

## Research article



## The IPASC data format: A consensus data format for photoacoustic imaging

Janek Gröhl<sup>a,b,1</sup>, Lina Hacker<sup>a,b,1</sup>, Ben T. Cox<sup>c</sup>, Kris K. Dreher<sup>d,e</sup>, Stefan Morscher<sup>f</sup>, Avotra Rakotondrainibe<sup>f</sup>, François Varray<sup>g</sup>, Lawrence C.M. Yip<sup>h,i</sup>, William C. Vogt<sup>j,2</sup>, Sarah E. Bohndiek<sup>a,b,\*,2</sup>, Members of the International Photoacoustic Standardisation Consortium (IPASC)<sup>a,b,3</sup>

<sup>a</sup> Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

<sup>b</sup> Department of Physics, University of Cambridge, Cambridge, United Kingdom

<sup>c</sup> Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

<sup>d</sup> German Cancer Research Center, Division of Computer Assisted Medical Interventions, Heidelberg, Germany

<sup>e</sup> Heidelberg University, Faculty of Physics and Astronomy, Heidelberg, Germany

<sup>f</sup> iThera Medical GmbH, Munich, Germany

<sup>g</sup> Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France

<sup>h</sup> Department of Medical Biophysics, Schulich School of Medicine and Dentistry, Western University, London, Canada

<sup>i</sup> Imaging Program, Lawson Health Research Institute, London, Canada

<sup>j</sup> Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, United States

## ARTICLE INFO

## Keywords:

Data format  
Photoacoustic imaging  
Open science  
Standardisation  
Metadata

## ABSTRACT

Photoacoustic imaging (PAI) is an emerging modality that has shown promise for improving patient management in a range of applications. Unfortunately, the current lack of uniformity in PAI data formats compromises inter-user data exchange and comparison, which impedes: technological progress; effective research collaboration; and efforts to deliver multi-centre clinical trials. To overcome this challenge, the *International Photoacoustic Standardisation Consortium* (IPASC) has established a data format with a defined consensus metadata structure and developed an open-source software application programming interface (API) to enable conversion from proprietary file formats into the IPASC format. The format is based on Hierarchical Data Format 5 (HDF5) and designed to store photoacoustic raw time series data. Internal quality control mechanisms are included to ensure completeness and consistency of the converted data. By unifying the variety of proprietary data and metadata definitions into a consensus format, IPASC hopes to facilitate the exchange and comparison of PAI data.

## 1. Introduction

Translation of photoacoustic imaging (PAI) into clinical trials has found a diverse landscape of potential applications for the technology, which unites the high contrast afforded by optical excitation with the resolution and penetration depth available through ultrasound detection [1–4]. PAI can be naturally applied across a range of length scales due to the broadband nature of photoacoustic waves, with systems ranging from microscopic resolution (at sub-millimetre imaging depth) to macroscopic, often tomographic, configurations (up

to centimetres imaging depth). Such diversity in system configurations enables application-specific customisation yet introduces complexity in data acquisition and management. In all cases, however, the physics of the photoacoustic effect and thus the overall nature of the data are the same. For these reasons, the algorithms used for data processing, image reconstruction and data display, as well as the respective metadata, have much in common. Despite these similarities, there is at present no community consensus on a data format for storing and sharing PAI data.

The value of a standardised data format has been recognised in other imaging communities, leading to the developments of tailored

\* Corresponding author at: Department of Physics, University of Cambridge, Cambridge, United Kingdom.  
E-mail address: [seb53@cam.ac.uk](mailto:seb53@cam.ac.uk) (S.E. Bohndiek).

<sup>1</sup> Shared First Authorship.

<sup>2</sup> Shared Last Authorship.

<sup>3</sup> <https://www.ipasc.science>.

formats such as the brain imaging data structure (BIDS) for magnetic resonance imaging (MRI) [5], Neuroimaging Informatics Technology Initiative (NIFTI) format for neuroimaging [6], the Shared Near-Infrared Spectroscopy Format (SNIRF) for functional near-infrared spectroscopy [7] or the ultrasound file format for ultrasound (UFF) [8]. The Digital Imaging and Communications in Medicine (DICOM) format is the current international standard for handling clinical imaging datasets [9], but the format is designed to store reconstructed images only and targeted towards clinical use.

The lack of a standardised format for raw PAI time series data impedes access, exchange and usability of acquired data between different researchers, vendors and organisations. A standard photoacoustic data format would have two substantial benefits. First, it would assist those working with photoacoustic data on a day-to-day basis by facilitating better data handling. For example, it could aid in comparison of data obtained from different sites or using instruments from different vendors. Second, a standard data format, if widely adopted, would help those developing photoacoustic technologies and visualisation software to present a united front to clinicians, data scientists, and other potential users. For example, a common format could accelerate the development and testing of new image reconstruction and processing methods by making PAI data more transparent.

Nonetheless, the development of a standardised format for PAI is associated with challenges. As highlighted above, system configurations for PAI are diverse, leading to a wide variability in the available (and relevant) metadata between systems. Moreover, time series data can be large in size and high-dimensional. Additionally, image acquisition parameters that can be highly relevant for post-processing algorithms may be unavailable (for example, because of hardware constraints or intellectual property protection) or subject to uncertainties.

The development, introduction, and adoption of a standard file format relies heavily on the widespread participation of researchers within the community. The International Photoacoustic Standardisation Consortium (IPASC) is a community-led effort committed to assisting and supporting the introduction of standards within the field of PAI. The work of IPASC is streamlined into three working groups, covering: study design; phantom development; and data acquisition and management (DAM). The DAM working group has taken up the task of developing a standardised way to store PAI data in a tailored digital format. The format underwent several optimisation procedures [10] in which feedback from the PAI community was collected and used to improve content and structure of the format. Here, these combined efforts are presented. First, relevant terminology and definitions are introduced. Second, the structure of the IPASC format and its associated metadata are highlighted. Finally, the organisation and workflow of an open-source application programming interface (API) for data conversion are presented. By introducing this format, IPASC hopes to facilitate PAI data handling and processing, thereby supporting future standards development within the PAI community and accelerating efforts in clinical translation.

## 2. Terminology and definitions

This section provides definitions of terms to avoid ambiguity. For further details, please refer to the IPASC ‘Terms and Definitions’ [11] and ‘Photoacoustic Data and Device Parameters’ consensus documents [12] that have been adopted by IPASC and are available on the IPASC website.

