

Publishing Integrated Data at Scale: Update on the IDR Prototype

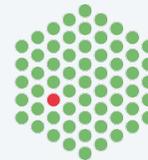
Josh Moore and Eleanor Williams

University of Dundee
The OME Consortium



Open Microscopy Environment
Centre for Gene Regulation & Expression
College of Life Sciences, University of Dundee
Dundee, Scotland, UK

EMBL-EBI



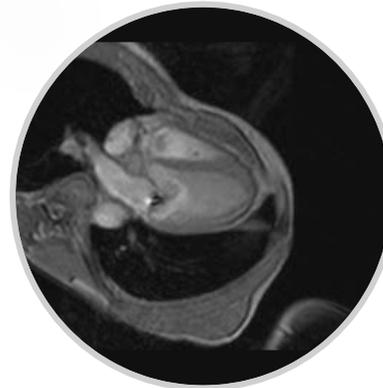
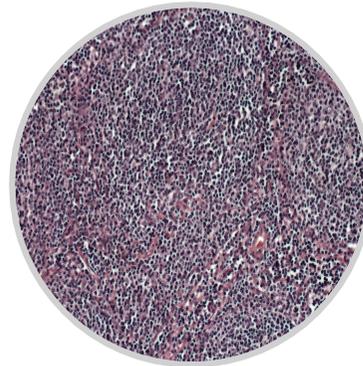
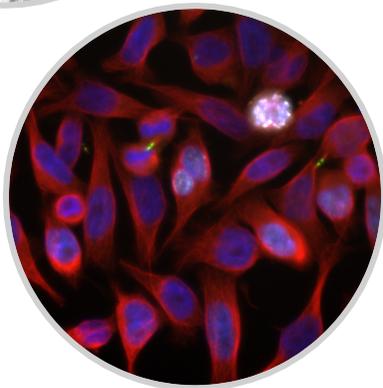
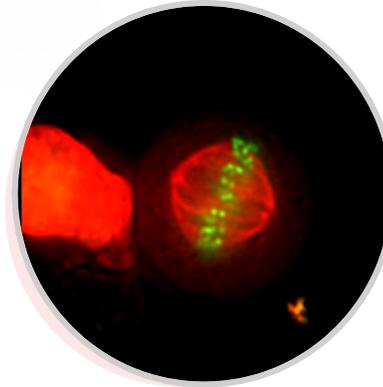
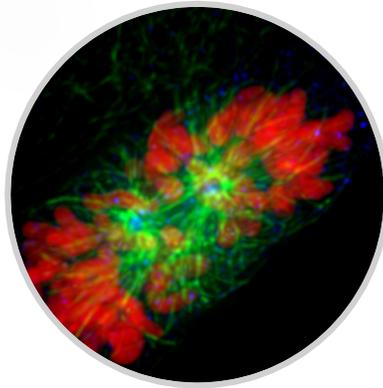
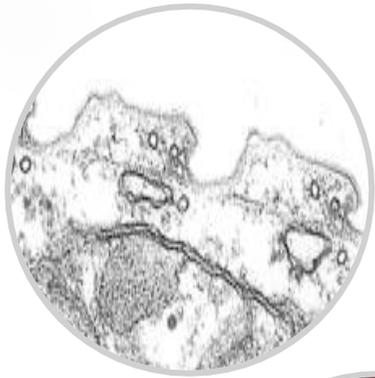
UNIVERSITY OF
CAMBRIDGE

