

Implementing digital pathology for biobanking

The flexibility of digital pathology hardware and software solutions allows institutions, such as the Brussels Free University, to create bespoke solutions to meet individual biobanking needs

Yves Sucaet PhD

Digital Pathology Manager,
Brussels Free University, Belgium

B iobanks play a central role in translational biomedical research. Large collections of high-quality samples can help researchers identify clinically useful markers of disease and develop novel drugs. In order to fulfil this role, however, it is essential that the samples are well-documented with up-to-date epidemiological, clinical and molecular data. The vast numbers of patients and controls contained in these biobanks needs to be easy to browse and search, too.

At Brussels Free University (VUB), and with the financial support of InnoViris, an online histopathology platform was built that presents a catalogue of tissues available in the biobank. This was done to contribute to the valorisation of the biobank, and to lead to new collaborations between academia and industry.

The outcome of this project is a hybrid solution that consists of commercial hardware and software, as well as open source packages. This article provides a brief introduction to digital pathology and how this new technology has significantly enhanced our research and education activities.

What is a biobank?

A 'biobank' is an organised collection of

physical human material. The material is stored with the intent to facilitate future scientific research, and must be well-documented and annotated. The term is typically used for material of human origin, but can also be used for collections of plants, animals, and microbes. Specimen types include blood, urine, saliva, skin cells, organ tissue, and other materials.

The main task of a biobank is to preserve specimens for the future. For this purpose, cryogenic storage facilities are provided for the samples, which can range in size from individual refrigerators to warehouses. Biobanks are maintained by institutions like our own, as well as by pharmaceutical companies. Disease-oriented biobanks, usually located at a university-based hospital, often have formalin-fixed paraffin-embedded (FFPE) tissue available, next to fresh frozen tissue. Typical problems encountered include (geographical) fragmentation, undefined access rules, lack of uniform quality standards, and the absence of a uniform legal and ethical framework. This may hamper international collaboration.

An important secondary task is the validation of study results. Publicly available image data allow independent verification of study outcomes in histopathology.¹

The VUB/UZB Biobank implemented a digital pathology system for the valorisation of our own biobanks in the domains of diabetes and oncology. These

biobanks already all had a significant international profile, including involvement in investigator-driven clinical trials and collaboration with industry.^{2,3} The (digital) biobank is made available to other university hospitals and research institutions alike. A separate portal was built and is accessible to the public at www.diabetesbiobank.org.

Digital pathology

Whole slide imaging (WSI) technology allows glass slides to be scanned and digitally stored as a computer image. The image can then be shown on a computer screen via dedicated software.

One hundred years on from the original ideas for slide scanning and image capturing, digital microscopy is now taking these earliest experimental designs to the next level. Modified optics and digital cameras are used to produce an integral image that represents the original glass slide onto a monitor display, assisted by software.

Today, essentially two scanning methods exist: tile scanning and line scanning. A tile scanner uses a fixed-area camera to capture thousands of individual image tiles and subsequently stitch a composite image. This implicates a stop-and-go nature: move the slide, capture the field of view, move the slide again, etc. Leica Biosystems' ScanScope technology (formerly belonging to Aperio; Leica Biosystems acquired Aperio in 2012) features line scanning technology: A

linear-array detector in conjunction with a microscope objective lens and other custom optics capture a small number of contiguous overlapping image stripes. The microscope-slide moves continuously and at a constant speed during the acquisition of an image stripe, and the scanner automatically adjusts the focus of the objective lens from one scan line to the next.

Other vendors each have their own variations on these basic schemes and have diverged from traditional microscopic optics to suit their needs: for example, DMetrix's DX-40 scanner contains an in-house manufactured array of 480 optical surfaces, and uses aspherical lenses instead of conventional spherical types.

Legally, many of the original underlying patents for the technology either reside with Olympus (with respect to tile-scanning) and Leica Biosystems (with respect to line-scanning).⁴

Resulting images can be large: scanning a piece of tissue of 40mm x 20mm at 20x magnification results in a resolution of 0.25 micron per pixel, or 100K x 50k pixels. When scanned at 40x magnification (0.125 micron per pixel), those numbers double to 200k x 100k pixels. The uncompressed image is then over 6GB in size (compression can bring the effective file size down significantly, depending on the complexity of the observed tissue features).

Virtual slides are typically organised as pyramid stacks: multiple subsampled images at various resolutions are created to facilitate rapid zooming into a desired resolution. This allows at any given time that the appropriate level of details is shown to the end-user: when a user chooses to view the entire slide at low magnification, data is transmitted at low resolution. However when a user chooses to inspect a small portion of a slide at high magnification, that portion alone is transmitted; cropped from the appropriate level in the stack, and at the proper high resolution. The applications for these datasets today are (mostly) in (digital) pathology and image analysis (bioimaging informatics).

