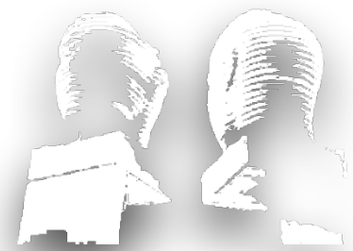
TEM data is not FAIR, how to make it play well with others?

Andy Stewart
University of Limerick

University of Limerick
Ollscoil Luimnigh

Bernal Institute

andy.stewart@ul.ie

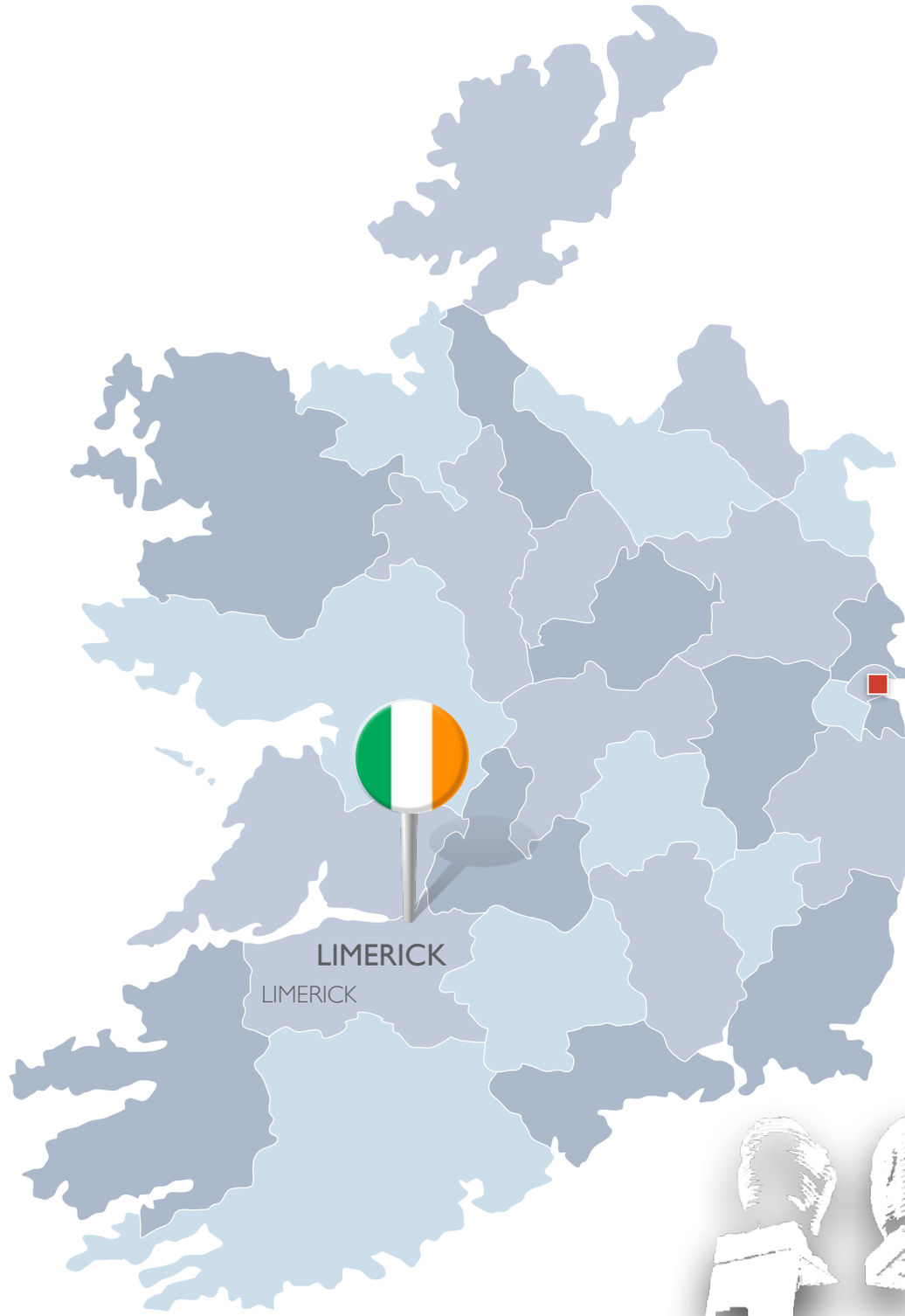


Where is Limerick?

University of Limerick
Ollscoil Luimnigh

Bernal Institute

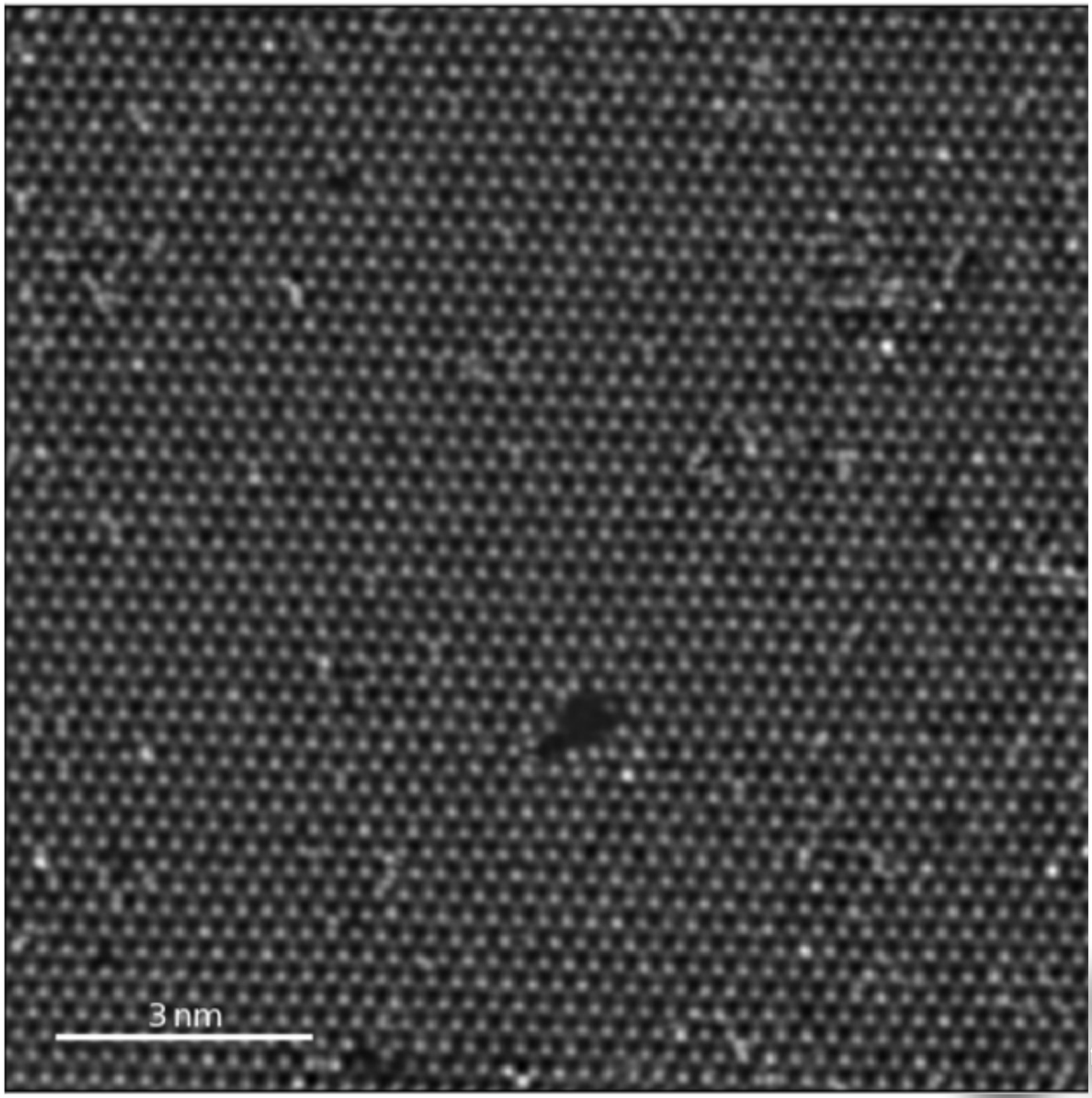
Where is Limerick?



UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute

TEM image



Courtesy of Eoghan O'Connell, University of Limerick

University of Limerick
Ollscoil Luimnigh

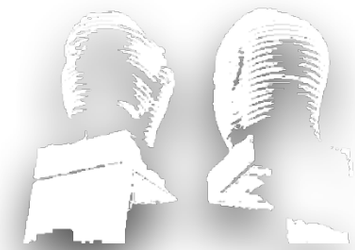
Bernal Institute

Data Processing Metadata

Data Processing Metadata

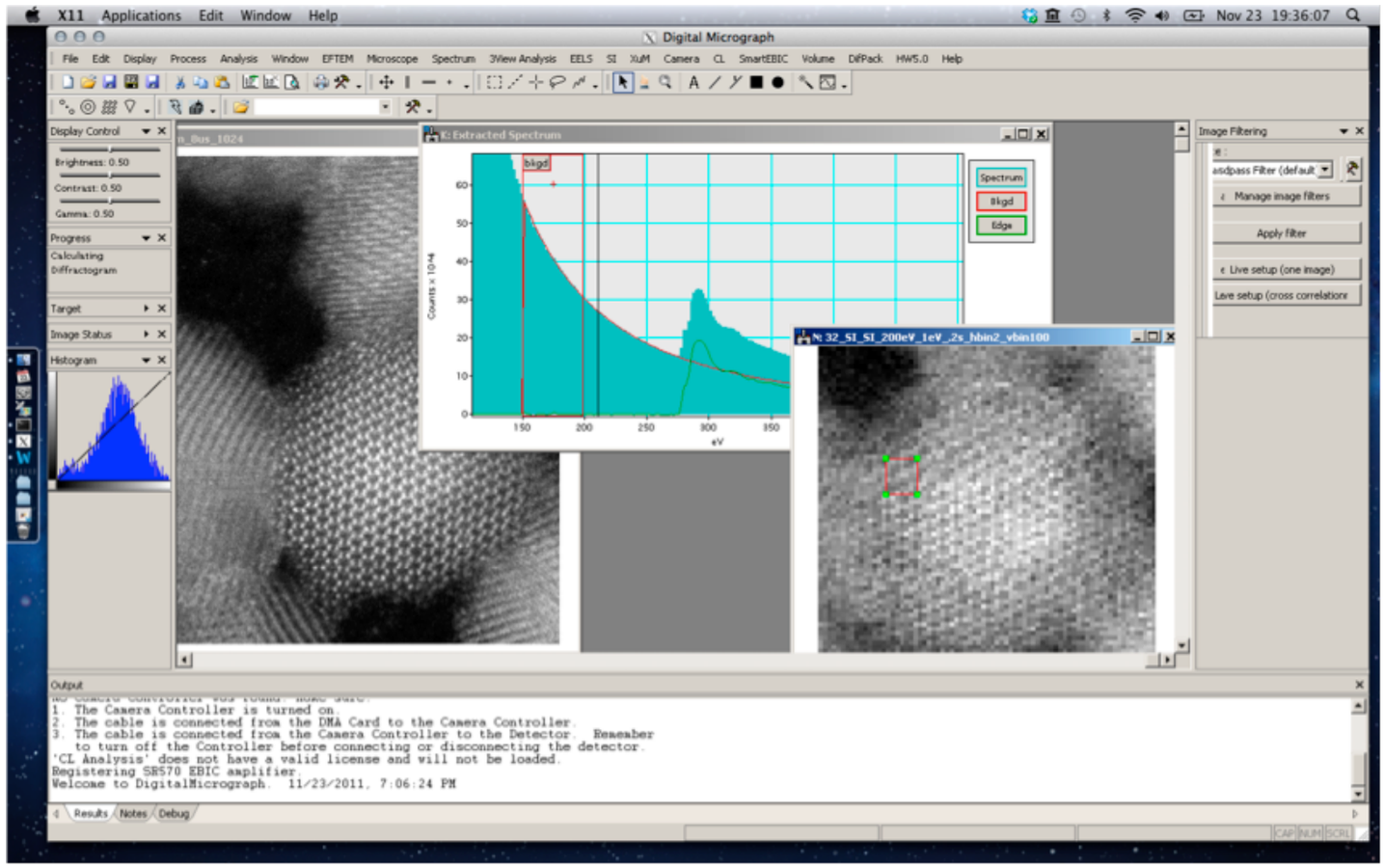
University of Limerick
Ollscoil Luimnigh

