Instruct: Innovation and Open Data

Ray Owens

(Instruct-UK & University of Oxford)

Corbel Meeting on Innovation by Open Access: Brussels, 20th June 2017

Instruct: a distributed Infrastructure for Integrated Structural Biology

Instruct is a consortium of 17 centres across Europe supporting 4988 users from 12 European countries.

Instruct enables European users to access cutting edge technologies with excellent scientific and technical guidance.

Instruct will become an European Research Infrastructure Consortium (ERIC) by 1st July 2017.
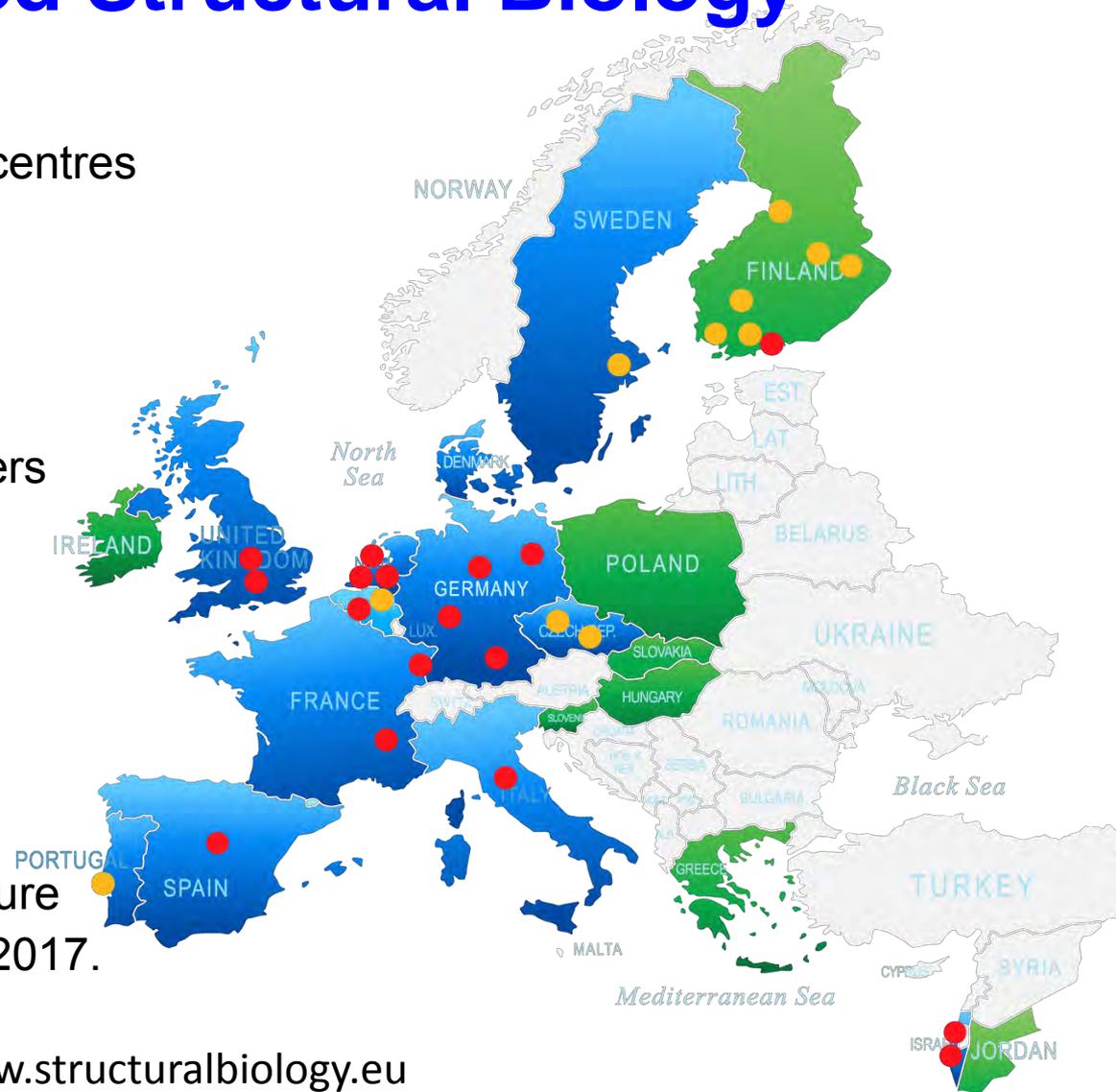
www.structuralbiology.eu

What is innovation ?