**Raw Time Series Data:** A *time series* refers to the time-sampled signal from one detection element. *Raw time series data* refers to an unaltered set of such time series, one for each detection element.

**Device:** A specific make and model of a photoacoustic device (hardware and/or software). This may include lab prototypes, systems sold for research use, or medical devices regulated in the United States by the Food and Drug Administration (FDA) and in the European Union (EU) by CE marking [13]. A distinction is drawn between

**full-scan** (image acquisition without sequential movement of illumination/detection array or target) and **composite-scan** devices (image acquisition by sequential change of position and/or orientation of illumination/detection array or target).

**Modality:** A category of imaging device, characterised by a distinct physical principle [13] (e.g. PAI).

**Measurement:** A set of raw time series data corresponding to a single illumination and acquisition step of the imaging device. For full-scan devices, only one measurement is taken, whilst for composite-scan devices, multiple measurements contribute to the same image. For composite scans, the *Measurement Spatial Pose* datum can be used to assign the relative spatial pose of a measurement.

**Image:** An array of values varying in two or more spatial dimensions derived from analysis of an imaging signal and corresponding to an array of spatial locations in the imaged object [13]. In the context of this document, a photoacoustic image refers to the result of mapping raw time series data into the spatial domain. An image can be derived from one or multiple *measurements*. An image might be subject to post-processing steps, such as envelope detection, display dynamic range, thresholding, bit depth discretisation, gain adjustment, colour mapping, or fluence correction.

**Detection Element:** A specific material element capable of converting mechanical energy to electrical energy and in some cases converting electrical energy to mechanical energy [14]. Here, it refers to e.g. a piezoelectric crystal or a laser interferometer.

**Illumination Element:** A specific material element capable of emitting light to illuminate a target, e.g. an optical fibre that conveys the light generated by a laser or light-emitting diode (LED) source. All defined illumination elements must be fixed throughout the entirety of the acquisition of one measurement.

## 3. The IPASC data format

A PAI dataset consists typically of one (or more) images representing the projection of a target volume onto an image plane (two-dimensional (2D) imaging), a series of images representing volume (three-dimensional (3D) imaging), or multiple acquisitions of the same 2D or 3D dataset over time and/or wavelength to produce a dynamic series of acquisitions (multi-dimensional imaging). The IPASC data format is designed to store measurements as raw time series data and not as reconstructed images to prevent loss of information and decrease the complexity of the data format. The raw time series data are accompanied by relevant metadata to enable reproducible reconstruction.

To accommodate the varied nature of PAI datasets, the Hierarchical Data Format (HDF5) [15] was chosen as the data format due to the following benefits: HDF5 (1) is able to store and organise large amounts of data within a single file container; (2) is platform-independent; (3) supports an unlimited variety of data types; (4) is widely used among the scientific community; (5) is openly available; (6) implements a high-level API with, for example, MATLAB,<sup>4</sup> Python [16], C++,<sup>5</sup> or Java interfaces;<sup>6</sup> and (6) contains the descriptive metadata within the data file, so metadata cannot get lost when exchanging files. HDF5 uses two objects types: datasets, which are multidimensional arrays of a homogeneous type, and groups, which are container structures holding datasets and other groups. In the IPASC data format, metadata are stored in the form of user-defined, named attributes attached to groups and datasets.

<sup>4</sup> MATLAB. The MathWorks Inc., Natick, Massachusetts, 2020.

<sup>5</sup> <http://h5cpp.org/> (accessed 03/03/2022).

<sup>6</sup> <https://portal.hdfgroup.org/display/HDF5/HDF5+Java+Documentation> (accessed 03/03/2022).

**Table 1**  
Overview of the metadata attributes used to describe each of the metadata items within the IPASC format.

Metadata attributes	
Necessity	Can be 'Minimal' or 'Report if present'.
dtype	Data type of the attribute.
Units	SI unit of the attribute if applicable.
Description	A short description of the attribute.
Method name	The name of the method or function that can be called in a programming language in order to obtain the information of this attribute.
Condition	Constraints of the attribute, if applicable.
Nested attribute	A sub-attribute that further describes an attribute.
Measurement device	A specific type of a nested attribute that further describes measurement device details if required.
Measurement device type:	A string literal describing the measurement device for this attribute, e.g. 'pyroelectric sensor' or 'wavemeter'.
Measurement device manufacturer:	A string literal describing the manufacturer of the measurement device, e.g. 'Thorlabs'.
Measurement device serial number:	A string literal comprising the serial number of the measurement device.
Calibration date:	A timestamp referring to the date when the measurement device was last calibrated.

#### 4. Metadata attributes

Metadata provide information about relevant aspects of data. Within the IPASC data format, each metadatum is characterised by a series of attributes to describe and define its use and boundary conditions (Table 1). If applicable, further specifications by nested attributes are given. All units of the metadata are given in the International System of Units (SI units) unless otherwise specified.

The metadata of the IPASC data format are organised into three sub-categories: (1) Minimal Metadata, (2) Acquisition Metadata and (3) Device Metadata. Whilst the format tries to accommodate the most relevant metadata parameters for full PAI data description, the wide variety of existing systems and constant technological progress may require the addition of further parameters that have not been considered yet in our current framework. To account for this need, the format allows for the addition of custom parameters by the end user, thereby providing the flexible structure that is required to accommodate the diverse nature of PAI data.

##### (1) Minimal Metadata

*Minimal Parameters* are indicated by the *Necessity* attribute and comprise all parameters that are required to read PAI data and reconstruct any image from the raw time series data. Any additional information should be reported in the metadata if available. The minimal parameter set contains:

**Container Format Metadata:** The inherent features of the file format, which specify mandatory parameters. They include the unique universal identifier (UUID), the type of compression, and the type of encoding.

**Binary Data Metadata:** The metadata that make the binary data machine-readable. They include specifications on data type, dimensionality, and the size of each dimension.

**A/D (Analogue/Digital) Sampling Rate:** A single value referring to the rate at which samples of the analogue signal are taken to be converted into digital form.

**Acquisition Wavelengths:** A 1D array that contains all wavelengths used for the image acquisition.

**Detector Positions:** The positions of each detection element in 3D Cartesian coordinates  $[x_1, x_2, x_3]$ .

**Field of View:** An array defining an approximate cuboid (3D) area that should be reconstructed in 3D Cartesian coordinates  $[x_{1\_start}, x_{1\_end}, x_{2\_start}, x_{2\_end}, x_{3\_start}, x_{3\_end}]$ . A 2D *Field of View* can be defined by setting the start and end coordinate of the respective dimension to the same value.