Bernal Institute



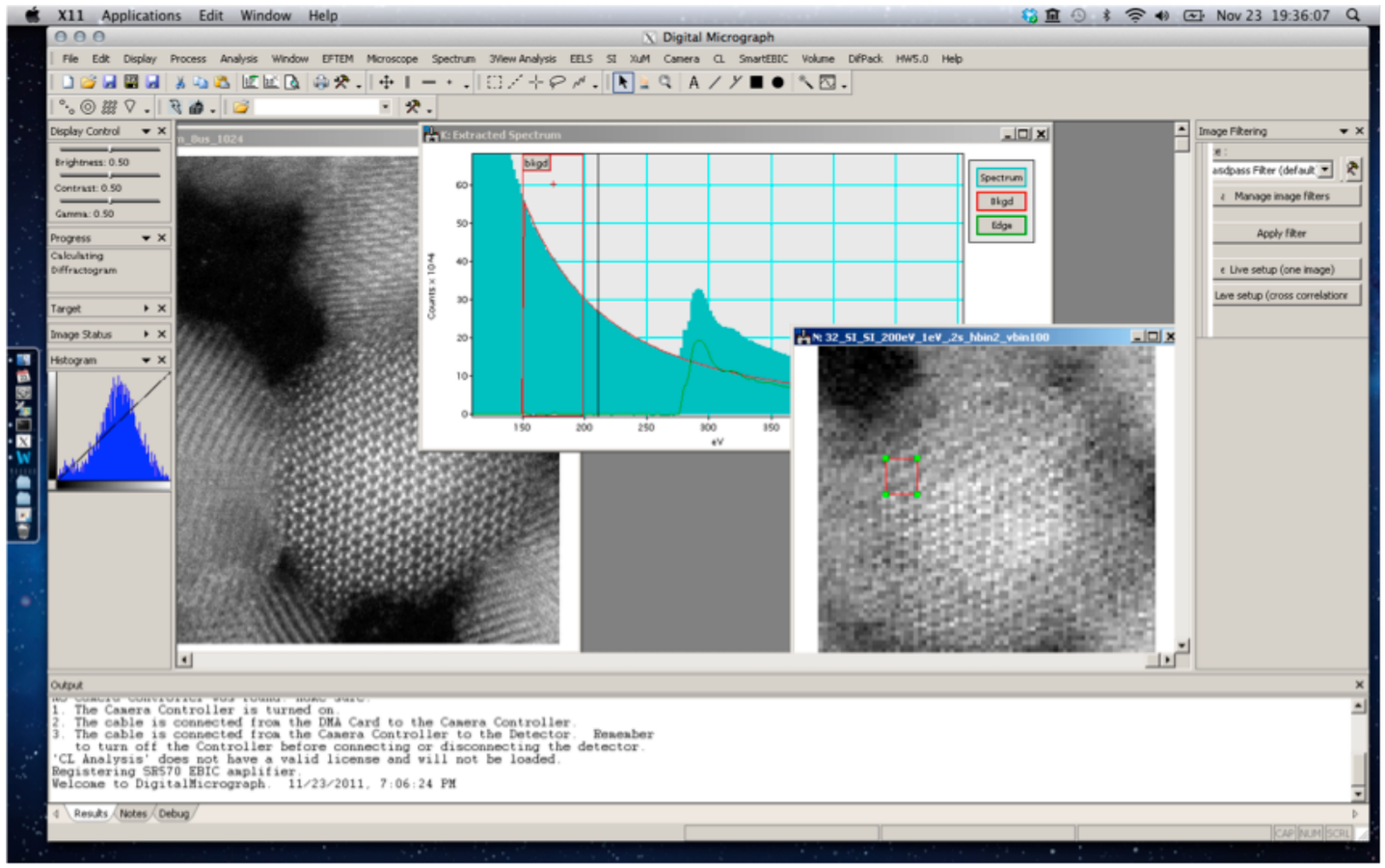
Data formats

Currently dm3/dm4



Data formats

Currently dm3/dm4



23 versions of the TIFF

University of Limerick
Ollscoil Luimnigh

Bernal Institute

TEM Metadata

- Acquisition
  - Device
    - Active Size (pixels): (4096,4096)
    - Camera Number: 2
    - CCD
      - Pixel Size (um): 15.0+i15.0
    - Configuration
      - Location: high_resolution
      - Transpose
    - Name: One View Camera
    - Source: One View Camera
    - Source ID: OneView-46565850035
    - Source Model: OneView
    - Temperature (C): -5.03906
  - Frame
    - Area
    - CCD
    - Intensity
    - Reference Images
  - Parameters
    - Acquisition Write Flags: 4294967295
    - Base Detector
    - Detector
    - Environment
    - High Level
    - Objects
    - Parameter Set Name: Record
    - Parameter Set Tag Path: Imaging:Acquire:Record
    - Version: 34078720
  - Calibration
  - DataBar
    - Acquisition Date: 12/21/2018
    - Acquisition Time: 11:31:32 AM
    - Acquisition Time (OS): 1.31899e+17
    - Binning: 2
    - Collection Number: 1
    - Device Name: One View Camera
    - Exposure Number: 2175
    - Exposure Time (s): 4.0
    - Signal Name: One View Camera
  - GMS Version
    - Created: 3.31.2169.1
    - Saved: 3.31.2169.1
  - Microscope Info
    - Actual Magnification: 1069.74
    - Cs(mm): 0.0
    - Emission Current (µA): 0.0
    - Emission Current (µA): 0.0
    - Formatted Actual Mag: 1kx
    - Formatted Indicated Mag: 580x
    - Formatted Voltage: 300kV
    - Illumination Mode: TEM
    - Illumination Sub-mode: 1
    - Imaging Mode: DIFFRACTION
    - Indicated Magnification: 580.0
    - Magnification Interpolated: false
    - Name: FEI Tecnai Remote
    - Operation Mode: DIFFRACTION
    - Operation Mode Type: 2
    - Probe Current (nA): 0.0
    - Probe Size (nm): 0.0
    - STEM Camera Length: 145.0
    - Stage Position
      - Stage Alpha: 0.00878402
      - Stage X: -90.0588
      - Stage Y: -40.6173
      - Stage Z: -35.5413
    - Voltage: 300000.0
  - Session Info
    - Items
    - Microscope:

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Sample Metadata missing?

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Sample Metadata missing?

Sample

Preparation methods and details

Specimen composition

Date and time of production

Colour

Shape

Thickness

Defects

Grain boundaries

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Sample Metadata missing?

Sample

Preparation methods and details
Specimen composition
Date and time of production
Colour
Shape
Thickness
Defects
Grain boundaries

Sample Storage

Environment
Relative humidity
Temperature
Light/Dark

Any variable which may cause sample to change over time

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Sample Metadata missing?

Sample

Preparation methods and details
Specimen composition
Date and time of production
Colour
Shape
Thickness
Defects
Grain boundaries

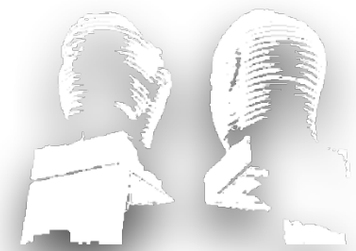
Sample Storage

Environment
Relative humidity
Temperature
Light/Dark
Any variable which may cause sample to change over time

Relationship between images and spectroscopy data

University of Limerick
Ollscoil Luimnigh

Bernal Institute



Sample Metadata missing?

Sample

Preparation methods and details
Specimen composition
Date and time of production
Colour
Shape
Thickness
Defects
Grain boundaries

Sample Storage

Environment
Relative humidity
Temperature
Light/Dark
Any variable which may cause sample to change over time

Relationship between images and spectroscopy data

Record any value/detail which will be necessary for evaluating with any other dataset or technique

University of Limerick
Ollscoil Luimnigh

Bernal Institute

TEM Missing Metadata

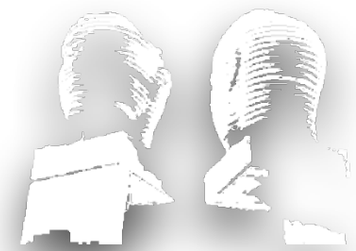
TEM Missing Metadata

<u>TEM variables</u>

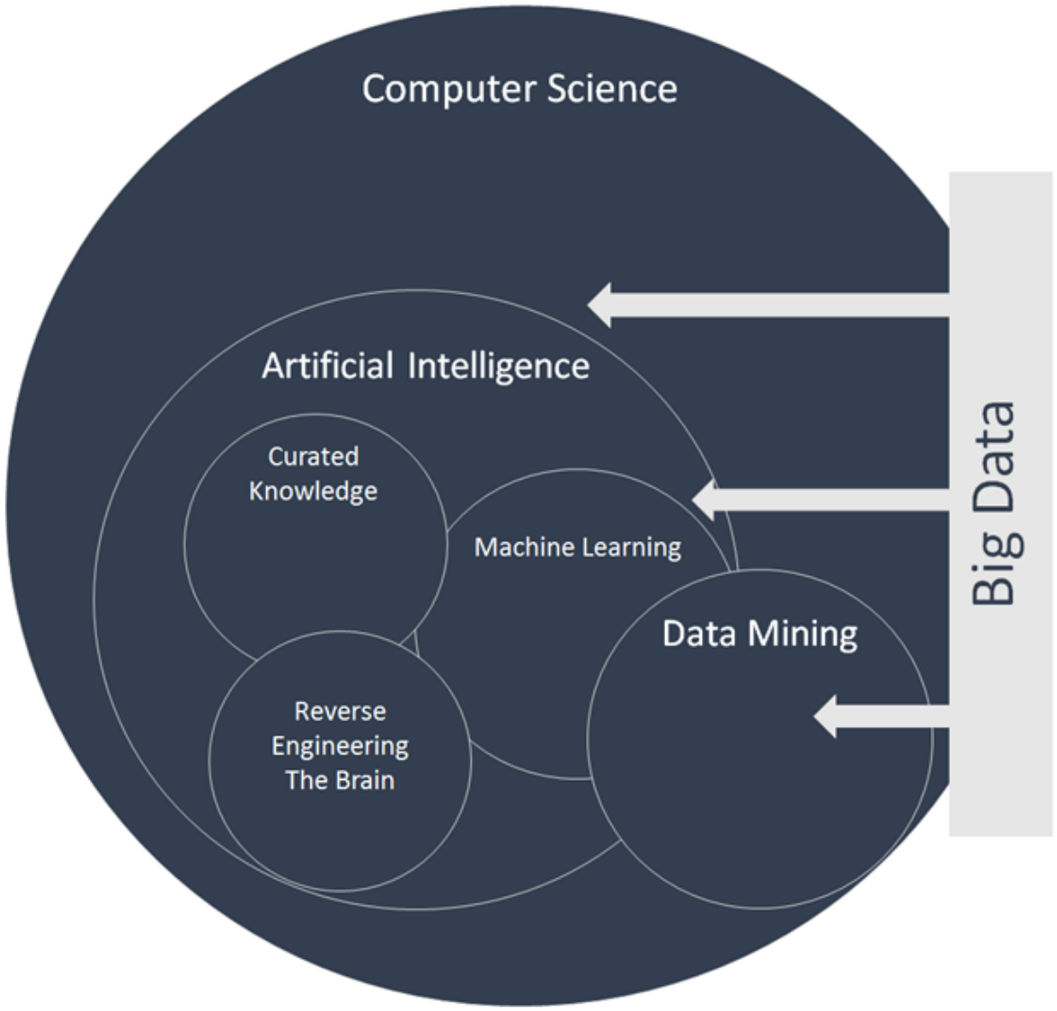
TEM Environment
Relative Humidity of room
Temperature of room
Fields in the room
Water temperature
Magnet temperature
Power Supply variability
Beam brightness
Detector temperature