"Being innovative is not necessarily inventing; innovation refers to renewing, changing or **creating more effective processes,** products or ways of doing things. It does not only refer to partnering with industry or providing industry with a service but rather providing the opportunity to change a business model and adapt to changes in the environment to deliver better products or services. Innovation is the key to competitive advantage for any business. For RIs of pan-European or global benefit, this premise could mean **implementing new ideas**, creating dynamic products or **improving existing services**. Innovation can be a catalyst for growth and sustainability in an ever-changing scientific and technological global environment and so play a key role in **ensuring the long-term value of research facilities** to the European academic and industrial research communities".

ESFRI

Working Group on Innovation Report to ESFRI, March 2016

Instruct activities

- **Publications:** a total of 453 peer review publications acknowledge Instruct.

- **R & D awards:** A total of 35 awards from 4 calls: awardees have up to 1 year to use award of up to €20K.

- **Training**: 43 courses (Instruct commissioned and funded) and 21 internships (3-5 months) in an Instruct Centre.

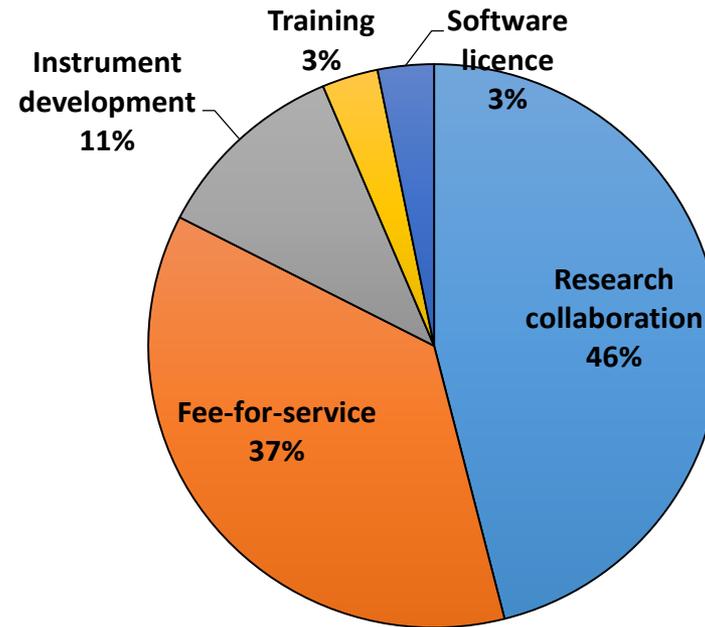- **Industry collaboration/ service at Centre level:** 63 projects involving 8 Instruct Centres.