##### (2) Acquisition Metadata

The *Acquisition Metadata* describe the acquisition settings at data capture. In addition to the *Minimal Metadata*, they include the following parameters:

**Regions of Interest:** A list of named regions within the underlying 3D Cartesian coordinate system (cf. *Device Metadata*). Strings containing the region names are mapped to arrays that define either an approximate cuboid area (cf. *Field of View*) or a list of coordinates describing a set of 3D Cartesian coordinates surrounding the named region.

**Photoacoustic Imaging Device Reference:** A string referencing the UUID of the PAI device description as defined in the *Device Metadata*.

**Pulse Energy:** A value specifying the pulse energy used to generate the photoacoustic signal. If the pulse energies are averaged over many pulses, the average value must be specified.

**Measurement Timestamps:** An array specifying the time at which a measurement was recorded.

**Measurement Spatial Pose:** Coordinates describing the position and orientation changes of the acquisition system relative to the measurement of reference (first measurement).

**Time Gain Compensation:** An array containing relative factors that have been used to correct the time series data for the effect of acoustic attenuation.

**Overall Gain:** A single value describing a factor used to modify the amplitude of the raw time series data.

**Element-dependent Gain:** An array that contains the relative factors used for apodisation or detection element-wise sensitivity corrections.

**Temperature Control:** An array describing the temperature of the imaged space (covering both the imaged medium and the coupling agent) for each measurement.

**Acoustic Coupling Agent:** A string representing the acoustic coupling agent that is used.

**Speed of Sound:** Either a single value representing the mean global speed of sound in the entire imaged medium or a 3D array representing a heterogeneous speed of sound map in the device coordinate system. This definition covers both the imaged medium and the coupling agent.

**Scanning Method:** A string representing the scanning method that is used. The following descriptions can be used: ("composite\_scan", "full\_scan"). This flag determines the way the metadatum "measurement" is defined.

**Measurements Per Image:** A single value describing the number of measurements that constitute the dataset corresponding to one image.

**Frequency Domain Filter:** An array defining the frequency threshold levels that are applied to filter the raw time series data, containing [lower, higher]  $-3$  dB points of the filter in Hertz. [lower,  $-1$ ] denotes a high-pass filter and [ $-1$ , higher] denotes a low-pass filter.

##### (3) Device Metadata

The *Device Metadata* carry all information necessary to describe a PAI device (Fig. 1). In contrast to the *Acquisition Metadata*, they aim to facilitate modelling of how individual systems acquire data, for example, to enable corrections for directivity, or the application of illumination geometry-specific fluence correction algorithms, such as described in [17] or [18]. By collecting these metadata, a database

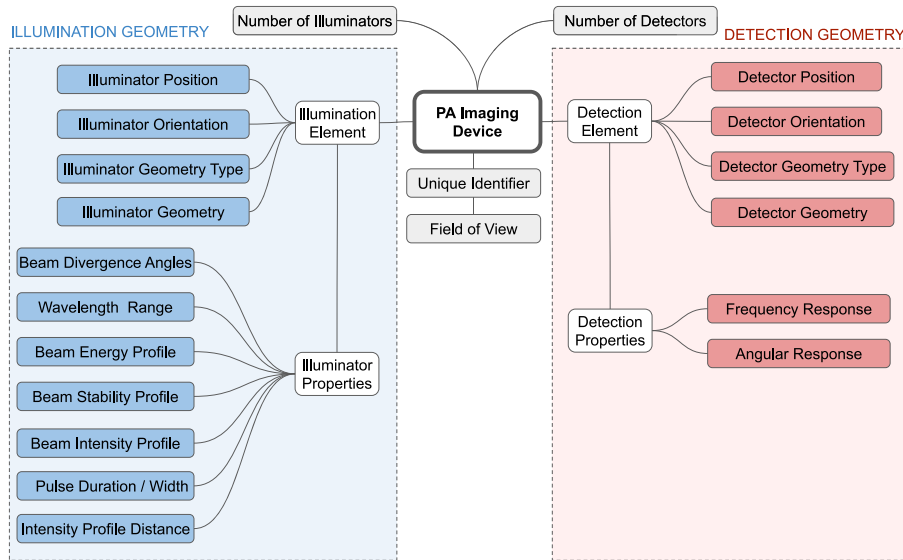


Fig. 1. Overview of the *Device Metadata* parameters. In this representation, a device is modelled to have a number of detection elements and illumination elements that each have characteristic properties.

of commercial and custom-built PAI devices can be created with all parameters necessary to construct a digital twin of the imaging device hardware. Each system is assigned with a unique identifier, which is referred to in the recorded photoacoustic data. The *Device Metadata* include:

**Universally Unique Identifier:** A randomly generated hexadecimal string that can be used to reference the device.

**Field of View:** Coordinates describing an approximate cuboid of the area detectable by the PAI device.

**Number of Illumination Elements:** The number of illuminators used in the PAI device.

**Number of Detection Elements:** The number of transducer elements used in the PAI device.

**Illumination Element**

**Illuminator Position:** Coordinates defining the position of the illuminator centroid.

**Illuminator Orientation:** Coordinates defining the direction unit vector of the illuminating beam.

**Illuminator Geometry Type:** A string describing the shape of the optical fibre (bundle) output.

**Illuminator Geometry:** Values defining the numerical geometry of the optical fibre (bundle) output. The data type and content of this metadatum are determined by the *illuminator geometry type* field.

**Wavelength Range:** An array of three values [minimum wavelength, maximum wavelength, accuracy] describing the wavelengths that can be generated by the illuminator.

**Laser Energy Profile:** An array of two double arrays [wavelengths, energies] which describes the laser energy of the illuminator.

**Laser Stability Profile:** An array of two double arrays [wavelengths, standard deviations] representing the standard deviation of the pulse-to-pulse laser energy of the illuminator.

**Pulse Duration/Width:** A value describing the total length of a laser pulse (measured as the time interval between the half-power points on the leading and trailing edges of the pulse.)

**Beam Intensity Profile:** Array of two double arrays [positions, intensities] specifying the relative laser beam intensity according to the planar emitting surface of the illuminator at the distance defined in *intensity profile distance*.

**Intensity Profile Distance:** An array describing the distance from the light source for measuring its *beam intensity profile*.

**Beam Divergence Angles:** A value specifying the opening angle of the laser beam with respect to the orientation vector.