University of Limerick
Ollscoil Luimnigh

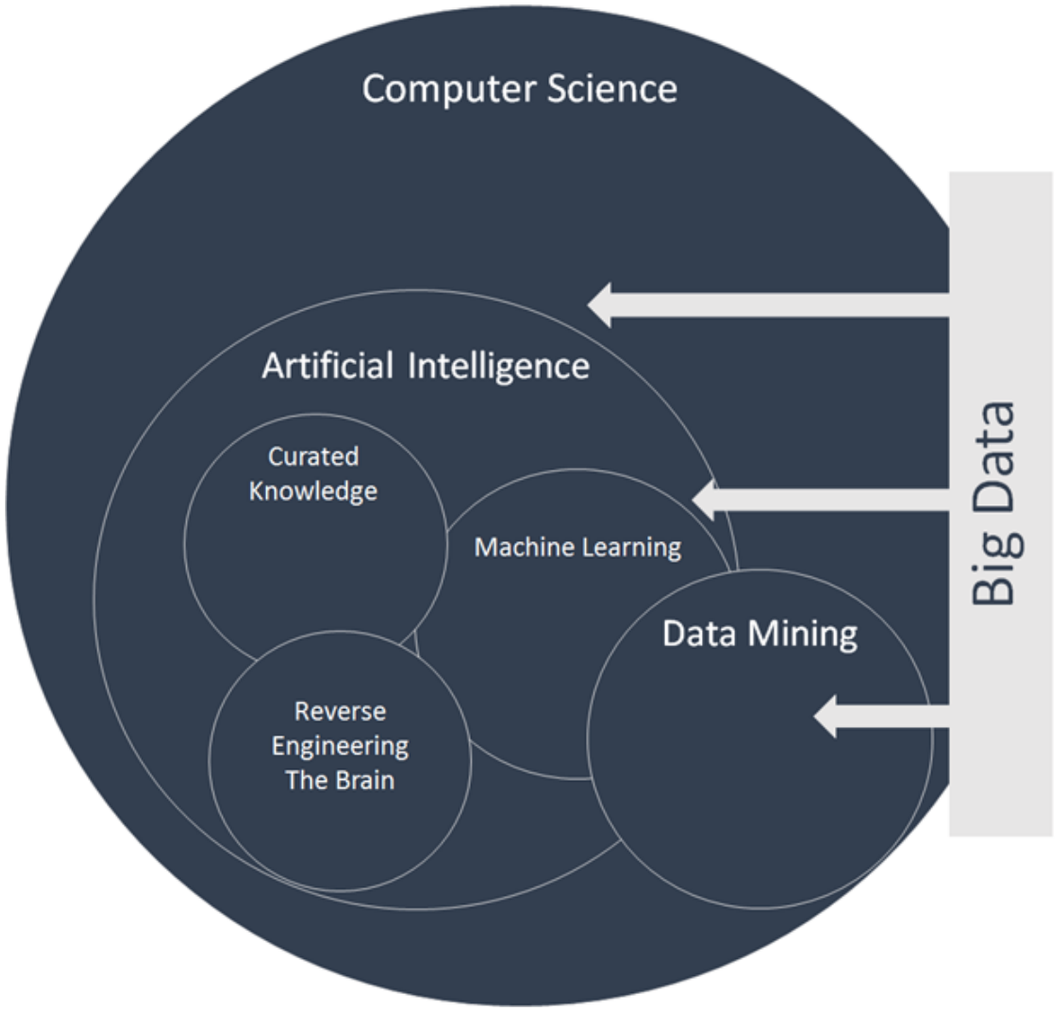
Bernal Institute



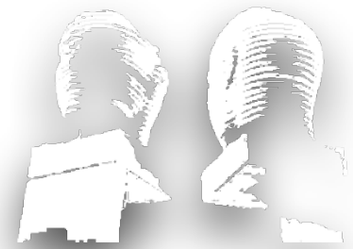
Why should we care?



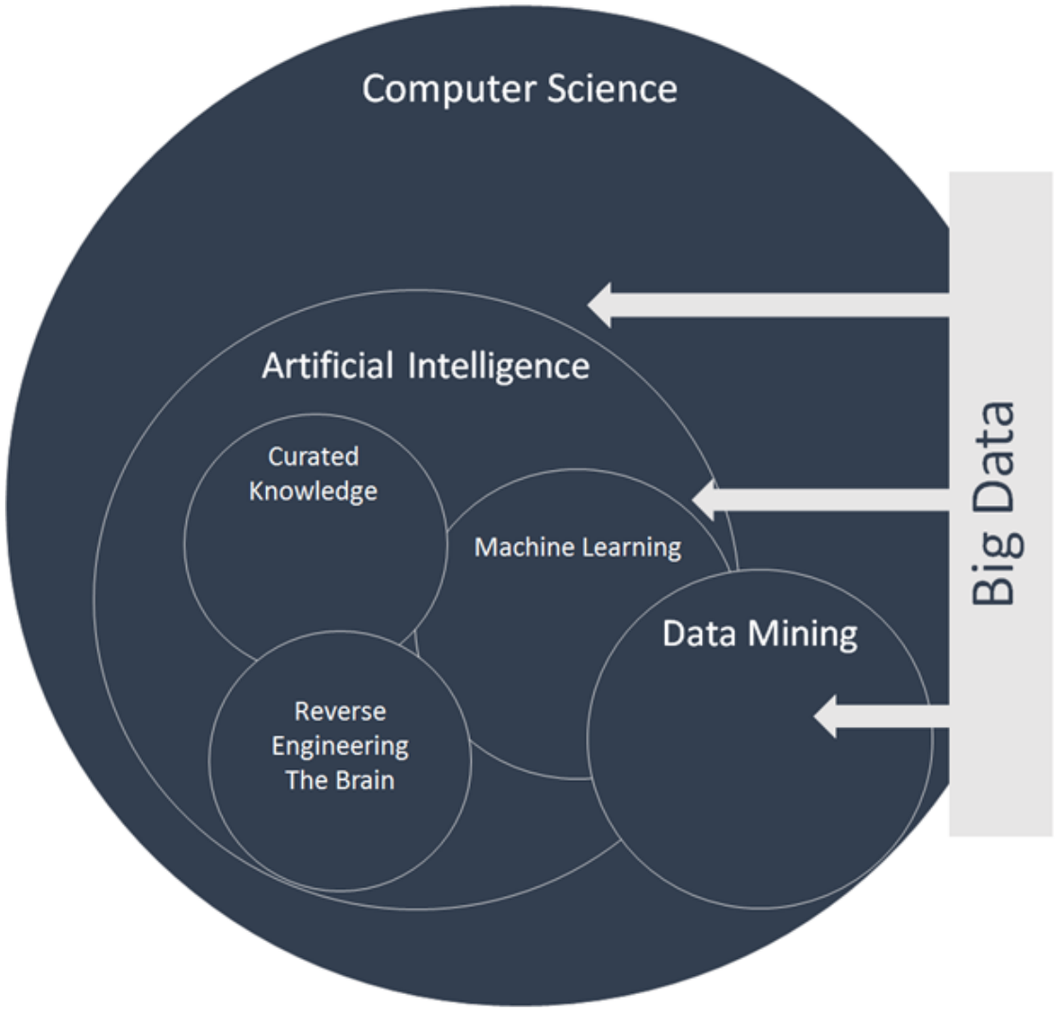
Why should we care?



Grow the need for microscopy



Why should we care?



Grow the need for microscopy

More opportunities to be cited

University of Limerick
Ollscoil Luimnigh

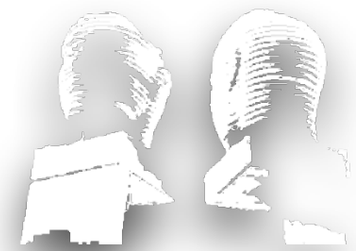
Bernal Institute

It’s NOT about you

It’s NOT about you

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute



It’s NOT about you

Wellcome Open Research

EPSRC
Engineering and Physical Sciences Research Council

European Commission
HORIZON 2020

HELMHOLTZ ASSOCIATION

Open Science

University of Limerick
Ollscoil Luimnigh

Bernal Institute

It’s NOT about you

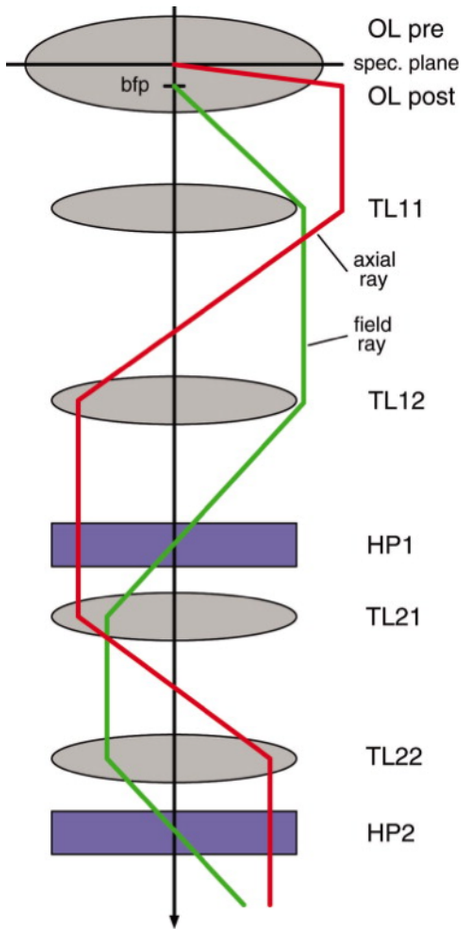
For TEM Simulations

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute

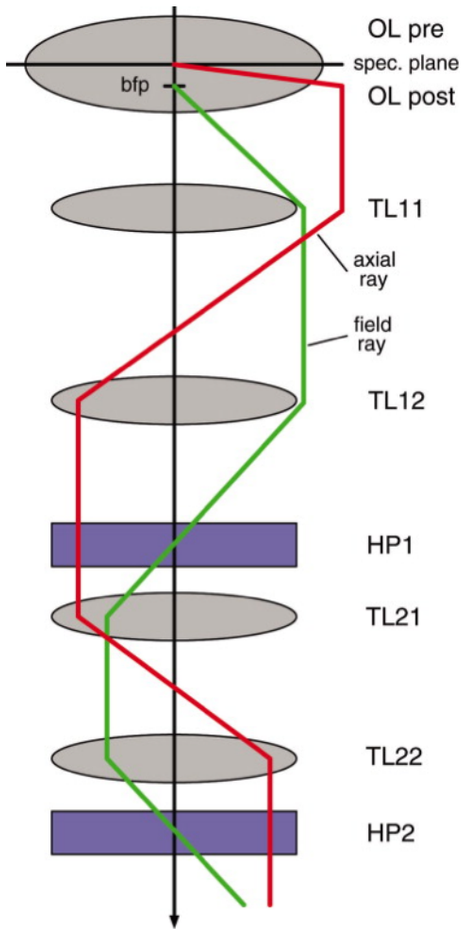
For TEM Simulations

No Corrector aberrations measurements Stored



For TEM Simulations

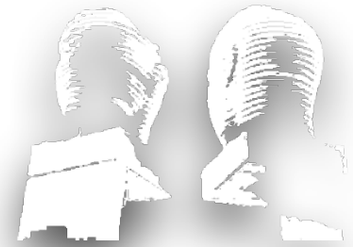
No Corrector aberrations measurements Stored