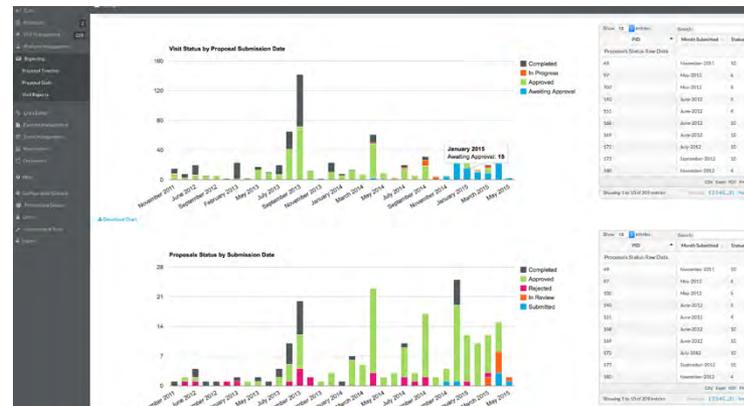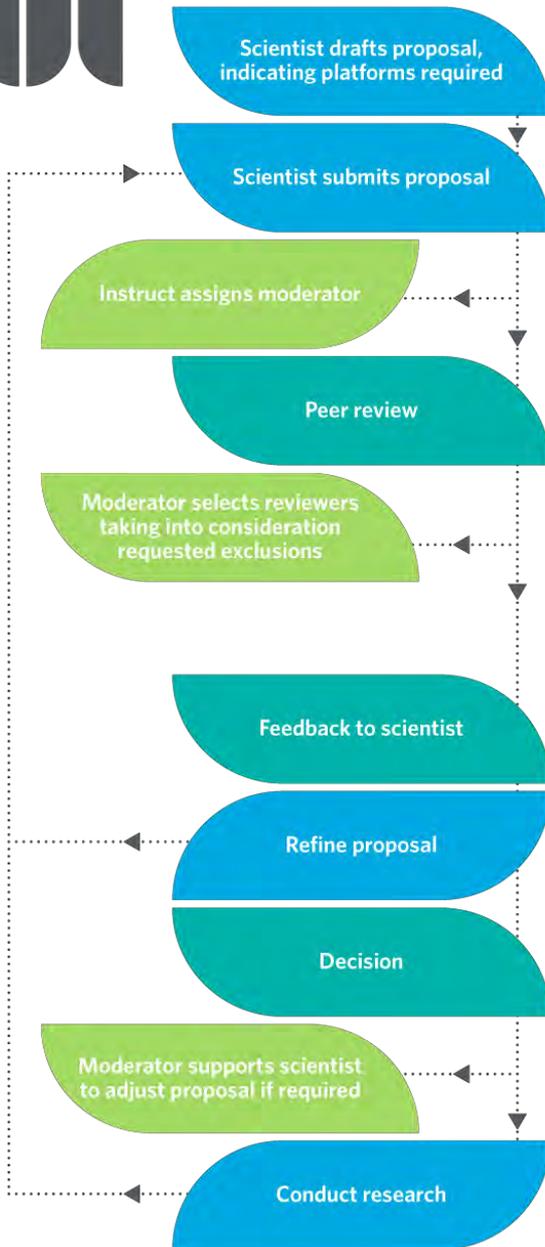# Links between Instruct Centres and Industry
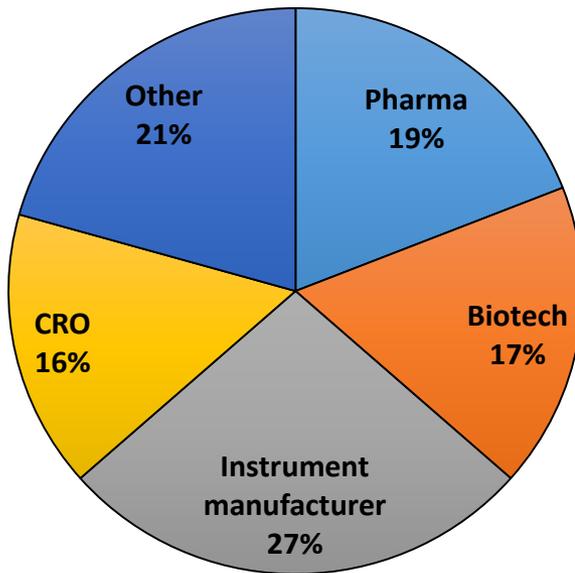
(a) Type of sector

(b) Type of relationship

Development of a unique software solution for managing access to Research Infrastructures (reviewing, tracking, reporting, metrics)

Instruct Data Management Policy

- Policy applies to Experimental, Reduced and Supporting data:

  - Experimental Data is Raw Data, Analysed Data and all associated Metadata produced during Instruct Access projects.

  - Reduced Data is data derived from experimental data, as an intermediate step in production of the final results.

  - Supporting data refers to the data necessary to reproduce the published conclusions, e.g. original X-ray data and electron micrographs.

- Instruct Centres do not claim any ownership of the Experimental Data that they produce. However, if user access requires substantial input of Facility staff (e.g. protocol development), then data ownership will be shared between Users and Facility.

- Instruct Centres aim to offer an archive to store data, especially in cases where the data volume makes this more practical than transferring the data to the user.

- Data remain confidential to the user until made available by them e.g. released into public repositories (e.g. Protein Data Bank). Users of Instruct Centre facilities are responsible for meeting any third party data management or transfer obligations that may be applicable.

