4Quant
BIG IMAGE ANALYTICS

# WHITEPAPER
# **HICCAP**

2018-02-22

Dr. Kevin Mader, Flavio Trolese, Dr. Amine Korchi, Joachim Hagger

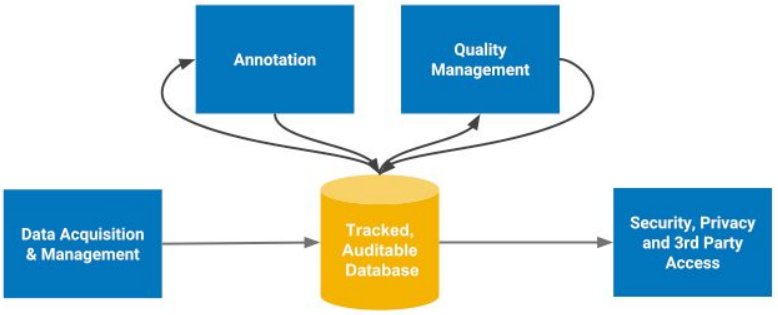
# MOTIVATION: MEDICAL AI UNLEASHED THROUGH CONTINUOUS HIGH QUALITY DATA STREAMS

AI promises to revolutionize every aspect of medicine with an ever larger number of algorithms matching and even exceeding human performance at clinically-relevant tasks. The bottleneck for these developments has become access to anonymized, systematically aggregated, structured, synthesised and enriched clinical (big) data. We work with a number of hospitals and clinics worldwide to provide these data streams to

- Accelerated development of new machine learning-based decision support and CADx applications
- Deliver deep clinical insights into the patient journey and the effectiveness and safety of pharmaceutical products.
- Identification of biomarkers through translational biomedical informatics
- Suggest appropriate treatment program for patients based on their genetic makeup and avoiding tests that are known to be unlikely to work
- Minimize therapy revisions, likely failures and repeat procedures, thereby saving effort, costs, and complications
- Use intelligent approach to drug development, by aborting trials that are unlikely to work, or to focus on efforts to develop drugs to specific population subsets
- Trial failures can be reduced by using statistical tools and algorithms to improve clinical trial design and patient recruitment to better match treatments to individual patients.

Making medical data Accessible, Findable, Interoperable and Reusable.

# HICCAP PLATFORM

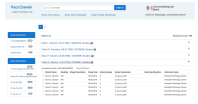
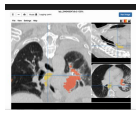
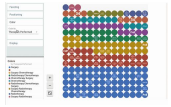



## SCOPE

This paper discusses 4Quant HICCAP annotation platform for quickly and continuously creating curated data in the clinical setting. The platform is built on a number of core components enabling us to integrate with existing systems, provide pixel-exact annotations with single clicks,

incorporate medical information into workflow and scale to thousands of annotators covering millions of patients.

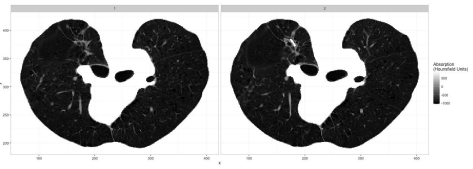
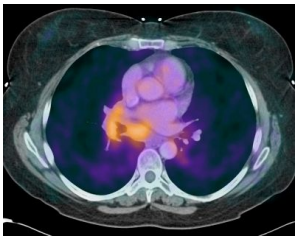
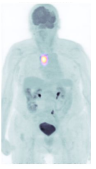
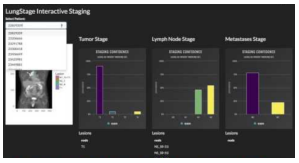
Our system has been applied to a number of different cases and image types from diffusion-weighted MRI images to PET/CT and has used a combination of standard image processing approaches (with integrations for ImageJ, ITK, OpenCV, and scikit-image) and the latest deep learning tools to solving problems. The platform enables all of these heterogeneous data sources and approaches to be easily explored in a single, distributed environment.