Time between aberration measurement and data acquisition

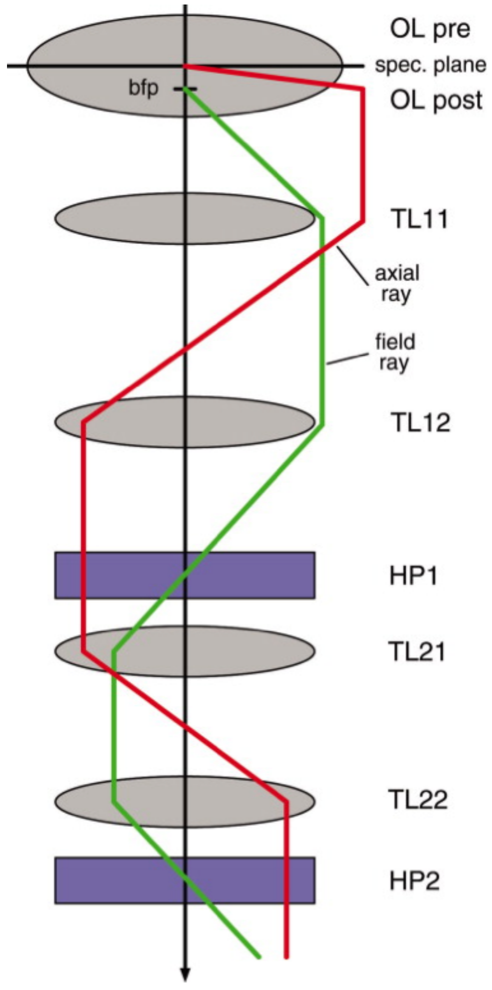
University of Limerick
Ollscoil Luimnigh

Bernal Institute



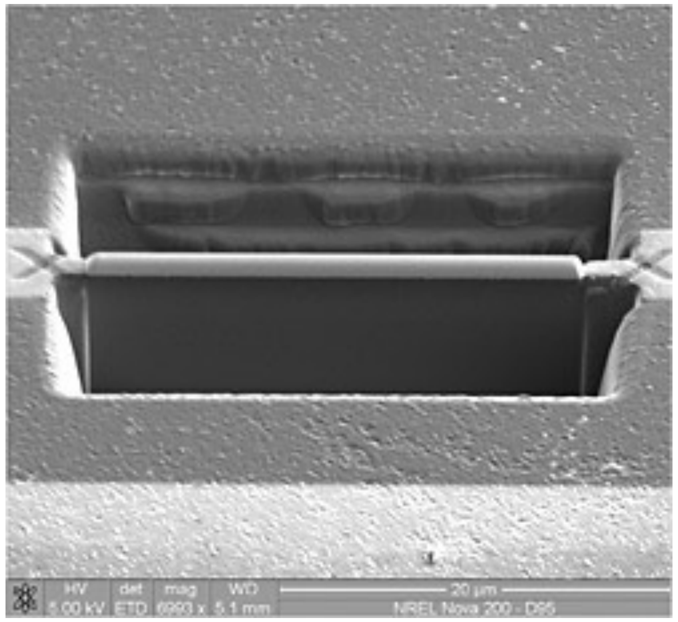
For TEM Simulations

No Corrector aberrations measurements Stored



Time between aberration measurement and data acquisition

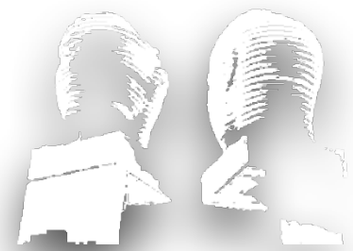
Sample thickness estimate/measurement



We are not the first and will not be the last

University of Limerick
Ollscoil Luimnigh

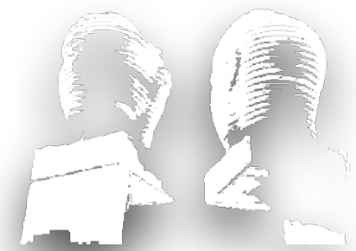
Bernal Institute



We are not the first and will not be the last

University of Limerick
Ollscoil Luimnigh

Bernal Institute

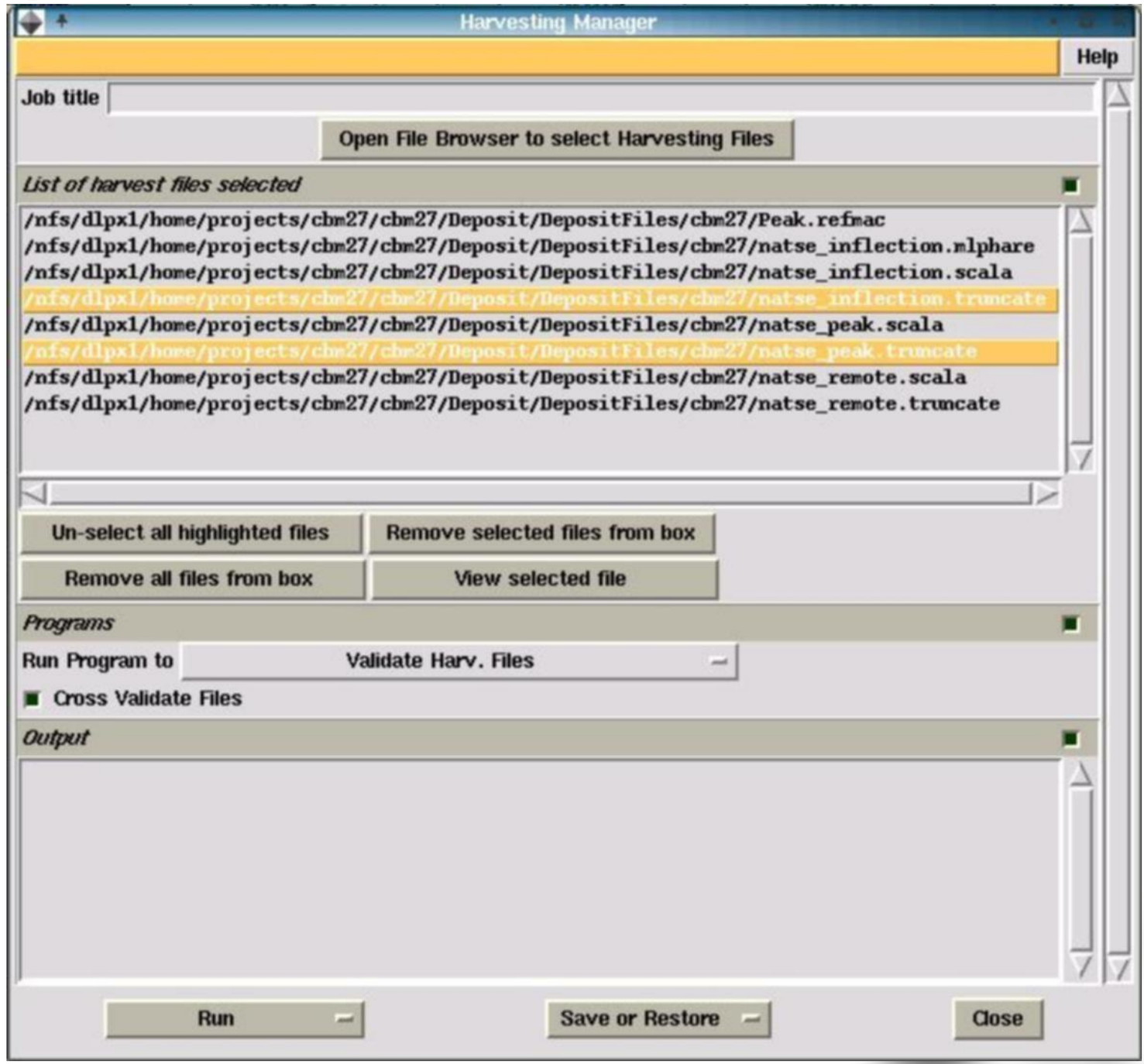


CCP4

MTZ format (1970's)

Data Harvesting (1990's)