Open Access to cryoElectron Microscopy

Number of EM structures

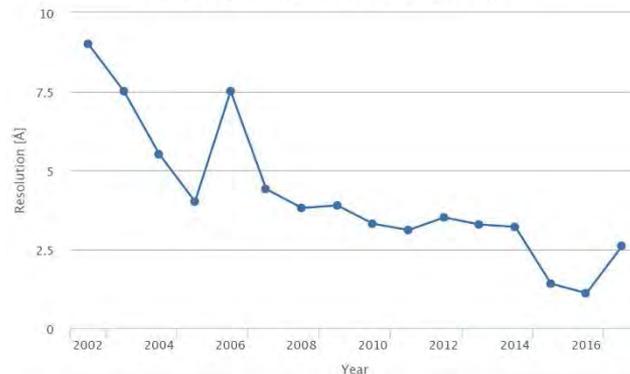Thanks to Martyn Winn for graphic
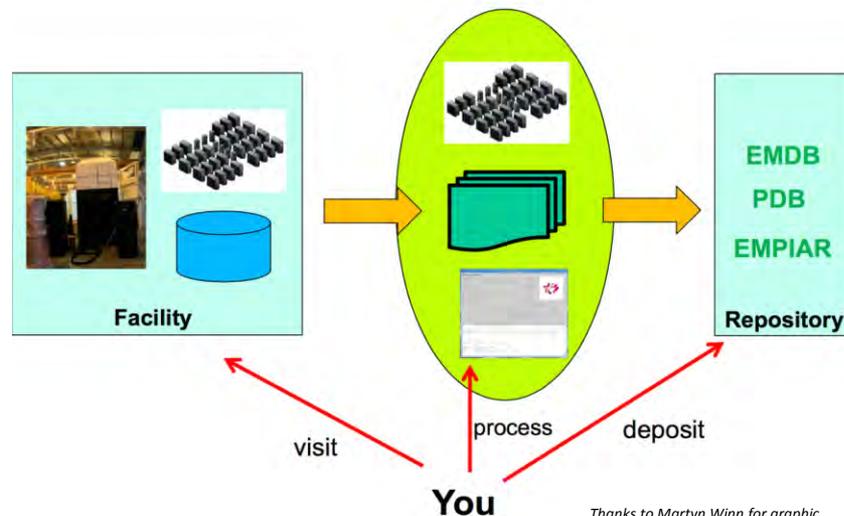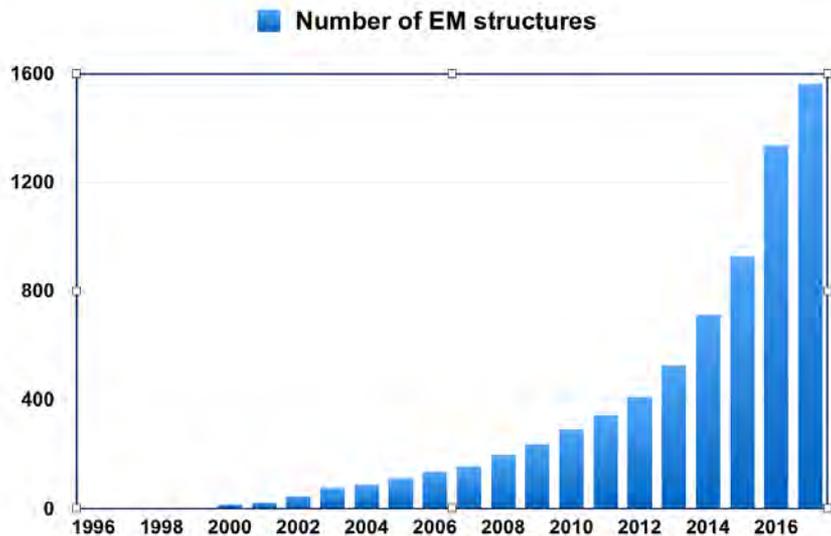
Public databases:

EMPIAR: **E**lectron **M**icroscopy **P**ublic **I**mage **A**rchive- raw 2D images

EMDB: **E**lectron **M**icroscopy **D**atabank- final EM maps

PDB: Protein Databank – fitted model

Highest resolution achieved in a given year

Instruct-ULTRA

- To date Instruct Centres have amassed over 4 PB of catalogued data in domains from Macromolecular crystallography, Electron Microscopy, XFEL, BioSAXS, Spectroscopy, Imaging and NMR.

- Two of the aims of Instruct-ULTRA are to increase the quality and integrity of structural data and metadata and to increase open data sharing.

- First Instruct Workshop on cryoEM best practices: May 9th-10th 2017 Harwell, UK.