IDR: Outline

- The vision
- Collection & integration
- Published & re-usable?
- The future

THE VISION

The Image Problem... is Ubiquitous



A pretty picture? A measurement? A resource?

```

3. jamoore@demo2-omero:~ (ssh)
name | name | value
-----+-----+-----
X_110222_S1 [Well A-1; Field #1] | Strain Name | empty
X_110222_S1 [Well C-11; Field #1] | Gene Identifier | SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Strain Name | rps2801
X_110222_S1 [Well C-11; Field #1] | Gene Symbol | rps2801
X_110222_S1 [Well C-11; Field #1] | Gene Identifier | SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Strain Name | rps2801
X_110222_S1 [Well C-11; Field #1] | Gene Symbol | rps2801
X_110222_S1 [Well C-11; Field #1] | Gene Identifier | SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC25G10.06
X_110222_S1 [Well C-11; Field #1] | Strain Name | rps2801
X_110222_S1 [Well C-11; Field #1] | Gene Symbol | rps2801
X_110222_S1 [Well E-4; Field #1] | Gene Identifier | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Strain Name | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Strain Name | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Strain Name | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier | SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC824.07
X_110222_S1 [Well E-4; Field #1] | Strain Name | SPAC824.07
X_110222_S1 [Well C-9; Field #1] | Gene Identifier | SPAC17G8.05
X_110222_S1 [Well C-9; Field #1] | Gene Identifier URL | http://www.pombase.org/spombe/result/SPAC17G8.05
X_110222_S1 [Well C-9; Field #1] | Strain Name | med20
X_110222_S1 [Well C-9; Field #1] | Gene Symbol | med20
X_110222_S1 [Well C-9; Field #1] | Phenotype 1 | abnormal microtubule cytoskeleton morphology during mitotic interphase
X_110222_S1 [Well C-9; Field #1] | Phenotype 1 Term Name a | abnormal
X_110222_S1 [Well C-9; Field #1] | Phenotype 1 Term Accession a | PATO_0000460
--More--

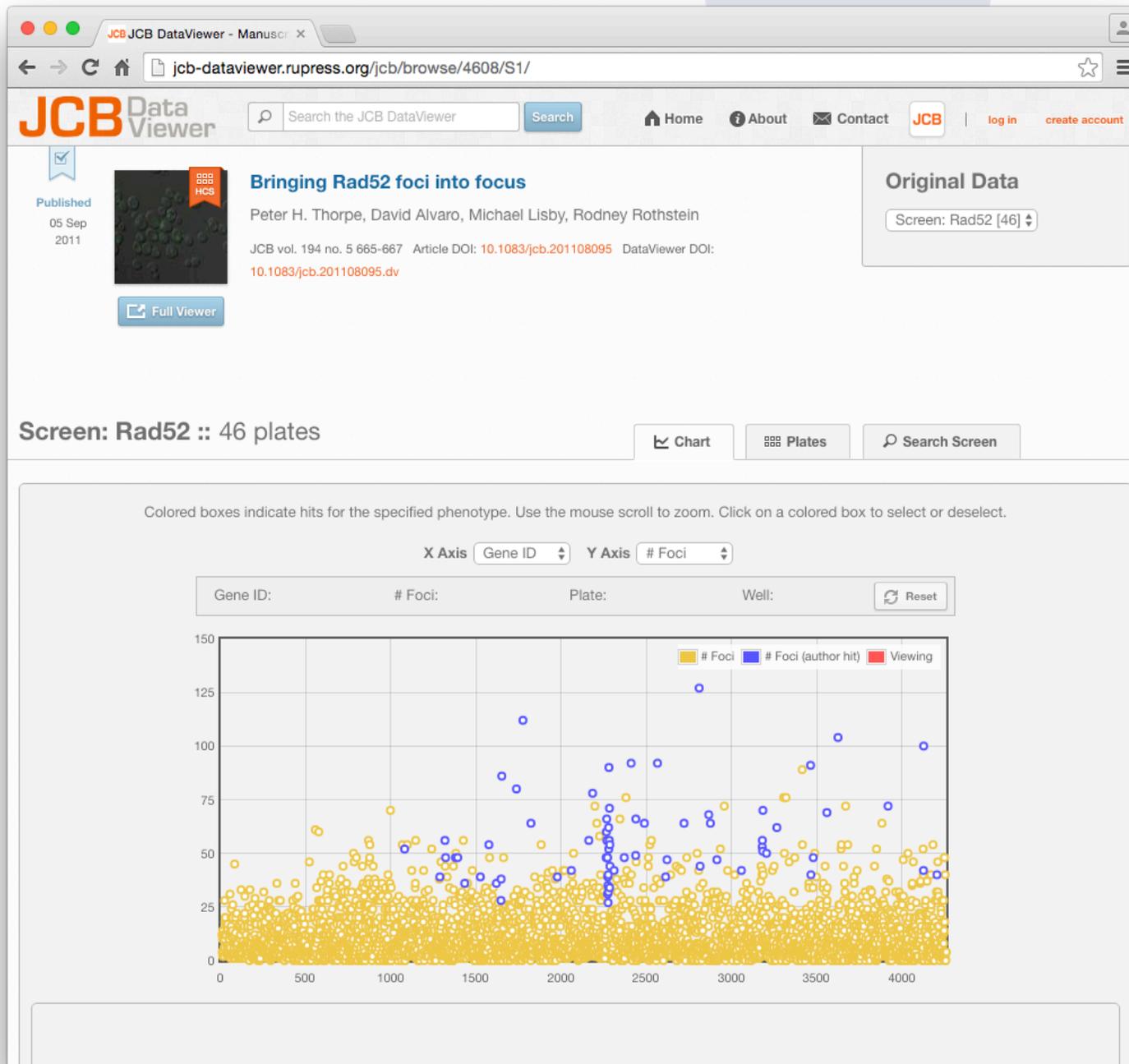
```