Also Nomad

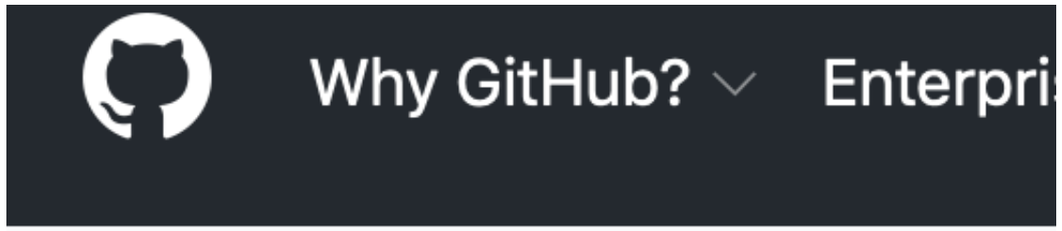


University of Limerick
Ollscoil Luimnigh

Bernal Institute

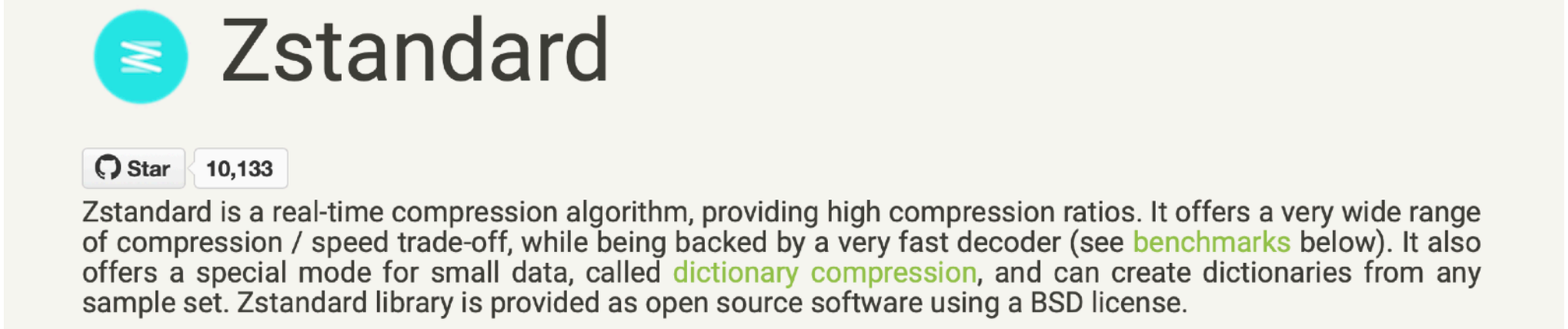
Data compression

MRCZ



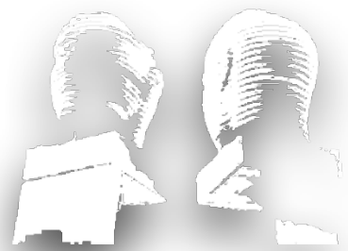
XFEL

On the file data compression



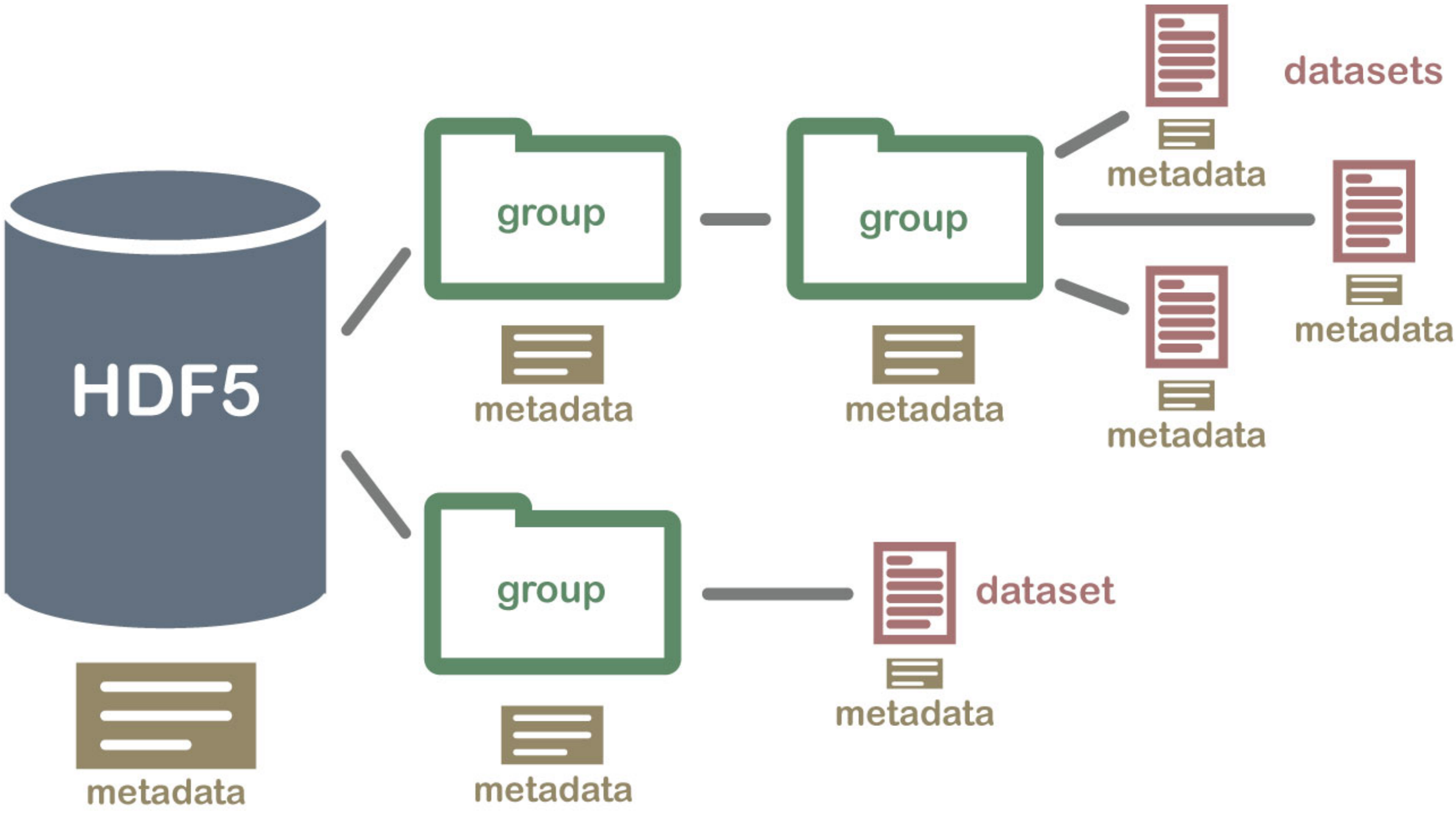
University of Limerick
Ollscoil Luimnigh

Bernal Institute



Recognised need for change Hierarchical Data Format #5

Already >25 variants in the EM community



Data format proliferation

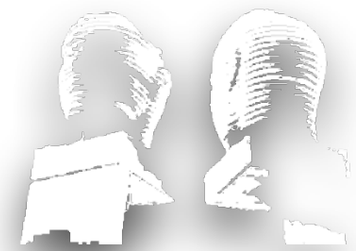
Monty Python, Life of Brian

Data format proliferation

Monty Python, Life of Brian

Adoption

Last thing we need is just another file format

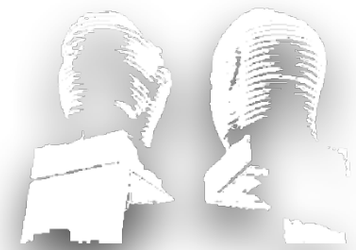


Adoption

Last thing we need is just another file format

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute



Adoption

Last thing we need is just another file format

- Public databases
- Organisations promote use of format

Adoption

Last thing we need is just another file format

- Public databases
- Organisations promote use of format

- Journals insist on data submission in a specific format
- Funding bodies insist on open data with all meta data

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Community Agreement on a single format

Or a series of convertors like Nomad?

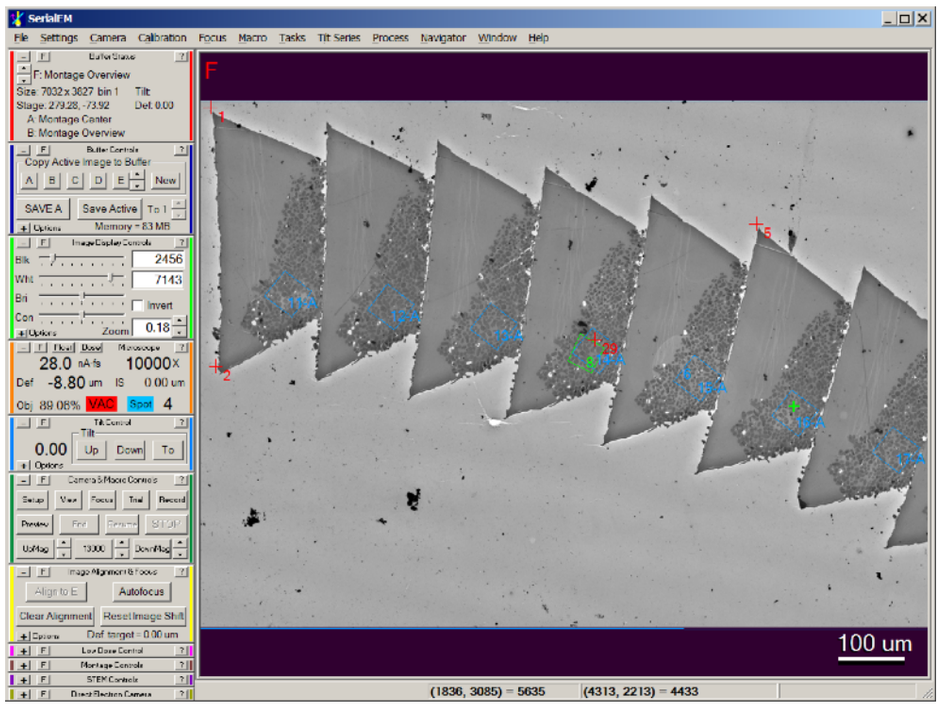
Missing metadata a problem….?

