemination of information from PMO and management board meetings as well as from related TF groups such as metadata, architecture, GDPR/Ethics and Node coordination in which members of WP3 are participating. Participants of these meeting are all WP3 lead including the Node Manager and Node leaders and their locums. Where appropriate and required, participants from other WPs or Task forces are invited at an ad hoc basis. Reports on the activities across WPs are of importance and inform on actions affecting the Node Network Organisation Model, NNOM, such as prioritization of data contribution. The meetings are led by the leader of WP3 and notes are shared within BIGPICTURE. Ad hoc communication is organised through a mailing list by e-mail.

3.2. Communication and dialogue on meetings at node level

For clinical node coordination a regular Web-based approach was taken for an iterative formation of the nodes with first focus on inclusion criteria in terms of datasets from beneficiaries within the BIGPICTURE consortium and potential SCTPs. The meetings are led by the node manager and the co-coordinating node manager. Notes from the meetings are shared within BIGPICTURE and a brief summary report is presented on WP3 meetings.

For both, WP3 and Node meetings, node coordinators and WP3 lead invite partners within the network to participate in and contribute to the needs and services of WP3 and the Node Network in a continuous

mix of expectation management, information management and agile approach to tackle tasks and needs.

3.3. Communication with SCTP

Information for SCTP is communicated at two levels:

- i) general information on the progress of the project is provided during WebEx conferences jointly organised by WP1 and WP3 (1st meeting virtual on 26 May 2021) on a regular basis every 6 to 9 months. During these meeting WP3 lead, node manager and node coordinators report on progress of the project in general terms and are available for questions. In each meeting, a specific topic will be discussed in more detail. During the next meeting, planned for Q4/2021, we will discuss the framework of a Slide contribution agreement. Contents of the material containing the information discussed will be provided to the participants by email.
- ii) Node coordinators are directly in contact with SCTP by email and telecommunication to answer specific questions with regard to the content and contribution of slide collections on an ad hoc basis.

The communication activities and channels thus serve collectively to identify and communicate expectations placed on WP3 in relation to other WPs and SCTPs and provide valuable input on the development of the BIGPICTURE repository.

4. Relation to Legal requirements and Task forces - Prerequisite for Slide Collection

4.1. GDPR Compliance, Ethical Considerations, and Data Transfer and Data Sharing

Compliance with legal and ethical requirements when working with and sharing human data is not only mandatory but at highest priority of BIGPICTURE data collection to protect the personal rights of those they are coming from and who have consented to their use. Thus, mechanisms have been implemented at several intersections with other WPs, orchestrated by them and associated task forces.

4.2. GDPR Compliance

Details of GDPR compliance are listed in the description of action (Task 5.1 - Ethical and legal compliance monitoring) prepared by WP5 with a view to Deliverable D5.1: Overview of the legal frameworks of all countries involved in BIGPICTURE, due in Month 9.

Prerequisite for the collection and transfer of data complies with local national and international standards for the use and transfer of data laid out in the *EU General Data Protection Regulation 2016/679* as well as national laws. These have been collected in the forefront by WP5, and a guidance document has been prepared. This is available to the node coordinators for communication to slide contributors.

4.3. Ethical Considerations

Ethical and legal bases for collecting and sharing WSI and associated metadata have been evaluated by WP5 and a guidance document prepared. The goal is a guidance document or a check list for slide contributors to follow the legal procedures for slide sharing. This document will be available in Month 9 (Deliverable 5.1).

The academic lead of WP3 (partner 21 – MUW) will collect documentation of compliance with legal requirements (protocols and applications to local/national ethics committees and votes) and ensures they are complete and regularly updated.

In no event will the specific mandatory (non-)clinical metadata include data associated with/and or identifying the compound (such as compound ID, compound structure), which has been administered to or is otherwise used in relation to the subject of animal to which the relevant slide relates, or any clinical pathology reports, macro reports, or study reports. Each partner may additionally decide to attach

metadata other than the mandatory submission metadata to its whole slide images WSI's for use within the BIGPICTURE platform.

4.4. Data Handling, Transfer and Sharing

Conforming with European legislation the consortium members will confirm that they have appointed a data protection officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For the members of the consortium not required to appoint a DPO under the general data protection regulation, a detailed data protection policy for the project must be elaborated. The corresponding documents will be kept on file and will be provided to IMI JU upon request.

Under the guidance of WP5, a document is prepared to detail regulations for data sharing and data transfer. Until available and fully implemented, only anonymized data will be shared. The work of WP5 will be implemented in the functions of Honest Broker, for requesting and providing access to the BIGPICTURE datasets.

Discussion

No deviations from the DoA were needed in the execution in this task.

Conclusion

The first and ultimately one of the most important tasks in WP3 is to establish a node network organisation model (NNOM) that ensures smooth and efficient implementation of the work proposed, namely to establish a central repository containing over three million whole slide images and associated metadata. Although central to WP3, this requires not only close collaboration between beneficiaries and SCTP within WP3 but also communication with partners in other WPs and Task forces working across WPs. Thus, the NNOM that has been implemented in the first months of the project provides both a practical framework for the development of the repository and a communication structure for all parties involved. Within this task the fundamental legal and logistic requirements, such as defining disease entities of cases to be collected and the associated metadata have been established and are continuously developed within the context of the NNOM.

Appendix 1

Relation to the Description of Action

The first and one of the major aims as outlined in the Description of Action (DoA) of the Grant Agreement (GA) of BIGPICTURE is to establish and employ a sustainable node network organisation model responsible for mediating and collecting datasets from beneficiaries and third - parties and to contribute them to the BIGPICTURE data repository in the description of WP3:

I: Organisation establishment and start-up of node network*

Task 3.1 - Establishing collector nodes– (M1- M6) (UMCU, MUW, UMR, LHTNHS, HUS, RÖ, DECI). The tasks of collecting and managing BIGPICTURE's large datasets will be distributed between **7 Nodes** that, as **functional units act as coordinators, collectors, and distributors of datasets**. Each node will be led by a coordinating manager. The nodes and coordinating manager are listed in Table WP3.1. The framework for the coordination and interaction between and within the nodes will be developed in collaboration with WP1. The **coordinating managers will establish a communication network amongst node contributors**, functioning as a board, **overseeing and coordinating the composition and diversity of the slide collection**. The node coordinators' board based on the **quality control committee's reports and tools, evaluate the contributed datasets and determine a prognosis on the upcoming annual needs of BIGPICTURE data sets**. Each node will promote the control process for the partners in the node to contribute the prospected **data sets with appropriate anonymized and pseudonymised metadata**, implemented automated quality control on images and data (developed by WP3 and WP4), adapted to the BIGPICTURE license for data use.

* In order to harmonise terminology used and to avoid confusion, all terms referring to the collector nodes have been amended to include the term "collector" in this document. This is in line with the BIGPICTURE Consortium agreement 11.9.

However in some instances, "collector" has been omitted. Thus the terms "Node Coordinator", "Node network board", "Node manager" and "Node Network board" refer to "Collector Node Coordinator", "Collector Node Network Board", "Collector Node Manager" and "Collector Node Network Board", respectively.

For clarity, the term "coordinating manager" refers to the "Collector Node Coordinator" which has been used throughout the document. Similarly, the term "collector node coordinators board" has replaced the term "collector node managing board".