| Workflow Step | Problem | Solution |
|---|---|---|
| **Acquisition**  | PACS and RIS systems are patient-focused, slow, poorly scalable and difficult to search | Integration with existing PACS and RIS systems as a DICOM and HL7 node providing high-speed, indexed, intelligent searches to create patient cohorts |
| **Annotation**  | Medical images are large and only tiny fractions of the image are clearly identifiable as diseased | State of the art tools for quickly and accurately segmenting complex volumes from single points by incorporating anatomical and physiological understanding of the images. |
| **Quality Control**  | Physicians have different opinions on cases and many patients and lesions are not always clearly identifiable | Project management tools for ensuring multiple readers per case, clarification of unclear cases, performance and quality assessment of readers and ultimately leading to much higher model accuracy from a given number of cases. |
| **Access**  | Medical formats for storing annotation and curations are incomplete and make training and validating models difficult. Proper anonymization and privacy require robust tools to ensure. | Clear boundaries between intrahospital and outside data with seamless transformation of data and structures from clinical formats to model training formats with a wide selection of automatically applied anonymization tools. |

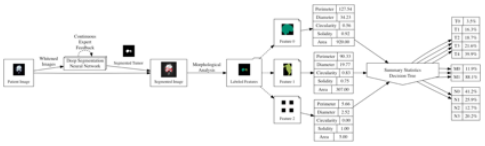
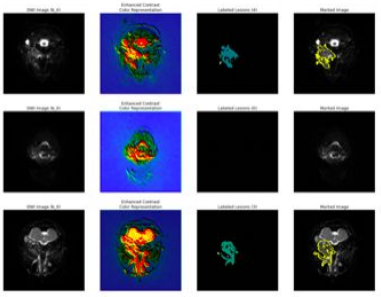
## Example Projects

We have used our annotation platform on a number of academic, research and commercial scale projects.

### Lung Product Suite

| Disease Target | Image Type | Training Size | Output |
|---|---|---|---|
| **COPD Analytics**  | Modality: Chest CT or Whole Body CT | 204,800 images from 3 countries and 3 manufacturers |  Collected data to build COPD algorithms requiring segmentation of lungs and quantification of disease progress. |
| **Lung Cancer Lesion Detection**  | Modality: PET/CT | 500 patients and 700,000 images | Collected voxel-level annotations of every malignant lesion in NSCLC patients.  |
| **Lung Cancer Staging**  | Modality: Full body PET/CT | 1000 patients and 1.4M images | Collected patient level information on NSCLC staging for radiological, clinical and pathological-based TNM staging. Acquired information on lesion specific mutations, treatment, and outcome for all patients.  |

## Other Sectors

| | | | |
|---|---|---|---|
| **Osteoporosis and Fracture Risk**  | Modality: CT (Chest) | 300,000 images | Collected information on risk  |
| **Head and Neck Cancer Staging**  | Modality: Diffusion Weighted MRI | 200 patients and 137,000 images | Collected spatial and patient level information on the lesions and staging for nasopharyngeal cancer patients.   |