**Detection Element**

**Detector Position:** Coordinates defining the position of the detection element centroid.

**Detector Orientation:** Coordinates defining the direction unit vector of the detector.

**Detector Geometry Type:** A string describing the type of detector geometry.

**Detector Geometry:** Values defining the numerical geometry of the detector. The data type and content of this metadatum are determined by the *detector geometry type* field.

**Frequency Response:** An array of two values [center frequency, bandwidth] characterising the frequency response of the detection element towards the incident pressure waves.

**Angular Response:** An array of two values [incident angle, response] characterising the angular sensitivity of the detection element to the incident angle (relative to the element’s orientation) of the incoming pressure wave.

To represent the spatial configuration of the device, a coordinate system is required. Here, the underlying coordinate system is assumed to meet the following conditions: The axes of the Cartesian coordinate systems are  $x_1 = (1, 0, 0)$ ,  $x_2 = (0, 1, 0)$ , and  $x_3 = (0, 0, 1)$  and the spatial location are given in units of metres [m]. Further conditions are not set, as it is sufficient if all axes definitions remain consistent with respect to the referenced *Device Metadata*. The field of view of the imaging device is given in the same coordinate system with six coordinate points ( $[x1_{start}, x1_{end}, x2_{start}, x2_{end}, x3_{start}, x3_{end}]$ ) and can be used to normalise the detector and illuminator positions. Thereby, the origin of the imaging system can be defined differently for different types of devices (cf. Fig. 2).

For ease of use, we suggest using the following convention for devices that collect time series data for reconstruction into 2D images:

1. The  $x_1$  axis should be defined as the horizontal (lateral) axis of the imaging plane.
2. The  $x_2$  axis should be defined as the vertical (axial) axis of the imaging plane.
3. The  $x_3$  axis should be defined as the normal (elevation) to the imaging plane.

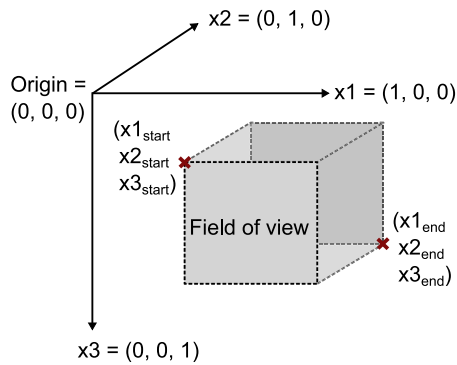


Fig. 2. Visualisation of the coordinate system definition used for the standard device definitions.

## 5. PACFISH: An open-source API for data access and conversion

To facilitate the use of the IPASC data format, a prototype Python-based software tool (PACFISH = Photoacoustic Converter for Information Sharing) was implemented (Fig. 3). PACFISH serves three purposes: (1) it helps vendors to integrate the IPASC data format export into their standard software; (2) it assists scientists to read and write data in the consensus HDF5 format; and (3) it helps the PAI community to create custom adapters that convert proprietary file formats into the consensus HDF5 format. PACFISH is available open-source on GitHub<sup>7</sup> under the commercially-friendly BSD-3-Clause licence<sup>8</sup> and contributions to the continued development of PACFISH are welcomed.

PACFISH is divided into the API, core, quality control, and iohandler modules. The API package (**pacfish.api** yellow module) can be used to facilitate the integration of adapters for conversion from arbitrary file formats into the IPASC data format. To create a conversion adapter, a Python representation of (1) the binary data, (2) the acquisition metadata dictionary, and (3) the device metadata dictionary needs to be implemented. The adapter must provide at least the *minimal metadata* as defined in the *Metadata Attributes* section. An option to add additional custom metadata items is also provided.

The core classes (**pacfish.core** green module) represent the metadata and data structure in Python. Each metadatum is described with specific device tags defining the name, data type, necessity and SI unit (if applicable), and setting a value constraint. Basic metadata constraints have been implemented to avoid accidental typos within the values field (e.g. only positive numbers larger than zero are applicable for acquisition wavelengths). If the value is not within the constraints a *TypeError* is raised. Metadatum-specific functions enable easy addition of the values for the specific metadata field.

The quality control functionalities (**pacfish.qualitycontrol** blue module) ensure the correctness of the conversion into the IPASC format: a *completeness checker* tests that all metadata are being called and a *consistency checker* ensures that all metadata are within their constraints. An automatically-generated output report gives a human-readable summary of the quality control checks and ensures that the likelihood of conversion mistakes are minimised. To assess the *Device Metadata*, the detector and illuminator positions can be represented in a 3D coordinate system as visual control (Fig. 4).

Finally, the I/O functionality (**pacfish.iohandler** red module) enables reading and writing of IPASC-formatted data files. Code listing 1 shows how to load and access the standardised data and metadata using PACFISH.

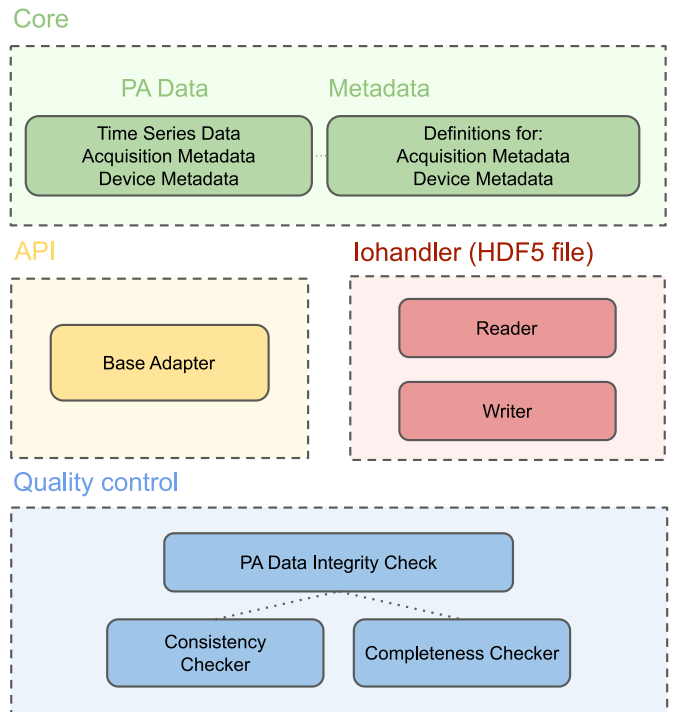


Fig. 3. Overview of the software components of PACFISH for Python-based access to the IPASC data format. Different colours represent different modules that encapsulate separate responsibilities of the API: The *api* module is displayed in yellow, the *core* module in green, the *qualitycontrol* module in blue and the *iohandler* in red.