University of Limerick
Ollscoil Luimnigh

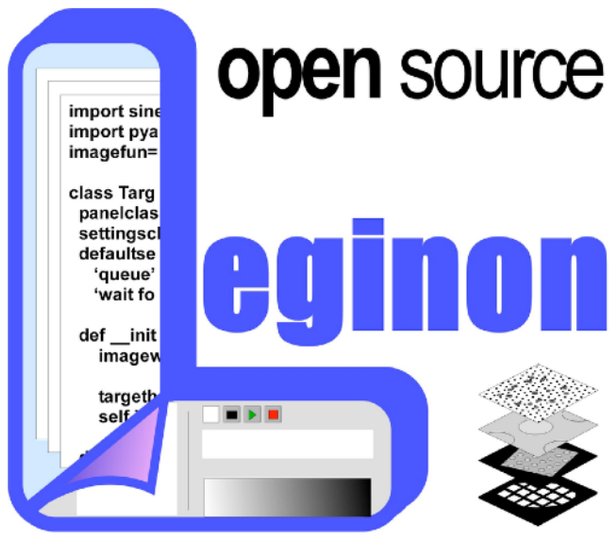
Bernal Institute

Making Metadata Easier

SerialEM

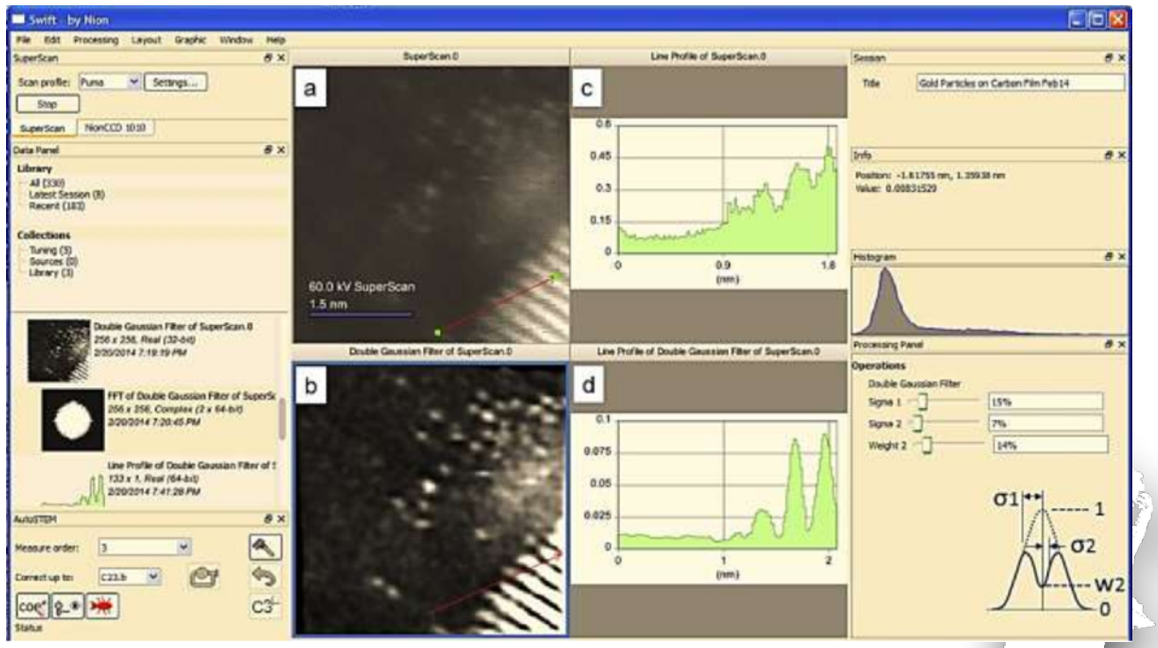


Leginon



LiberTEM

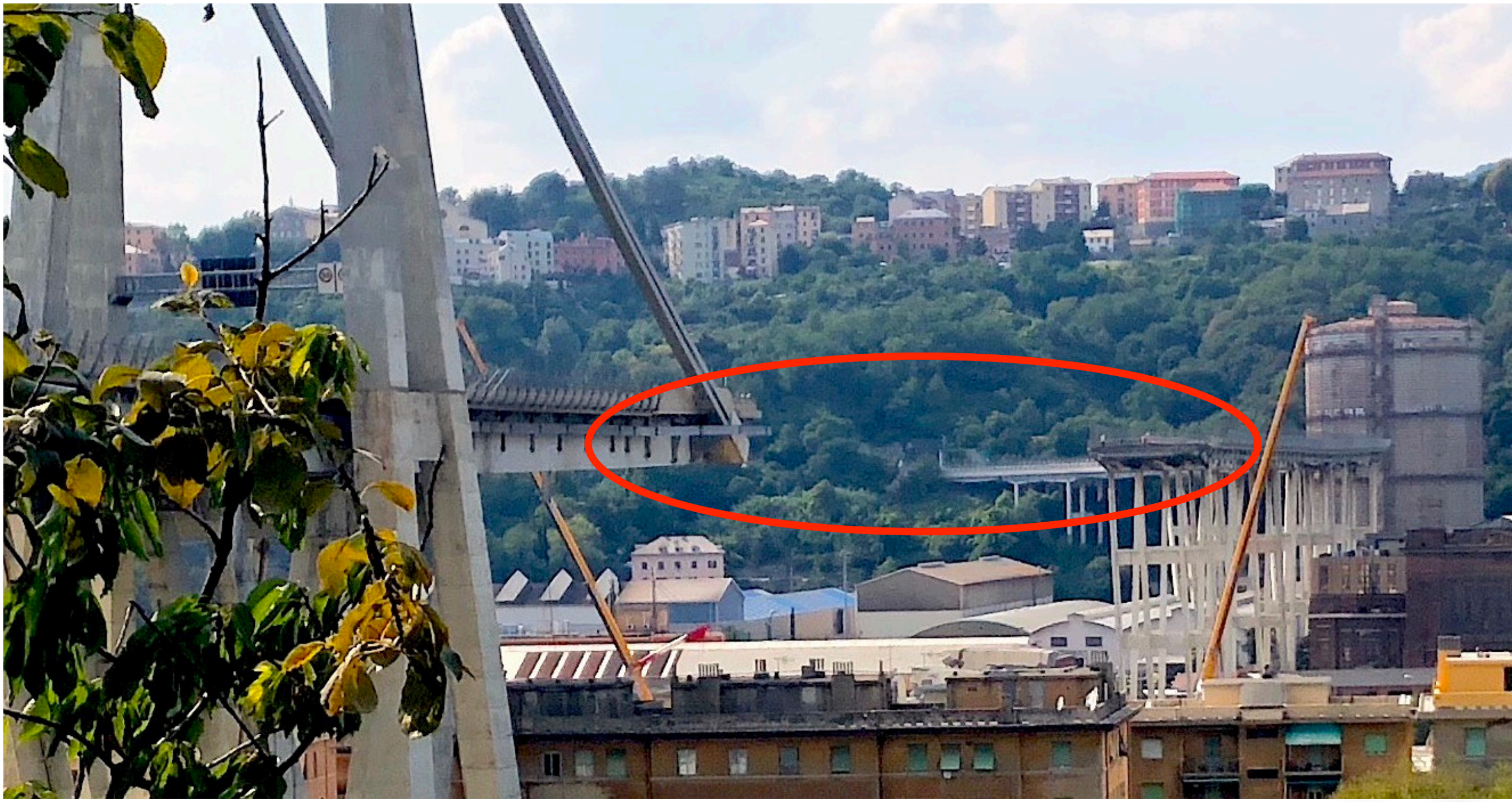
Swift Nion



University of Limerick
Ollscoil Luimnigh

Bernal Institute

Morandi Bridge Collapse

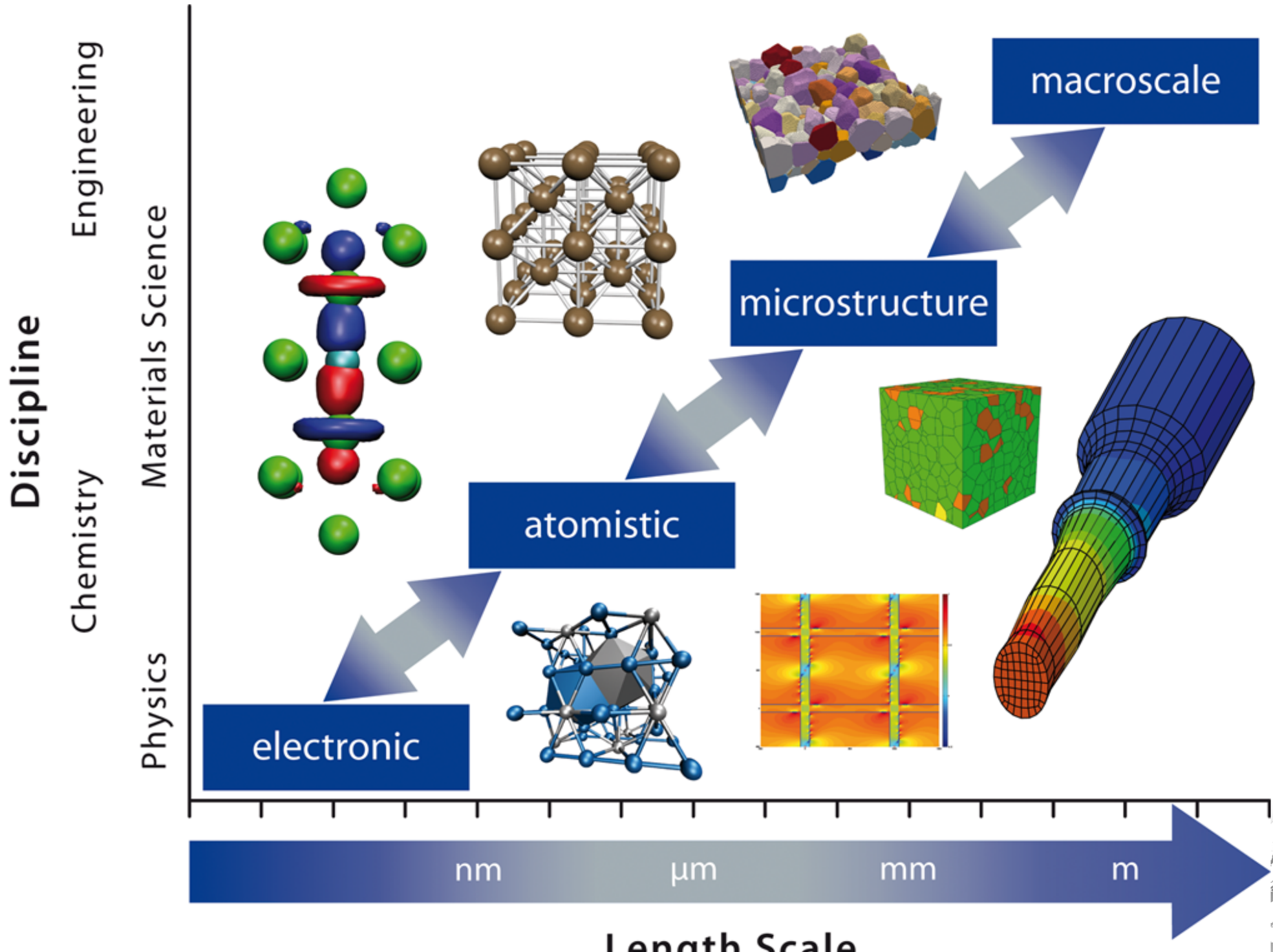


University of Limerick
Ollscoil Luimnigh

Bernal Institute

Davide Papalini

Integration across the length scales

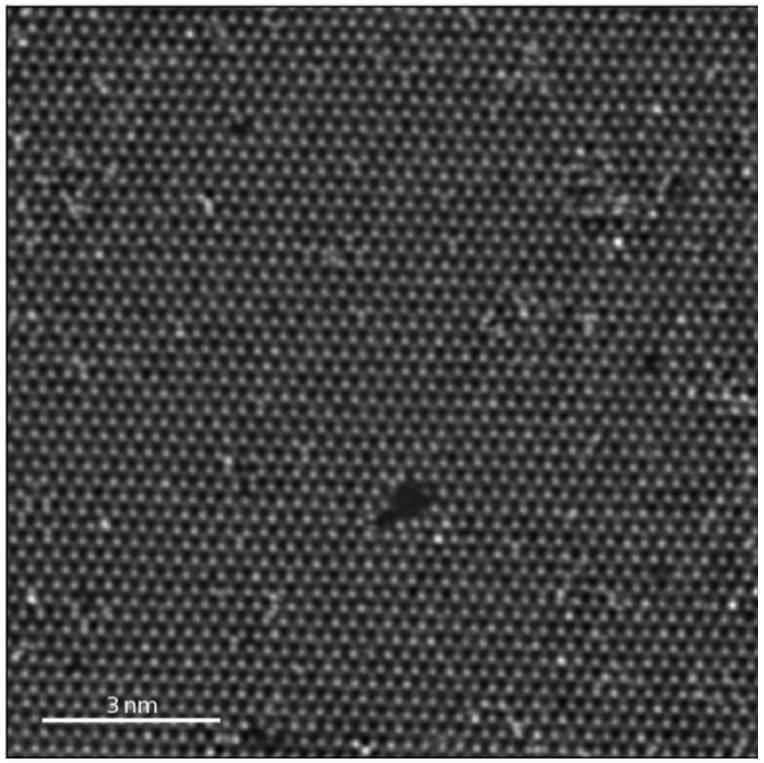


University of Limerick
Ollscoil Luimnigh

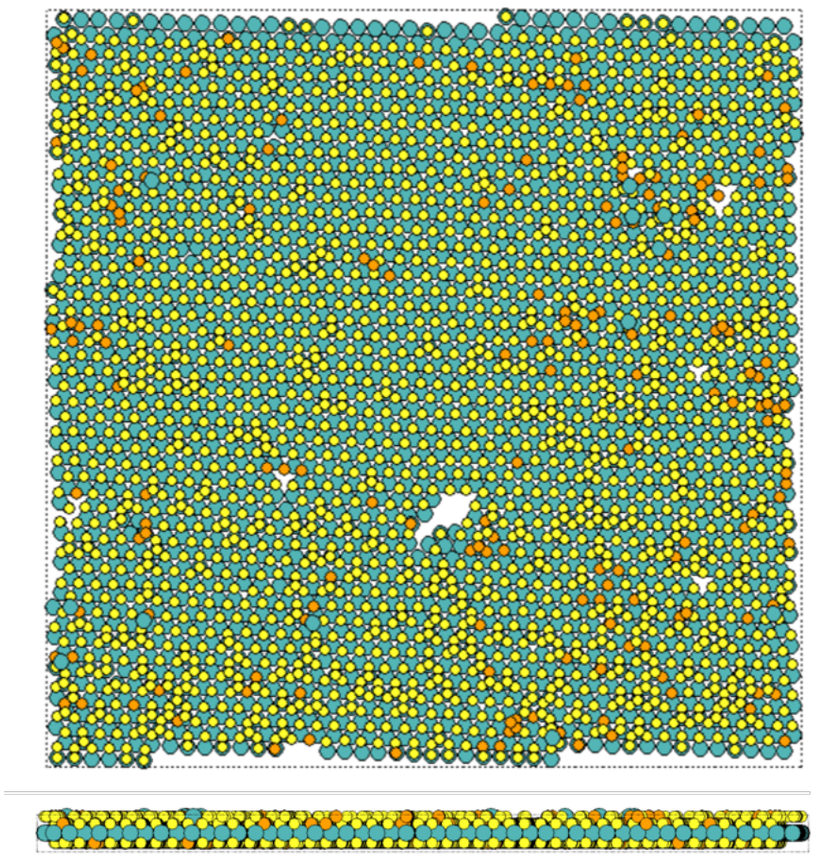
Berna
Institu

Mixing fields

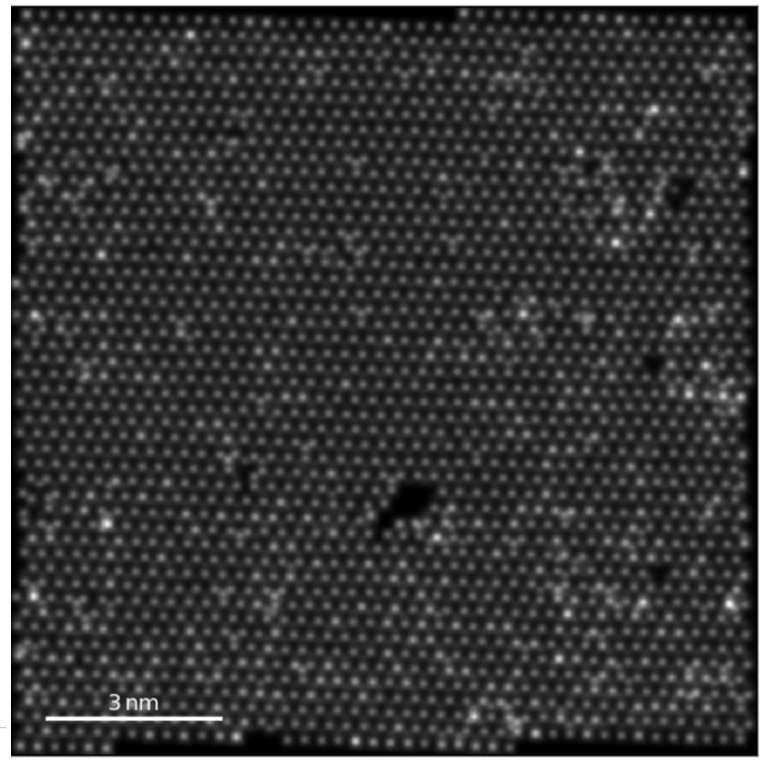
Filtered Experiment



Initial Model



Filtered Simulation



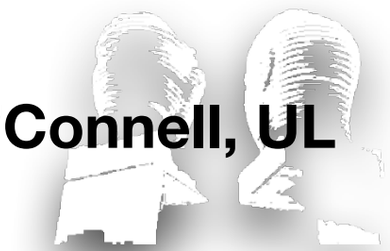
Multislice simulation

For Low Dimensional materials we can easily generate models from TEM data
Next link with simulations

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute

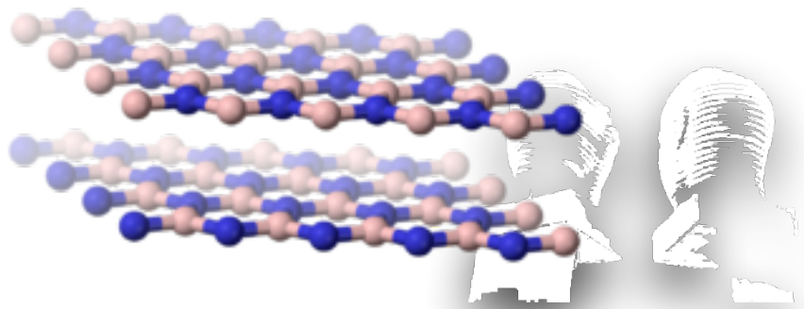
Eoghan O'Connell, UL



Proposal/Suggestion

Experimentally simulate and test a material across the length scales for a demonstration paper

Suggestion Material?
Boron Nitride
Magnesium Alloy…



Proposal/Suggestion

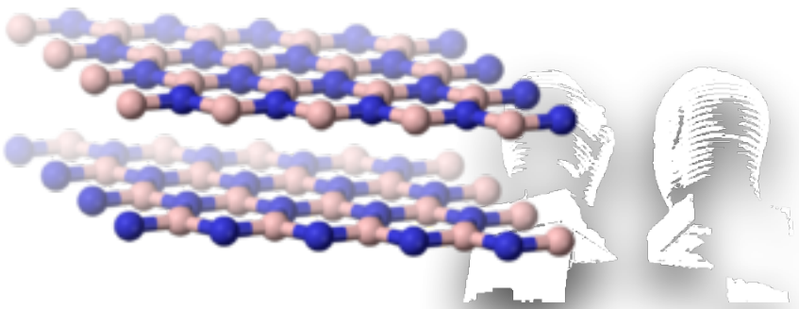
Experimentally simulate and test a material across the length scales for a demonstration paper

Additionally use the experience to define and refine the metadata required to make the exchange of data between experiments and simulation as smooth as possible

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Suggestion Material?
Boron Nitride
Magnesium Alloy…



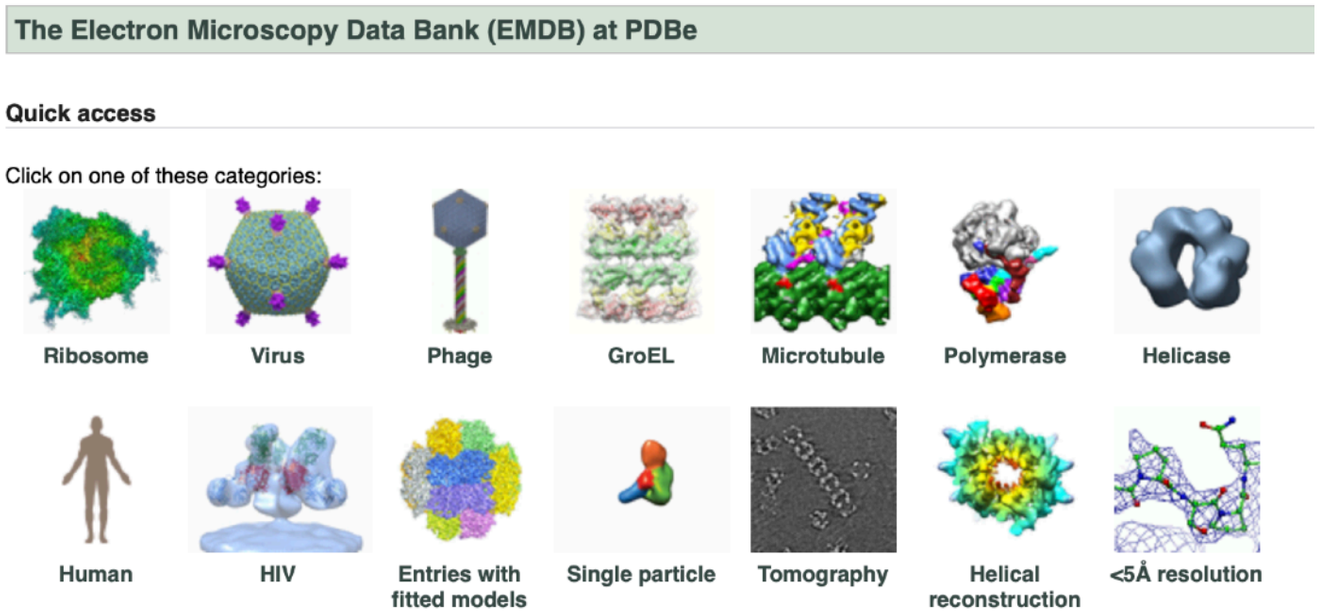
Databases

CCDC

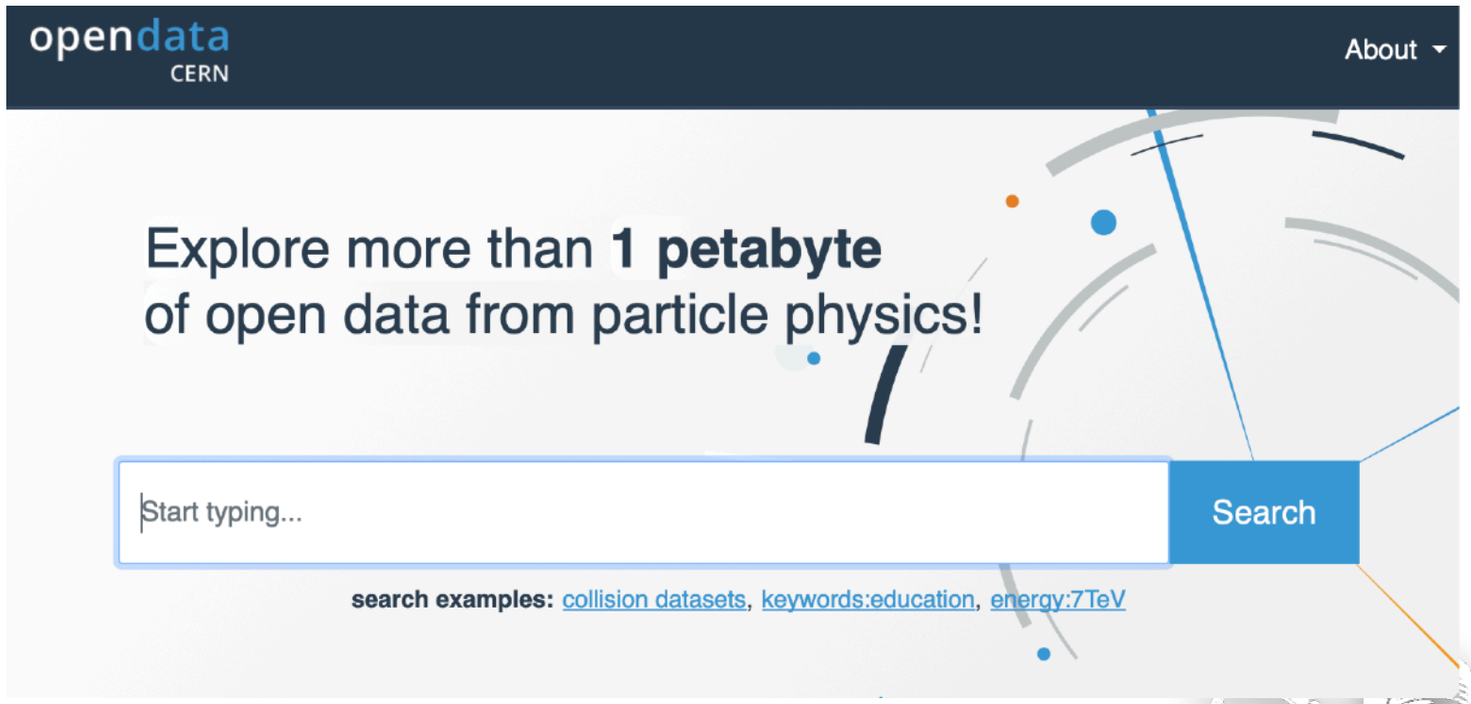