Digital pathology is an application of whole slide imaging. Digital pathology can help find solutions to make old scientific research available to the public by

publishing a catalogue of historical slide collections. Figure 1 shows a custom built portal for our digital pathology-related activities at Brussels Free University.

Why digitise a biobank?

Funders of biobanks are looking to obtain the most value from their investments in sample and data collection. There is, therefore, a need to increase the availability to researchers of large numbers of high quality, well-annotated samples. Despite the development of strict protocols for the inclusion of material into biorepositories, specimens in a biobank may still prove to be of little value for downstream testing. In one large study⁵ investigating 1138 samples from the University of Indiana tissue bank, only 59% were found to be at least 65% tumour versus non-neoplastic tissue. Meanwhile, 23% had a tumour volume that accounted for < 65% of the gross specimen, 17% were entirely negative for tumour, and 1% was completely necrotic. These findings underscore the importance of instituting adequate measures for histological sample quality control before the release of banked samples for downstream testing. Biobanks follow the protocols for material to be included, yet it is the researcher's responsibility to preview the samples digitally to confirm they are adequate.

The availability of an online database of whole slide images for specimens in a biobank then makes it possible for researchers to preselect specimens based on tissue composition.

Key advantages of this are:

- Researchers interested in retrieving material from the collection can visually inspect a sample first and find out for themselves if the sample really contains the material that they are looking for.
- Pathologists and others can now gather customised collections of virtual slide material for studying specific phenotypes (possibly even from different (virtual) tissue banks around the globe). No strong background in computer sciences or IT is needed (everything can be visualised within a web browser interface); the interested party need only worry about selecting the material and asking the right questions. When no physical material is needed, 'virtual' studies on imaging material alone may be conducted.
- Computationally oriented scientists can also bypass the (physical) retrieval of material from the biobank. These people that know how to perform Delauney triangulations based on nuclear cell counts, convert those structures into graphs and subsequently do network topology-

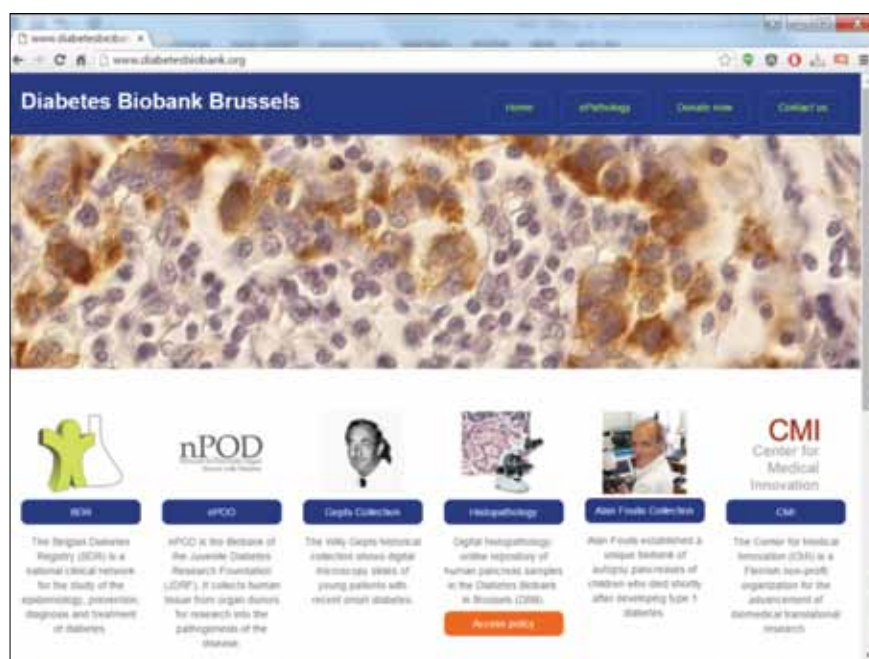


Figure 1. Customised digital pathology portal at Brussels Free University.

studies in software such as Cytoscape.⁶

Apart from opening the collection to a new sort of audience, this also permits the same material to be consulted again and again. Retrieval restrictions can be somewhat relaxed, too, because (in contrast to the physical patient samples), the digital data by their very nature are reusable by many groups around the world. This has practical implications, too: access policies for the physical and virtual material may now vary and be laid out differently. This follows from the observation that exclusivity of availability of the virtual material is no longer an issue to consider when one requests access.