HTM_ACQUISITIONS_REGISTRY_TARA_HCS1_H5.xlsx

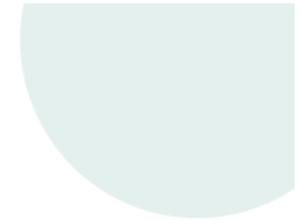
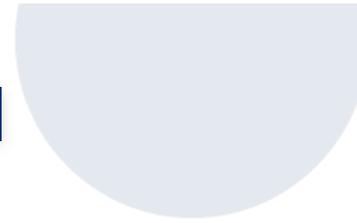
Search in Sheet

	A	B	C	D	E
107					
108	//METADATA				
	HTM_SAMPLE_LABEL_aggregated	EVENT_LABEL	EVENT_DATETIME_Start	EVENT_DATETIME_End	EVENT_LATITUDE_Start
109					EVENT_LATITUDE_End
110	S052--D0--R27--G100001472	TARA_20100517T0410Z_052_EVENT_PUMP	2010-05-17T04:10Z	2010-05-17T07:40Z	-16,957
111	S052--D1--R27--G100001473	TARA_20100517T1142Z_052_EVENT_PUMP	2010-05-17T11:42Z	2010-05-17T13:50Z	-16,9534
112	S064--D0--R27--G100002567	TARA_20100707T0448Z_064_EVENT_PUMP	2010-07-07T04:48Z	2010-07-07T07:45Z	-29,5019
113	S064--D1--R27--G100002568	TARA_20100708T0621Z_064_EVENT_PUMP	2010-07-08T06:21Z	2010-07-08T08:50Z	-29,5333
114	S065--D0--R27--G100002655	TARA_20100712T0559Z_065_EVENT_PUMP	2010-07-12T05:59Z	2010-07-12T09:00Z	-35,1728
115	S065--D1--R27--G100002656	TARA_20100712T1106Z_065_EVENT_MCASTS	2010-07-12T11:03:22Z	2010-07-12T12:12Z	-35,2421
116	S066--D0--R27--G100001988	TARA_20100715T1222Z_066_EVENT_PUMP	2010-07-15T12:22Z	2010-07-15T14:47Z	-34,9449
117	S066--D1--R27--G100001989	TARA_20100715T1539Z_066_EVENT_MCASTS	2010-07-15T15:36:10Z	2010-07-15T17:40Z	-34,8901
118	S067--D0--R27--G100002479	TARA_20100907T1016Z_067_EVENT_NET	2010-09-07T10:16Z	2010-09-07T10:21Z	-32,12
119	S068--D0--R27--G100002411	TARA_20100913T2237Z_068_EVENT_NET	2010-09-13T22:37Z	2010-09-13T22:51Z	-31,0536
120	S068--D1--R27--G100002481	TARA_20100914T1330Z_068_EVENT_PUMP	2010-09-14T13:30Z	2010-09-14T15:07Z	-31,027
121	S070--D0--R27--G100002163	TARA_20100921T1202Z_070_EVENT_NET	2010-09-21T12:02Z	2010-09-21T12:10Z	-20,4107
122	S071--D0--R27--G100002201	TARA_20100928T1214Z_071_EVENT_NET	2010-09-28T12:14Z	2010-09-28T12:23Z	-9,2612
123	S072--D0--R27--G100002978	TARA_20101005T1100Z_072_EVENT_NET	2010-10-05T11:00Z	2010-10-05T11:15Z	-8,7277
124	S072--D1--R27--G100002980	TARA_20101005T1828Z_072_EVENT_NET	2010-10-05T18:28Z	2010-10-05T18:50Z	-8,6831
125	S076--D0--R27--G100003346	TARA_20101016T1245Z_076_EVENT_NET	2010-10-16T12:45Z	2010-10-16T13:07Z	-20,9724
126	S076--D1--R27--G100003348	TARA_20101016T1953Z_076_EVENT_NET	2010-10-16T19:53Z	2010-10-16T20:48Z	-21,092
127	S078--D0--R27--G100003406	TARA_20101104T1424Z_078_EVENT_NET	2010-11-04T14:24Z	2010-11-04T14:42Z	-30,2163
128	S081--D0--R27--G100003739	TARA_20101202T1028Z_081_EVENT_PUMP	2010-12-02T10:28Z	2010-12-02T11:30Z	-44,5342
129	S081--D1--R27--G100003741	TARA_20101202T1814Z_081_EVENT_NET	2010-12-02T18:14Z	2010-12-02T18:27Z	-44,5838
130	S082--D0--R27--G100003584	TARA_20101206T1323Z_082_EVENT_NET	2010-12-06T13:23Z	2010-12-06T13:27Z	-47,2002
131	S082--D1--R27--G100003584	TARA_20101206T1323Z_082_EVENT_NET	2010-12-06T13:23Z	2010-12-06T13:27Z	-47,2002
132	S082--D1--R27--G100003586	TARA_20101206T1858Z_082_EVENT_PUMP	2010-12-06T18:58Z	2010-12-06T22:11Z	-47,2007