RCSB PDB PROTEIN DATA BANK

153836 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 | WORLDWIDE PDB PROTEIN DATA BANK | EMDataResource Unified Data Resource for 3DEM | ndb NUCLEIC ACID DATABASE | Worldwide Protein Data Bank Foundation



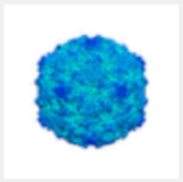
EMPIAR



University of Limerick
Ollscoil Luimnigh

Bernal Institute

EMPIAR

Dataset	Title	Authors and references	Size	Resolution	Release date
EMPIAR-10284 	Extracellular albumin and endosomal ions prime enterovirus particles for uncoating that can be prevented by fatty acid saturation [multiple data sets in MRC format]	Domanska A, Ruokolainen VP, Pelliccia M, Laajala MA, Marjomäki VS, Butcher SJ [Pubmed: 31189702] [DOI: 10.1128/JVI.00599-19] EMD-4903, EMD-0565, 6rjf, 6o06	2.4 TB	3.5-3.6	2019-06-19
EMPIAR-10281 	Cryo electron micrographs of human cystic fibrosis transmembrane conductance regulator (CFTR) in complex with GLPG [3335 multi-frame micrographs composed of 50 frames each in TIFF format]	Zhang ZZ, Liu FL, Chen JC [Pubmed: 31221859] [DOI: 10.1126/science.aaw7611] EMD-0606, 6o1v	1.4 TB	3.2	2019-06-21
EMPIAR-10270 	Single particle reconstruction of 52 kDa biotin-bound state streptavidin at 3.2 Angstrom resolution [3309 multi-frame micrographs composed of 32 frames each in MRC format]	Fan X, Wang J, Zhang X, Yang Z, Zhang JC, Zhao L, Peng HL, Lei J, Wang HW [Pubmed: 31160591] [DOI: 10.1038/s41467-019-10368-w] EMD-0689, 6j6j	5.5 TB	3.2	2019-06-07

University of Limerick
Ollscoil Luimnigh

Bernal Institute

Block Chain database

Caltech

Caltech Electron Tomography Database

A public repository featuring 11293 electron tomography datasets of intact bacterial and archaeal cells, representing 85 species.