# TECHNICAL DETAILS

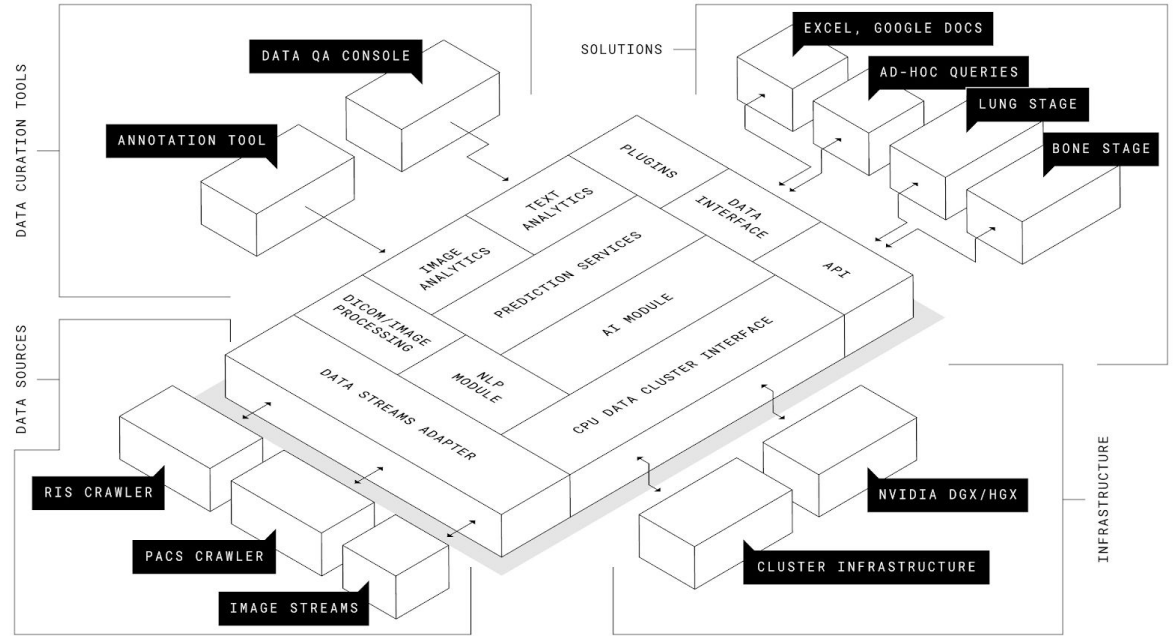
HICCAP provides open interfaces to the outside world:

*Database adaptors to various systems*: PACS Crawler allows for a full faceted search on all DICOM metadata of your PACS system. PACS Crawler is scalable to millions of studies and provides result sets in a fraction of a second. RIS Crawler does the same on clinical data coming from RIS (or other clinical information system). HICCAP combines the various data sources, including real-time streams of images or video data, for any analytics task.

*Integration of Data Curation Tools*: Data curation addresses key tasks for big data analysis and machine learning. Data needs to be structured, labeled, annotated, cleansed, combined to gain the quality requirements for ground truth data used as training data. Add-ons to the platform include applications for efficient labeling and annotation of medical imaging data as well as performing data quality checks.

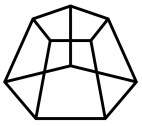
*Operating infrastructure*: HICCAP scales on a variety of standard platforms, from notebook computers to computing and storage clusters including GPU-accelerated servers. The platform can either be deployed on-site for very data sensitive or regulatory demands or on generic cloud infrastructure for better scalability.

Solutions built on HICCAP either use a mix of generic and application-specific API or consume analytics results using an interface for structured data. Plug-ins implementing complex analytics tasks and complementing applications have full access on all data and analysis services, modules, and image or data processing services and profit directly from near-linear scalability of the platform.

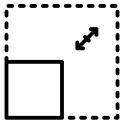



*The* HICCAP *Annotation and Analytics Platform*

## HIGHLIGHTS OF THE PLATFORM

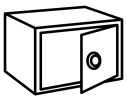


*Volumetric analysis and quantification*: The analysis is fully 3D and delivers quantitative metrics based on volume in addition to simple slice-based diameter and shape to provide a greater degree of precision for characterizing disease and establishing highly accurate diagnostic criteria.

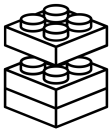


*Simple, scalable Integration*: The results can easily be output in a number of different formats. For small studies this comes in the form of a simple REST API, JSON, or as Excel workbooks. For larger studies this can be integrated into a Big Data workflow using MongoDB, Elasticsearch or other database-backends for batch processing or Kafka for real-time streaming analysis.

*Big Data ready*: Scales linearly and elastically to available resources, interactive, fault-tolerant, processes terabytes to petabytes of image and video data, near real-time analytics available. Supports the “5 V’s” in big data: Value, Volume, Velocity, Veracity, Variety.



*Protects previous investments*: HICCAP fully supports existing investments in ImageJ, Matlab, or R and parallelizes single-CPU algorithms. HICCAP integrates code modules and algorithms directly using a plug-in mechanism.



*Ready to use modules*: HICCAP comes with a variety of ready-made modules of common tasks in medical imaging, such as organ segmentation, multi-modality registration, lesion detection and classification, diameter and volume measuring, contrast simulation etc.