Central repository of digital pathology slides to support the development of artificial intelligence tools
Digital Revolution in Pathology

~1 kB

Minor Salivary gland
Mild inflammatory infiltrate
Acinar atrophy
Suggestive Sjögren's syndrome
FOCUS score = 4

>1 GB

Automated Gleason grading of prostate cancer tissue microarrays via deep learning

Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning

Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer
Objectives of the full project
Infrastructure

Data Centre

Annotations/Meta data

Honest Broker

Portal

Format

OpenSlide

Annotations/Meta data

OpenSlide

Annotations/Meta data

OpenSlide

Annotations/Meta data
Data

Preclinical
- Toxicology
- Models

Clinical
- Trials
- Series
- Archives

2 Mio Slides

1 Mio Slides
Tools

Access

Visualisation

Augmented intelligence

open source initiative

innovative medicines initiative
Regulations

Submission

Litigation
Expected impact

- Hospitals & Patients
- Digital Pathology Repository & Tools
- Pharma Companies
- Data Companies
- Health Authorities
Suggested architecture of the project

| Phase 1: Establish an honest broker and infrastructure. |
| Phase 2: Data collection, tools for access and visualisation. |
| Phase 3: Artificial intelligence models and tools for morphological data mining and assisted diagnosis. |

Temporal overlapping to ensure optimal use of the resources

<table>
<thead>
<tr>
<th>Work packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Project management, coordination, and sustainability</td>
</tr>
<tr>
<td>2 – Infrastructure and database hosting</td>
</tr>
<tr>
<td>3 – Data collection &amp; management</td>
</tr>
<tr>
<td>4 – Tools for accessing, annotating and mining digital slides</td>
</tr>
<tr>
<td>5 – Regulatory framework for digital slides and AI-based methods</td>
</tr>
</tbody>
</table>
Work package 1 – Project management, coordination, and sustainability

Industry:
- Assurance of the coherence of consortium activity
- Supervision of project management
- Project risk management
- Communication and dissemination

Applicant:
- Detailed follow-up and tracking
- Regular work package reports
- Early reports of any unexpected organisational or structural issues or delays
Work package 2 – Infrastructure and database hosting

Industry:
- Advice for the harmonisation of metadata associated with the digital slides

Applicant:
- infrastructure (data centre) to host three million digital slides and implement a database to register the corresponding files and associated metadata.
Work package 3 – Data collection & management

Industry:
- ~ 2 Mio glass or digital slides from nonclinical toxicology studies, animal models of diseases, or clinical trials
- Corresponding metadata, compliant with INHAND/ICD nomenclature, structured under the standardisation for exchange of nonclinical data (SEND) format.

Applicant:
- honest broker:
  - database, encoding mechanisms and registering of digital slides
  - removal of sensitive information
  - information security/ access rights/ encryption/ retrieval
- digital or glass slides from clinical series (1 Mio) with metadata (e.g. ICD)
- scanning of glass slides
Work package 4 – Tools for accessing, annotating and mining digital slides

Industry:
• defining the functionalities required
• guiding the development of tools to ensure implementation according to required functionalities
• testing tools and providing feedback.

Applicant:
• tools to interact with databases and managing digital slides & metadata
• end-user applications for the visualisation, annotation, and analysis of digital slides
• deep learning models for histopathology (CNNs)
**Work package 5 – Regulatory framework for digital slides and AI-based methods**

**Industry:**
- Guidance for the interaction with health authorities with respect to the qualification of digital and computational pathology in drug development.

**Applicant:**
- engage health authorities
- lead discussions for the adoption of frameworks or roadmaps:
  - digital slides as surrogate of glass slides
  - validation/qualification of AI tools
  - use of clinical archives to develop AI tools
  - regulatory context for the sharing of rare cases or published cases series
Key deliverables of the full project

- sustainable infrastructure to host a large series of digital slides
- mechanisms for managing with digital slides and associated data
- nonclinical slide collection
- clinical slide collection
- open-source, cross-platform software tools
- AI models
- engagement with regulatory authorities
- sustainability plan
Thank you