133	S082--D1--R27--G100003586	TARA_20101206T1858Z_082_EVENT_PUMP	2010-12-06T18:58Z	2010-12-06T22:11Z	-47,2007
134	S084--D0--R27--G100004906	TARA_20110103T1817Z_084_EVENT_NET	2011-01-03T18:17Z	2011-01-03T18:32Z	-60,4016
135	S084--D0--R27--G100004906	TARA_20110103T1817Z_084_EVENT_NET	2011-01-03T18:17Z	2011-01-03T18:32Z	-60,4016
136	S085--D0--R27--G100004339	TARA_20110106T1038Z_085_EVENT_PUMP	2011-01-06T10:38Z	2011-01-06T10:30Z	-62,0385
137	S085--D1--R27--G100004341	TARA_20110106T1857Z_085_EVENT_NET	2011-01-06T18:57Z	2011-01-06T19:21Z	-62,21
138	S091--D0--R27--G100004727	TARA_20110225T1532Z_091_EVENT_NET	2011-02-25T15:32Z	2011-02-25T13:38Z	-34,173
139	S092--D0--R27--G100004940	TARA_20110226T1440Z_092_EVENT_NET	2011-02-26T14:40Z	2011-02-26T14:57Z	-33,6737
140	S096--D0--R27--G100008990	TARA_20110324T1607Z_096_EVENT_NET	2011-03-24T16:07Z	2011-03-24T16:17Z	-29,6597
141	S100--D0--R27--G100009437	TARA_20110415T1625Z_100_EVENT_NET	2011-04-15T16:25Z	2011-04-15T16:40Z	-12,9488
142	S102--D0--R27--G100009617	TARA_20110421T2222Z_102_EVENT_NET	2011-04-21T22:22Z	2011-04-21T22:38Z	-5,252
143	S106--D0--R27--G100008907	TARA_20110503T1518Z_106_EVENT_NET	2011-05-03T15:18Z	2011-05-03T15:32Z	0,0247
144	S109--D0--R27--G100009725	TARA_20110512T1631Z_109_EVENT_NET	2011-05-12T16:31Z	2011-05-12T16:44Z	2,0399
145	S113--D0--R27--G100010454	TARA_20110619T1925Z_113_EVENT_NET	2011-06-19T19:25Z	2011-06-19T19:56Z	-23,1107
146	S122--D0--R27--G100010237	TARA_20110727T0127Z_122_EVENT_NET	2011-07-27T01:27Z	2011-07-27T01:45Z	-8,996
147	S122--D0--R27--G100010237	TARA_20110727T0127Z_122_EVENT_NET	2011-07-27T01:27Z	2011-07-27T01:45Z	-8,996
148	S122--D1--R27--G100010241	TARA_20110728T2305Z_122_EVENT_NET	2011-07-28T23:05Z	2011-07-28T23:46Z	-8,9543
149	S122--D1--R27--G100010241	TARA_20110728T2305Z_122_EVENT_NET	2011-07-28T23:05Z	2011-07-28T23:46Z	-8,9543
150	S122--D1--R27--G100010241	TARA_20110728T2305Z_122_EVENT_NET	2011-07-28T23:05Z	2011-07-28T23:46Z	-8,9543
151	S123--D0--R27--G100010173	TARA_20110731T1916Z_123_EVENT_NET	2011-07-31T19:16Z	2011-07-31T19:26Z	-8,8984
152	S123--D0--R27--G100010173	TARA_20110731T1916Z_123_EVENT_NET	2011-07-31T19:16Z	2011-07-31T19:26Z	-8,8984
153	S123--D3--R27--G100010177	TARA_20110801T2148Z_123_EVENT_NET	2011-08-01T21:48Z	2011-08-01T23:00Z	-8,903
154	S123--D3--R27--G100010177	TARA_20110801T2148Z_123_EVENT_NET	2011-08-01T21:48Z	2011-08-01T23:00Z	-8,903
155	S123--D3--R27--G100010177	TARA_20110801T2148Z_123_EVENT_NET	2011-08-01T21:48Z	2011-08-01T23:00Z	-8,903
156	S124--D0--R27--G100010607	TARA_20110804T1833Z_124_EVENT_PUMP	2011-08-04T18:33Z	2011-08-04T22:00Z	-9,1504
157	S124--D0--R27--G100010607	TARA_20110804T1833Z_124_EVENT_PUMP	2011-08-04T18:33Z	2011-08-04T22:00Z	-9,1504
158	S124--D0--R27--G100010607	TARA_20110804T1833Z_124_EVENT_PUMP	2011-08-04T18:33Z	2011-08-04T22:00Z	-9,1504
159	S125--D0--R27--G100010633	TARA_20110808T2343Z_125_EVENT_NET	2011-08-08T23:43Z	2011-08-08T00:00Z	0,0134