### API Usage Examples

In the following, we provide some example code stubs that can be used to understand the handling of PACFISH, using the version number available at the time of the publication date of this paper. For up-to-date examples, please visit the PACFISH GitHub page.<sup>9</sup>

#### 1. Using the API to work with data in the IPASC format:

This listing includes Python code showcasing how to use PACFISH to read data in the IPASC format, to unpack metadata relevant for image reconstruction, and to write data to the hard drive.

```

1 import pacfish as pf
2
3 # Loading data from the hard drive
4 pa_data = pf.load_data("path/to/hdf5file.hdf5")
5
6 # Visualising the device geometry
7 pf.visualize_device(pa_data.meta_data_device)
8
9 # example unpacking of metadata of file
10 field_of_view = pa_data.get_field_of_view()
11 detection_elements = dict()
12 detection_elements['pos'] = pa_data.get_detector_position()
13 detection_elements['ori'] = pa_data.get_detector_orientation()
14 detection_elements['geom'] = pa_data.get_detector_geometry()
15 detection_elements['type'] = pa_data.get_detector_geometry_type()
16 time_series_data = pa_data.binary_time_series_data

```

<sup>7</sup> <https://github.com/IPASC/PACFISH> (accessed 03/03/22).

<sup>8</sup> <https://opensource.org/licenses/BSD-3-Clause> (accessed 03/03/22).

<sup>9</sup> <https://github.com/IPASC/PACFISH> (accessed 03/03/22).

```

17
18 # Call to a reconstruction algorithm (not
    implemented in PACFISH)
19 reconstruction = reconstruct(time_series_data
    =time_series_data,
20
    detection_elements=detection_elements,
21                               field_of_view=
    field_of_view
22
23 # Writing of data to hard drive
24 pf.write_data("path/to/new/file.hdf5",
    pa_data)
    
```

Listing 1: Example listing how to use PACFISH to read and write data in the IPASC format and how to access the data and metadata.

2. Using the API to write a custom device-specific adapter:

```

1 from pacfish import BaseAdapter, MetaDatum
2 from pacfish import DeviceMetaDatumCreator
3 from pacfish import DetectionElementCreator
4 from pacfish import
    IlluminationElementCreator
5
6 # Create a custom Adapter class
7 class CustomAdapter(BaseAdapter):
8
9     def __init__(self, file_name):
10         # Include code for reading the
            proprietary file and pre-process the read
            data.
11         super(CustomAdapter).__init__()
12
13     def generate_binary_data(self):
14         # Include code to read the time
            series data.
15         return data
16
17     def generate_meta_data_device(self):
18         # The DeviceMetaDatumCreator supports
            the collection of the device-specific
            metadata
19         device_creator =
            DeviceMetaDatumCreator()
20
21         # To add detection and illumination
            elements use the DetectionElementCreator
            and IlluminationElementCreator helper
            classes, respectively, e.g.:
22
23         detection_element_creator =
            DetectionElementCreator()
24
25         detection_element_creator.
            set_detector_position(YOUR_POSITION)
26
27         device_creator.add_detection_element(
            detection_element_creator.get_dictionary
            ())
28
29     def set_metadata_value(self, metadata_tag
        : MetaDatum):
30         # With this method the acquisition
            metadata are provided for the pa_data
            object. A MetaDatum is given and
            appropriate value needs to be returned:
31
32         if metadata_tag ==
            MetadataAcquisitionTags.UUID:
33             return "YOUR_UUID"
34         elif metadata_tag ==
            MetadataAcquisitionTags.DATA_TYPE:
35             return YOUR_DATA_TYPE
36         else:
37             return None
    
```

Listing 2: Example listing how to use PACFISH to create a custom adapter for reading and conversion of a proprietary data format.

3. Using the custom adapter to access/convert photoacoustic data:

```

1 from pacfish import write_data
2 from custom_module import CustomAdapter
3 from pacfish import quality_check_pa_data
4
5 # Instantiating the adapter
    
```

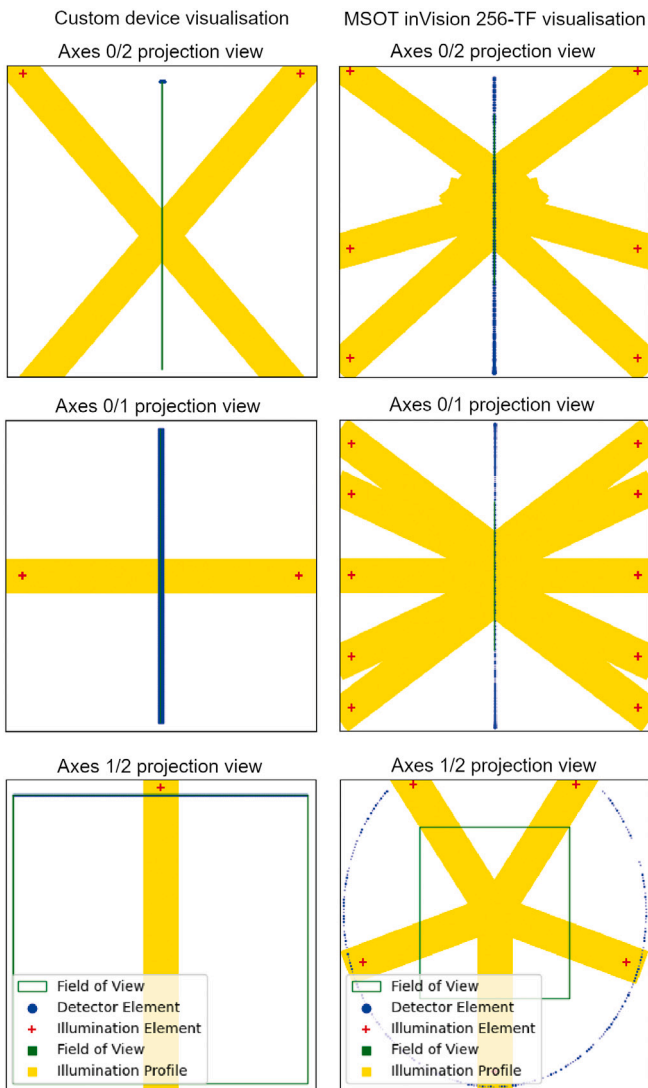


Fig. 4. Visualisation of example photoacoustic devices generated from the IPASC metadata. The upper image shows a linear transducer with two illumination slits at either side as described in [19]. The lower image is a representation of the MSOT inVision 256-TF (iThera Medical GmbH, Munich, Germany).

```

6 adapter = CustomAdapter("path/to/MY_DATA.
  CUSTOM_FORMAT")
7
8 # Generate the pa_data object with the
  adapter
9 pa_data = adapter.generate_pa_data()
10
11 # Check quality of the resulting pa_data
  object
12 quality_check_pa_data(pa_data, verbose=True,
  log_file_path="MY_LOGFILE_PATH")
13
14 # Writing of pa_data object as HDF5 file to
  hard drive
15 write_data("path/to/new/file.hdf5", pa_data)