Instruct Ultra — About · News · Work Packages · Events

First Instruct Workshop on cryoEM best practices

The First Instruct Workshop on cryoEM best practices took place on 9-10 May at the Research Complex at Harwell and was attended by 29 participants, including from industry (FEI, GSK, Vertex, Astrazeneca). The meeting covered both operational and computational methods through short scene-setting presentations from Instruct Centres (Peijun Zhang, eBIC; Ludo Renault, NeCEN, Bruno Klaholz, IGBMC, Martyn Winn CCP-EM, Jose-Maria Carazo, CNB-CSIC and Juha Huiskonen, OPIC) followed by discussion sessions. A plenary talk by Bridget Carragher (Simons

Funded by : Horizon 2020 Infradev-03-2016
Coordinator: Prof. Ray Owens (University of Oxford).

EOSC Science Demonstrator

Acronym: CryoEM workflows

"In this Science-Demonstrator we want to address the proper reporting of cryoEM image processing workflows, ensuring provenance at the level of data and analysis tools, linking workflow information with raw data either at cryoEM facilities, individual laboratories or public repositories, so that reproducibility of scientific results were enhanced, data and analysis workflows could be reused and properly mined, allowing for a deeper level of interoperability among information sources."

Support from: Instruct, EMDB-EBI and EMPIAR, WestLife VRE, CCP-EM, Spanish NGO.
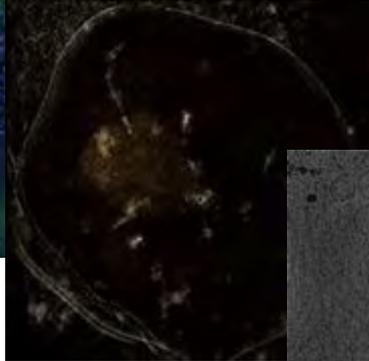**Project leader**: Prof. Jose-Maria Carazo (CNB-CSIC).

Future challenges in the Integration of Cell and Structural Biology
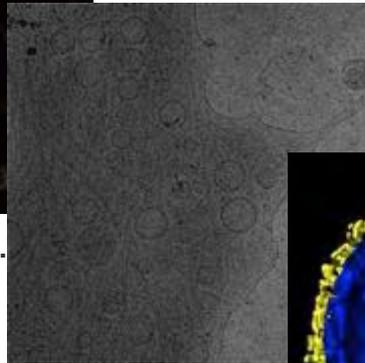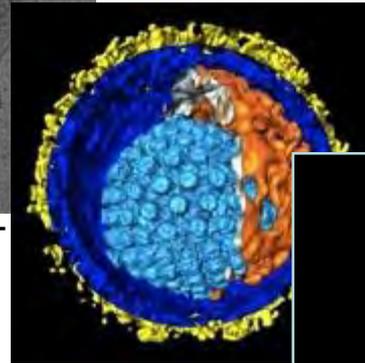

Increasing biological complexity
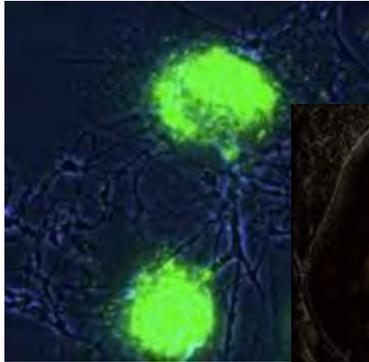
Fluor. microsc.

X-ray microsc.

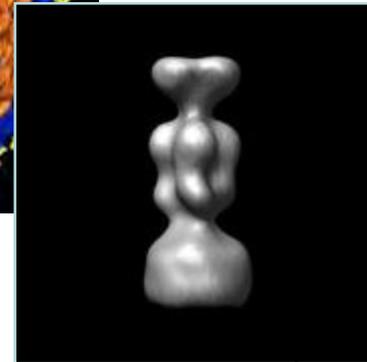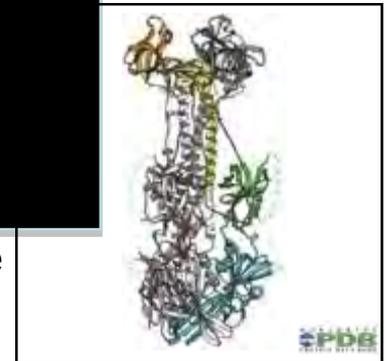- Multiple Centres

- Data management incl. provenances

Cellular cryoET

CryoET

- New algorithmns for Integrative approaches

Single particle

Increasing resolution

Crystallography/NMR