Implemented hardware

In a first phase, an Aperio CS2 scanner from Leica Biosystems was purchased after going through a public tender process. Criteria that led to the election of the CS2 scanner included pricing and openness to be plugged into other environments (with regards to the output file format). Furthermore, the slide capacity (loader for five slides) of the CS2 suit our purposes better than other high-volume scanners. The CS2 is a line scanner (cf. supra) that outputs a variety of formats including Aperio's own (TIFF-derived) SVS, with options for either JPEG or JPEG2000 compression. JPEG2000 compression, while enabled by default, was switched back to JPEG. Computationally the specific algorithms are cubic (n^3) in nature, and 'regular' JPEG has a reduced complexity (n^2). While significant savings in storage capacity can be achieved, the performance of image data presentation degrades. A trade-off, therefore, must be made, which will result in different outcomes based on individual circumstances.

Implemented software

Integration with existing organisational processes is the key factor that distinguishes successful solutions from stand-alone software programs. In the context of this article, connectivity with the laboratory information system (LIS) for example is becoming more important. Similarly, integration between different imaging modalities (which may reside at various physical locales throughout a campus) becomes critical to allow

multi-scale experiments to be managed in an efficient manner. The goal is to provide a physician or researcher access to data related to an experiment or a patient, rather than offering access to a dataset only linked to a specific piece of equipment.

In the case of microscopy, this means that for a specific sample, it should be possible to oversee all imaging data linked to that sample, whether this originates from a brightfield or fluorescent platform. Providing this service is still challenging today. There is no universal data format for whole-slide images,⁷ so each vendor (and sometimes even piece of equipment) has its own file format. Support for non-Windows platforms can be problematic, too. Last but not least: having to download specific (viewer) software is not always possible (and usually at least experienced as annoying) for regular users.

At our institute, too, different departments were using digital imaging platforms from different vendors. Therefore, in a second phase, existing digital imaging equipment was unified into a single platform. Due to the requirement to support file formats from multiple scanning platforms, a hardware-agnostic digital pathology management software needed to be introduced. The Pathomation software platform for digital pathology added significant additional value to our microscopy digitisation efforts, including:

- Allow access to digital microscopy data from any location and for any purpose.
- Allow integrated access to any file type irrespective of its (hardware vendor) origin
- No need to install separate viewing software on the local computer
- Embed digital microscopy content in dedicated community-oriented portal websites.

Building the portal website

The Diabetes Biobank website is built using Microsoft ASP.Net technology and the C# programming language. To publish the website on the internet, the Microsoft Internet Information Server (IIS) web server is employed. Besides this, no special plugins are required to

be installed or enabled (either on the client- or server-side). The site requires user authentication, which takes place against Aperio's ImageServer web service (so no redundant user databases need to be maintained). Only read access rights to the contents of the site are required to be granted to the executing process.

Whole slide image data is fetched from a Pathomation PMA.core server via SOAP web service calls. This means that the server hosting the site has to be able to access the target PMA.core installation over HTTP. Pathomation handles the imaging data. Cases that group selected slides at the patient-level are defined and managed in the Aperio eSlide Manager software. The back-end SQLServer database that safeguards the meta-data is accessed via ADO.Net and therefore is not publicly exposed: all (meta)data retrieval operations happen server-side.

Beyond biobanking: Image analysis and automated morphometry

Image analysis tools can be used to derive objective quantification measures from digital slides. Pattern recognition and visual search tools can be used to classify specimen imagery and identify medically-significant regions of digital slides. Incorporating digital pathology into biobank quality assurance procedures, using automated pattern recognition morphometric image analysis to quantify tissue features in digital WSI of tissue sections, can minimise the variability and subjectivity associated with routine pathologic evaluations in biorepositories. Whole-slide images and pathologist-reviewed morphometric analyses can be provided to researchers to guide specimen selection.⁸

For our diabetes biobank, we envision an intelligent querying system in the future that allows interested users to ask such queries such as: "Give me pancreatic tissue from patients with recent onset diabetes and a susceptible HLA-DQ genotype. At least 10% of the islets still have to contain insulinitis".

While not quite at this level yet, several groups have already expressed interest in using our infrastructure for image analysis. A key problem with using whole slide imaging data for image analysis software is to make sure that the

original data can be processed by the analytical program of choice. Because all groups at our Institute use the open-source software ImageJ for image analysis, Pathomation's HistoJ plugin is therefore used to once again bring any type of imaging data into the same environment. The resulting comprehensive digital pathology infrastructure (used for both biobanking and image analysis) at our institute is shown in Figure 2.

In order to further promote and explore the development of such systems, a workshop on 'Digital Pathology meets Bioinformatics' will be held for the first time in Den Hague on 4 September 2016, in the context of the European Conference on Computation Biology.