```

Listing 3: Example listing how to use a custom adapter to read and quality check photoacoustic data.

Sample data in the IPASC format can be found on Zenodo,<sup>10</sup> and application examples can be accessed on Github.<sup>11</sup> Fig. 5 shows the Zenodo sample data, displaying the device visualisation, simulated time series data, and reconstruction result of a conventional back projection algorithm [20] for four different sample data sets with different detection geometries.

<sup>10</sup> <https://zenodo.org/record/5938838> (accessed 03/03/22).

<sup>11</sup> <https://github.com/IPASC/PACFISH/tree/main/examples> (accessed 03/03/22).

## 6. Discussion

We have presented the IPASC data format for PAI data storage. The IPASC format is designed to store raw time series data with associated metadata as HDF5 files. To facilitate the use of the IPASC data format, a Python-based software tool (PACFISH = Photoacoustic Converter for Information Sharing) has been created and tested by members of the DAM working group; it is provided to the PAI community to use and further develop.

PAI systems can have vastly different illumination and detection geometries, including linear array, cylindrical, and hemispherical designs. For any hardware configuration, the scanning protocol can also vary (full-scan vs. composite-scan). While the IPASC format is designed to accommodate all needs, limitations remain.

First, the IPASC data format does not yet support the inclusion of data from bimodal systems, such as combined photoacoustic and ultrasound systems. Future implementations should account for these hybrid system types, as they become increasingly common (cf. e.g. [21,22]), especially in clinical settings.

There are also limitations associated with the use of the underlying data format HDF5, such as the risk of corruption or potential performance issues when dealing with large datasets [23]. Whilst these drawbacks are currently outweighed by the advantages of HDF5 – such as its flexibility, organisation, open access, and its capacity to store large annotated datasets within a single file – the limitations should be kept in mind to ensure safe use of the format.

The IPASC data format is designed to accommodate raw time series measurements with all relevant metadata. Storing raw data is

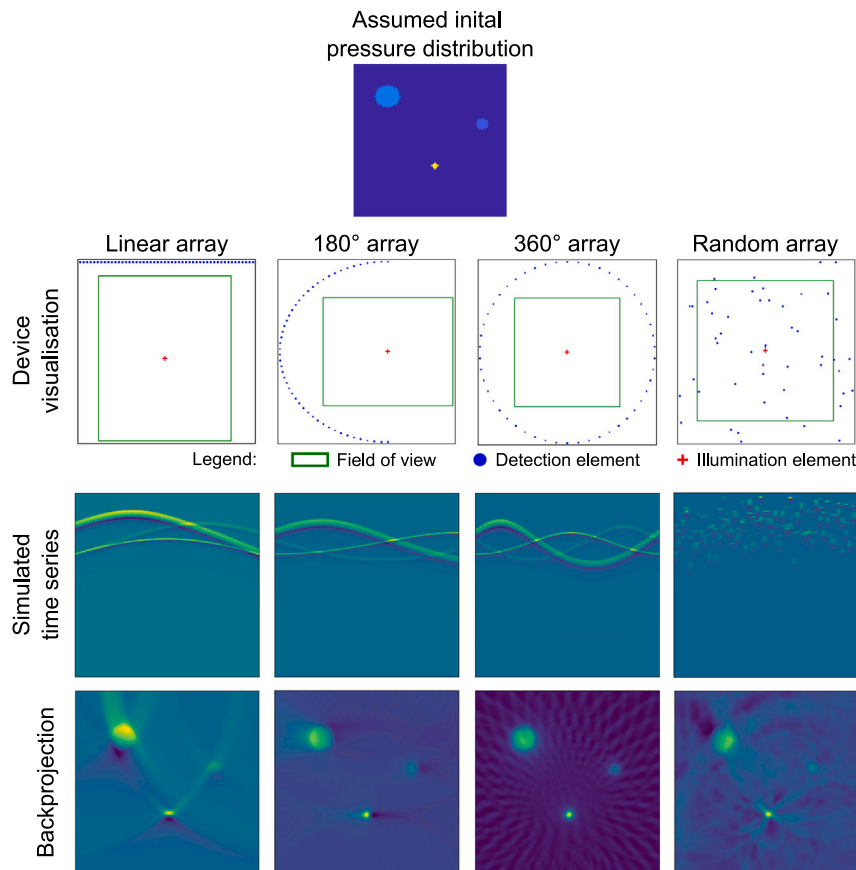


Fig. 5. Representative reconstructions of simulated data in the IPASC data format using different detector geometries. The simulated assumed initial pressure distribution is visualised at the top. Device visualisation, simulated time series, and reconstructed data using back projection are visualised underneath in the top, middle and bottom row, respectively. Four different detection arrays have been used: Linear array (first column), semi-circular (180°) array (second column), full circular (360°) array (third column), and a random array (fourth column).

preferable to storing reconstructed images, as reconstruction can lead to image artefacts and loss of information. For reconstructed images, compatibility with the DICOM format is desirable, as it is the most widely used format in management of medical image data. The IPASC industry board, which includes vendors from commercial PAI systems, is currently working towards integrating photoacoustic image data into DICOM. As the DICOM tags will be based on the metadata naming convention presented here, IPASC hopes to enable appropriate mapping between the two formats and to ensure their compatibility.

The PACFISH Python tool is still under development and is currently being tested by members of the photoacoustic community and by vendors to assess its applicability in an industrial setting. Adapters to other programming languages, such as MATLAB, could be added to increase the accessibility of the PACFISH within the user community. Moreover, the direct integration of PACFISH into software tools relevant for numerical forward modelling, such as k-Wave [24] or Monte Carlo eXtreme (MCX) [25] would be beneficial.