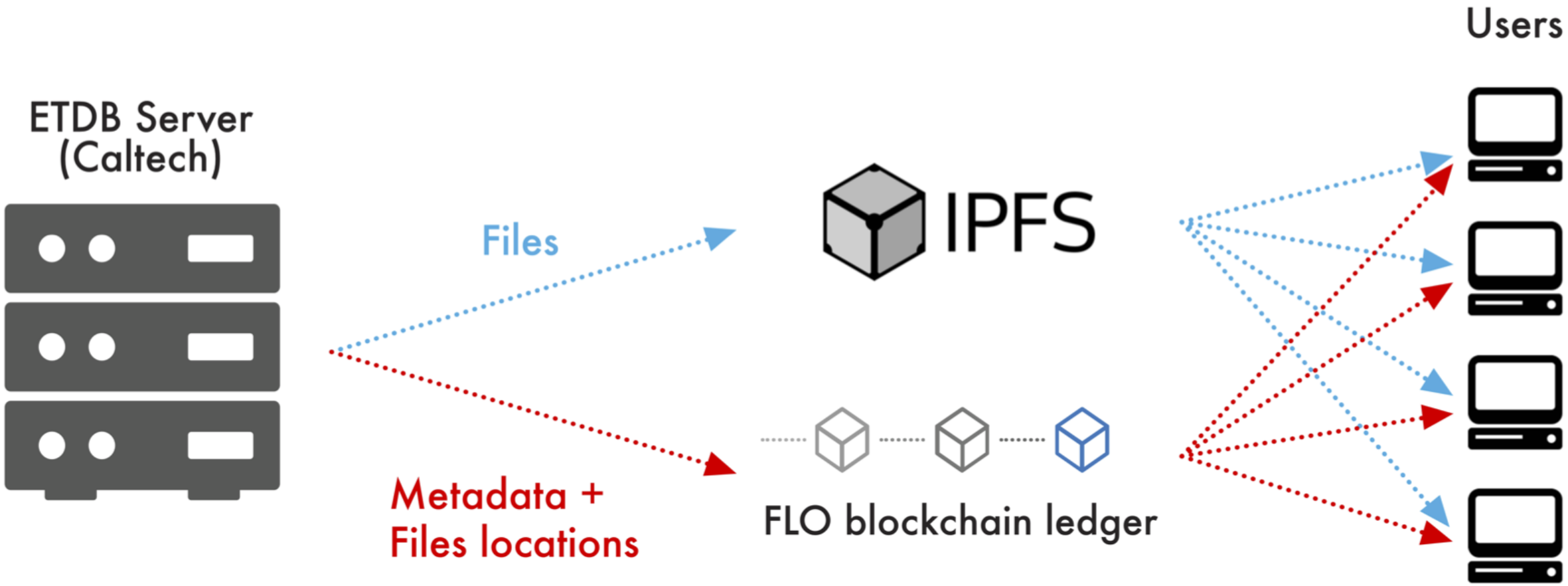
Browse database →

View our terms and conditions

>30TB of Cryo-Tomography data available, variable quality

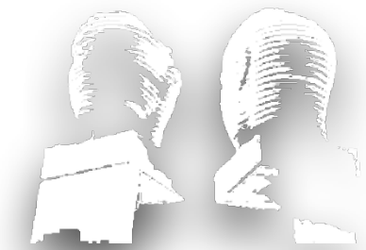
ETDB-Caltech: A blockchain-based distributed public database for electron tomography
Davi R. Ortega, et al., PLOS ONE
https://doi.org/10.1371/journal.pone.0215531

Block Chain database



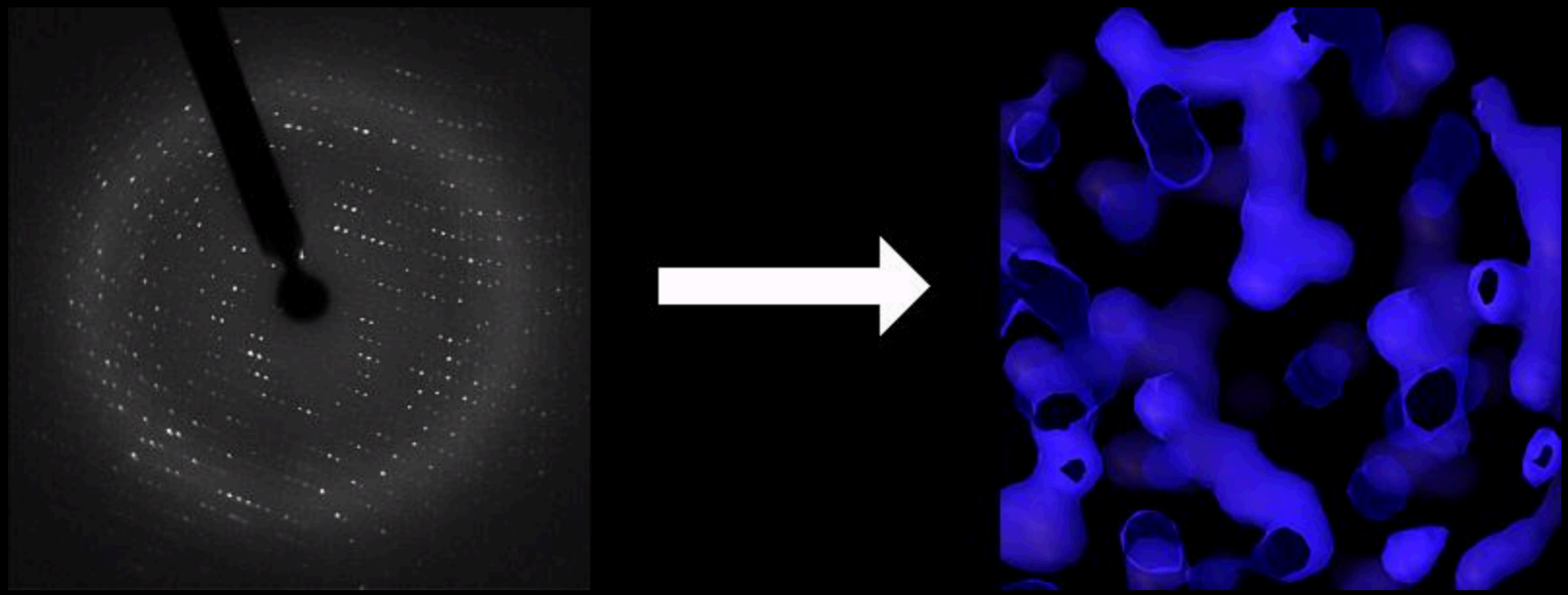
>30TB of Cryo-Tomography data available, variable quality

ETDB-Caltech: A blockchain-based distributed public database for electron tomography
Davi R. Ortega, et al., PLOS ONE
https://doi.org/10.1371/journal.pone.0215531



ELVES

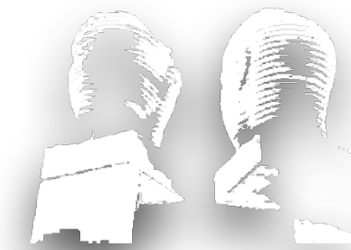
"Because You Have Better Things To Do"™



University of Limerick
Ollscoil Luimnigh

Bernal Institute

James Holten



Standard Operating Procedures

COMPLIANCE POLICY GUIDE (CPG)

Guidance for Industry, Q7A Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients

University of Limerick
Ollscoil Luimnigh

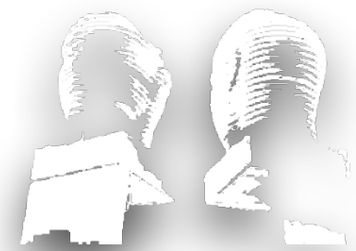
Bernal Institute

Food and Drug Administration

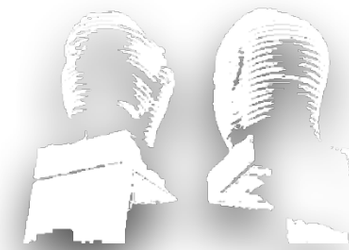
Governance of standard

University of Limerick
Ollscoil Luimnigh

Bernal Institute



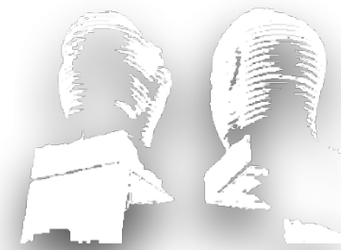
Governance of standard



Governance of standard

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute



Governance of standard

International Metrology Institutes

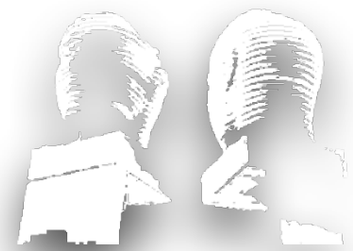
European Molecular Biology Laboratory

EMBL

European Molecular
Biology Laboratory

University of Limerick
Ollscoil Luimnigh

Bernal Institute



European Molecular Biology Laboratory

EMBL

European Molecular
Biology Laboratory

EM L

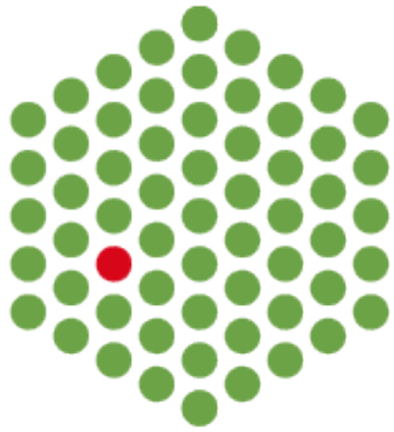
European Materials
Laboratory

UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Bernal Institute

European Molecular Biology Laboratory

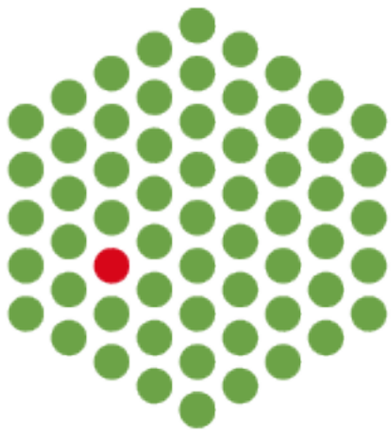
EMBL



European Molecular
Biology Laboratory

CERN

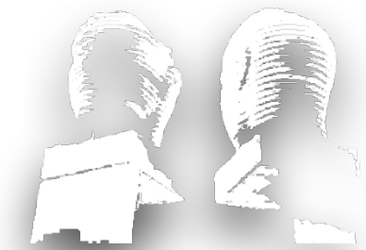
EM L



European Materials
Laboratory

University of Limerick
Ollscoil Luimnigh

Bernal Institute



Thank you for listening

University of Limerick
Ollscoil Luimnigh

Bernal Institute