This workshop wants to facilitate bridging opportunities between whole slide imaging (digital pathology) and the bioinformatics community. Examples of topics include machine learning, computer vision, and software development methodologies to ease the exploitation of large images, for example, the recognition and quantification of cells, tissues and other morphological 'objects' in large-scale microscopic imaging datasets. In addition, we welcome work that takes these virtual extraction procedures to another level, for example, by analysing the network patterns that emerge from Delaunay triangulation of such extracted features. More information can be found via the organisation's website at www.eccb2016.org/programme/workshops-tutorials/.

Conclusions

Thanks to high-performance digital slide scanners such as the Aperio CS2 from Leica Biosystems, and flexible software solutions such as the Pathomation software platform for digital pathology, fully digital microscopy and pathology workflows pathology departments are now within reach. A point has now been reached where the flexibility of digital pathology hardware and software allows institutions to create their own bespoke solution to meet their needs, rather than being tied completely to a single vendor.

Whole slide imaging is gaining in popularity with applications in education and pathology reviews.⁹ More labs will have access these systems and will make whole

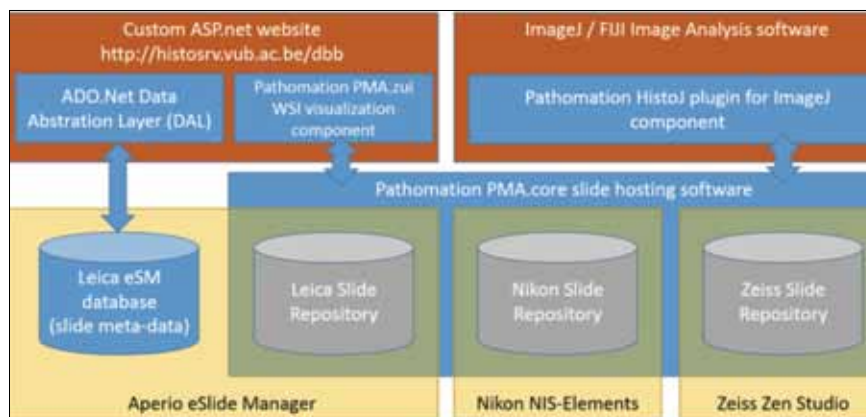


Fig 2: A schematic overview of the resulting digital pathology infrastructure after a few rounds of iterative development, implementation testing. A central role is played by Leica Biosystems' Aperio eSlide Manager, which houses brightfield whole slide imaging data, as well as slide metadata. Different fluorescent data silos have been added to this and all imaging data can now be accessed via Pathomation, a universal digital microscopy platform that lets stakeholders access imaging data in the same manner, irrespective of its origin.

slide image repositories as add-ons of existing biorepositories. From a regulatory point of view, this may even become mandatory in a not-so-distant future.

ISO/TC 276 Biotechnology was launched in 2013 to address standardisation in the field of biotechnology processes. Its topics of focus include biobanks and bioresources, analytical methods, data processing including annotation, analysis, validation, comparability and integration, as well as metrology. While no formal documents have yet to be made available to the public, it is believed that digital pathology can play an important part in this development.

We expect whole slide image database repositories that accompany biorepositories to quickly spread and become a new 'standard of care' for biobanks. ♦

Acknowledgements

Thanks to Peter In't Veld (Brussels Free University), Silke Smeets (Brussels Free University), and Gráinne Moroney (Leica Biosystems) for reviewing this manuscript, and InnoViris for financial support.

References

1. Bonner-Weir S, In't Veld PA, Weir GC. Reanalysis of study of pancreatic effects of incretin therapy: methodological deficiencies. *Diabetes Obes Metab* 2014;16(7):661–6.
2. Gorus FK et al. Predictors of progression to Type 1 diabetes: preparing for immune interventions in the preclinical disease phase. *Expert Rev Clin Immunol* 2013;9(12):1173–83.
3. In't Veld P. Insulinitis in human type 1 diabetes: a comparison between patients and animal models. *Semin Immunopathol* 2014;36(5):569–79.
4. Cucoranu IC et al. Digital pathology: A systematic evaluation of the patent landscape. *J Pathol Inform* 2014;5(1):16.
5. Sandusky G, Dumauval C, Cheng L. Review paper: Human tissues for discovery biomarker pharmaceutical research: the experience of the Indiana University Simon Cancer Center-Lilly Research Labs Tissue/Fluid BioBank. *Vet Pathol* 2009;46(1):2–9.
6. Cytoscape. www.cytoscape.org.
7. Singh R et al. Standardization in digital pathology: Supplement 145 of the DICOM standards. *J Pathol Inform* 2011;2:23.
8. Webster JD et al. Quantifying histological features of cancer biospecimens for biobanking quality assurance using automated morphometric pattern recognition image analysis algorithms. *J Biomol Tech*. 2011 Sep;22(3)
9. Sucaet Y, Waelput W. *Digital Pathology* (SpringerBriefs in Computer Science): ISBN 3319087797.