Our focus in the near future will be to widen the adaptation of the IPASC format for both the vendor and research communities. Besides implementing additional adapters to proprietary formats, programming languages and software tools, we would like to make established image reconstruction algorithms compatible with the IPASC format to streamline post-processing pipelines using the format. A further aim is to create an open-access database that includes data held in the IPASC format, to facilitate reproducibility studies and structured evaluation and benchmarking of reconstruction algorithms. By introducing the IPASC format, we hope to facilitate technological advancement, foster community collaboration, and ultimately accelerate clinical translation and adoption of this modality.

This article serves only as an introduction to the IPASC data format. The *Agreed Proposal* consensus document [12] with a complete description of the IPASC data format can be found on the IPASC website: <https://www.ipasc.science/>, which also contains further resources and guidance on how to provide feedback on the documents.

#### CRedit authorship contribution statement

**Janek Gröhl:** Writing – original draft, Software. **Lina Hacker:** Writing – original draft, Software. **Ben T. Cox:** Writing – review & editing, Supervision. **Kris K. Dreher:** Writing – review & editing, Software. **Stefan Morscher:** Writing – review & editing, Supervision. **Avotra Rakotondrainibe:** Writing – review & editing, Software. **François Varray:** Writing – review & editing, Software. **Lawrence C.M. Yip:** Writing – review & editing, Software. **William C. Vogt:** Writing – review & editing, Supervision. **Sarah E. Bohndiek:** Writing – review & editing, Supervision. **Members of the International Photoacoustic Standardisation Consortium (IPASC):** Supervision.

#### Declaration of competing interest

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.pacs.2022.100339>. Sarah Bohndiek reports a relationship with EPFL Center for Biomedical Imaging that includes: speaking and lecture fees. Sarah Bohndiek reports a relationship with PreXion Inc that includes: funding grants. Sarah Bohndiek reports a relationship with iThera Medical GmbH that includes: non-financial support. Avotra Rakotondrainibe reports a relationship with iThera Medical GmbH that includes: employment. Stefan Morscher reports a relationship with iThera Medical GmbH that includes: employment.

#### Acknowledgements

The authors would like to thank all members of the IPASC DAM working group for their helpful input on this study shared in project discussions. Moreover, they would like to thank all IPASC members for their valuable contributions to the general ongoing standardisation activities.

#### Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the US Department of Health and Human Services. This article reflects the views of the authors and should not be construed to represent FDA views or policies.

B.T.C. is an Editor of this journal and S.E.B. serves on the Editorial Board, but they were only involved in the peer review process to the extent that they are authors. They had no access to information regarding the review process that would not ordinarily be available to an author, and full responsibility for the peer-review process for this article was delegated to another Editor.

#### References

- [1] P. Beard, Biomedical photoacoustic imaging, *Interface Focus* 1 (4) (2011) 602–631.
- [2] J. Xia, J. Yao, L.V. Wang, Photoacoustic tomography: principles and advances, *Electromagn. Waves (Cambr. Mass.)* 147 (2014) 1.
- [3] P.K. Upputuri, M. Pramanik, Recent advances toward preclinical and clinical translation of photoacoustic tomography: a review, *J. Biomed. Opt.* 22 (4) (2016) 041006.
- [4] A.B.E. Attia, G. Balasundaram, M. Moothanchery, U. Dinish, R. Bi, V. Ntziachristos, M. Olivo, A review of clinical photoacoustic imaging: Current and future trends, *Photoacoustics* 16 (2019) 100144.
- [5] K.J. Gorgolewski, T. Auer, V.D. Calhoun, R.C. Craddock, S. Das, E.P. Duff, G. Flandin, S.S. Ghosh, T. Glatard, Y.O. Halchenko, et al., The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments, *Sci. Data* 3 (1) (2016) 1–9.
- [6] R. Cox, J. Ashburner, H. Breman, K. Fissell, C. Haselgrove, C. Holmes, J. Lancaster, D. Rex, S. Smith, J. Woodward, et al., A (sort of) new image data format standard: Nifti-1: WE 150, *Neuroimage* 22 (2004).
- [7] Q. Fang, Snirf file format, 2018, <https://github.com/fNIRS/snirf/> [Accessed: 03/03/2022].
- [8] O. Bernard, D. Bradway, H.H. Hansen, P. Kruizinga, A. Nair, D. Perdios, S. Ricci, O.M.H. Rindal, A. Rodriguez-Molares, M.B. Stuart, et al., The ultrasound file format (UFF)-first draft, in: 2018 IEEE International Ultrasonics Symposium (IUS), IEEE, 2018, pp. 1–4.
- [9] C. Parisot, The DICOM standard, *Int. J. Card. Imaging* 11 (3) (1995) 171–177.
- [10] IPASC, Consensus finding procedure, 2019, [https://www.ipasc.science/ipasc.science/documents/IPASC\\_Consensus\\_Workflow\\_2019\\_09.pdf](https://www.ipasc.science/ipasc.science/documents/IPASC_Consensus_Workflow_2019_09.pdf) [Accessed: 03/03/2022].
- [11] IPASC, Proposed list of terms and definitions, 2019, [https://www.ipasc.science/documents/20191210\\_Terms\\_Definitions\\_For\\_PAT.pdf](https://www.ipasc.science/documents/20191210_Terms_Definitions_For_PAT.pdf) [Accessed: 03/03/2022].
- [12] IPASC, Photoacoustic data and device parameters, 2020, [https://www.ipasc.science/ipasc.science/documents/20210916\\_IPASC\\_Format\\_V2.pdf](https://www.ipasc.science/ipasc.science/documents/20210916_IPASC_Format_V2.pdf) [Accessed: 03/03/2022].
- [13] J.P. O'Connor, E.O. Aboagye, J.E. Adams, H.J. Aerts, S.F. Barrington, A.J. Beer, R. Boellaard, S.E. Bohndiek, M. Brady, G. Brown, et al., Imaging biomarker roadmap for cancer studies, *Nat. Rev. Clin. Oncol.* 14 (3) (2017) 169.
- [14] IEC TR 60854:1986, Methods of measuring the performance of ultrasonic pulse-echo diagnostic equipment, IEC TR 60854.
- [15] M. Folk, G. Heber, Q. Koziol, E. Pourmal, D. Robinson, An overview of the HDF5 technology suite and its applications, in: Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases, 2011, pp. 36–47.
- [16] A. Collette, Python and HDF5: Unlocking Scientific Data, O'Reilly Media, Inc, 2013.
- [17] F.M. Brochu, J. Brunker, J. Joseph, M.R. Tomaszewski, S. Morscher, S.E. Bohndiek, Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in photoacoustic tomography, *IEEE Trans. Med. Imaging* 36 (1) (2016) 322–331.
- [18] X. Zhou, N. Akhlaghi, K.A. Wear, B.S. Garra, T.J. Pfeifer, W.C. Vogt, Evaluation of fluence correction algorithms in multispectral photoacoustic imaging, *Photoacoustics* 19 (2020) 100181.



- [19] J. Gröhl, T. Kirchner, T. Adler, L. Maier-Hein, Confidence estimation for machine learning-based quantitative photoacoustics, *J. Imaging* 4 (12) (2018) 147.
- [20] M. Xu, L.V. Wang, Universal back-projection algorithm for photoacoustic computed tomography, *Phys. Rev. E* 71 (1) (2005) 016706.
- [21] R.G. Kolkman, P.J. Brands, W. Steenbergen, T.G. van Leeuwen, Real-time in vivo photoacoustic and ultrasound imaging, *J. Biomed. Opt.* 13 (5) (2008) 050510.
- [22] G.-S. Jeng, M.-L. Li, M. Kim, S.J. Yoon, J.J. Pitre, D.S. Li, I. Pelivanov, M. O'Donnell, Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction, *Nature Commun.* 12 (1) (2021) 1–12.
- [23] S.-A. Dragly, M. Hobbi Mobarhan, M.E. Lepperød, S. Tennøe, M. Fyhn, T. Hafting, A. Malthe-Sørensen, Experimental directory structure (exdir): An alternative to HDF5 without introducing a new file format, *Front. Neuroinform.* 12 (2018) 16.
- [24] B.E. Treeby, B.T. Cox, K-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, *J. Biomed. Opt.* 15 (2) (2010) 021314.
- [25] Q. Fang, D.A. Boas, Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units, *Opt. Express* 17 (22) (2009) 20178–20190.



**Janek Gröhl** received his M.Sc. degree in medical informatics from the University of Heidelberg and Heilbronn University of Applied Sciences in 2016. He received his Ph.D. from the medical faculty of the University of Heidelberg in April 2021. In 2020, he worked as a postdoctoral researcher at the German Cancer Research Center in Heidelberg, Germany and is currently working as a research associate at the Cancer Research UK Cambridge Institute in Cambridge, United Kingdom. He does research in computational biophotonics focusing on data-driven methods for data processing and signal quantification in photoacoustic imaging.



**Lina Hacker** received her B.Sc. degree in Molecular Biomedicine at the University of Bonn (Germany) and her M.Sc. degree in Biomedical Engineering at RWTH Aachen (Germany). Currently, she is pursuing a Ph.D. in Medical Sciences at the University of Cambridge (UK) focusing on the technical and biological validation of photoacoustic imaging biomarkers.



**Kris Dreher** received his M.Sc. degree in Physics from the University of Heidelberg in 2020. He is currently pursuing a Ph.D. at the Division of Computer Assisted Medical Interventions (CAMI), German Cancer Research Center (DKFZ) and does research in synthetic photoacoustic image generation and deep learning-based domain adaptation methods to tackle the inverse problems of photoacoustic imaging.



**Stefan Morscher** completed his degree at the faculty for electrical engineering at Technische Universität München with his Master thesis on optoacoustic imaging under Prof Vasilis Ntziachristos. Thereafter, he joined the R&D team of the newly founded iThera Medical to make MSOT available to the scientific community, where he is now leading R&D efforts for software and algorithms as the “V.P. Software Development” to translate the preclinical success story of MSOT to clinical applications. He maintains close ties with many researchers in the field of optoacoustics to enable a full understanding of all aspects of optoacoustic imaging and pushes for standardisation of photoacoustic imaging under IPASC as well as DICOM.



**Avotra Rakotondrainibe** completed her Master degree at Telecom Physique Strasbourg (Université de Strasbourg), with a focus on engineering for life science. She is currently enrolled as “R&D Engineer” at iThera Medical to support R&D activities in the Software space. She contributes to the development of iThera Medical’s data analysis library to enable advanced applications and fast innovative algorithm prototyping for the next generations of MSOT devices.



**François Varray** was born in Montpellier, France, in 1985. He received the Engineering Diploma degree and the master’s degree in image and signal processing from the École des Mines de Saint-Étienne, France, in 2008, and the Ph.D. degree with a focus on nonlinear ultrasound simulation in 2011. His Ph.D. research was realised in co-agreement between the Centre de Recherche en Acquisition et Traitement de l’Image pour la Santé (CREATIS, University of Lyon 1), Lyon, France, and the MSD Laboratory, Florence, Italy (University of Firenze). Since 2013, he has been an Associate Professor with CREATIS. His research interests include nonlinear ultrasound propagation simulation, motion estimation, cardiac imaging, photoacoustic imaging and anisotropy imaging.



**Lawrence CM Yip** is a Ph.D. candidate in Medical Biophysics at Western University in London, Ontario, Canada, where he also completed his Honours BMSc in Medical Science. His current research interests include developing medical imaging research devices with a focus on photoacoustic tomography and near-infrared spectroscopy.



**William C. Vogt** received his BS degree in mechanical engineering from the University of Massachusetts Amherst in 2009 and his Ph.D. in biomedical engineering from Virginia Polytechnic Institute and State University in 2013. Since joining the FDA in 2013, he has been conducting regulatory science to develop tools and test methods for evaluating the safety and effectiveness of photoacoustic imaging devices. His research interests include photoacoustic imaging, tissue phantoms, nanoparticles, standardisation, and biophotonic medical device characterisation and evaluation.



**Sarah Bohndiek** completed her Ph.D. in Radiation Physics at University College London in 2008 and then worked in both the UK (at Cambridge) and the USA (at Stanford) as a postdoctoral fellow in molecular imaging. Since 2013, Sarah has been a Group Leader at the University of Cambridge, where she is jointly appointed in the Department of Physics and the Cancer Research UK Cambridge Institute. She was appointed as Full Professor of Biomedical Physics in 2020. Sarah was recently awarded the CRUK Future Leaders in Cancer Research Prize and SPIE Early Career Achievement Award in recognition of her innovation in biomedical optics.