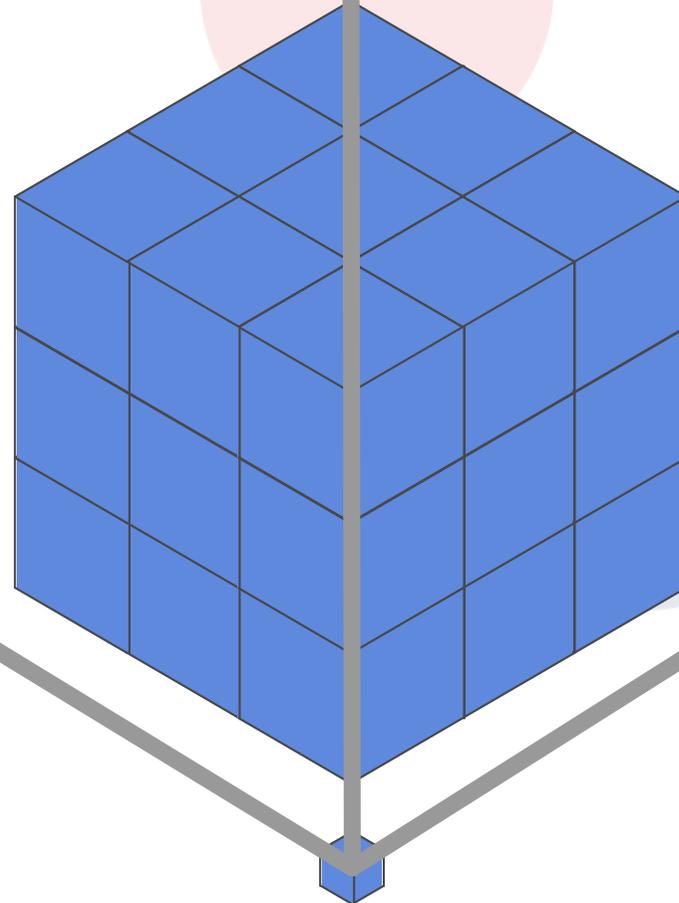
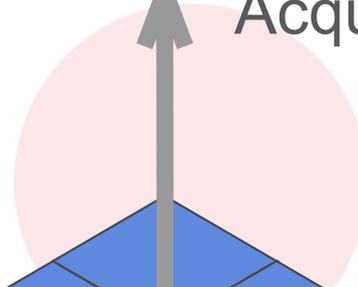
registry renaming missing samples Sheet3



The goal



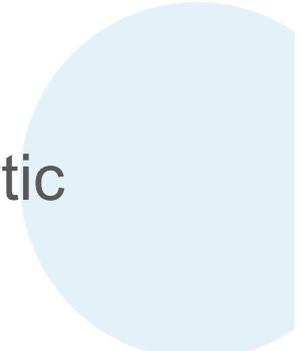
Acquisition



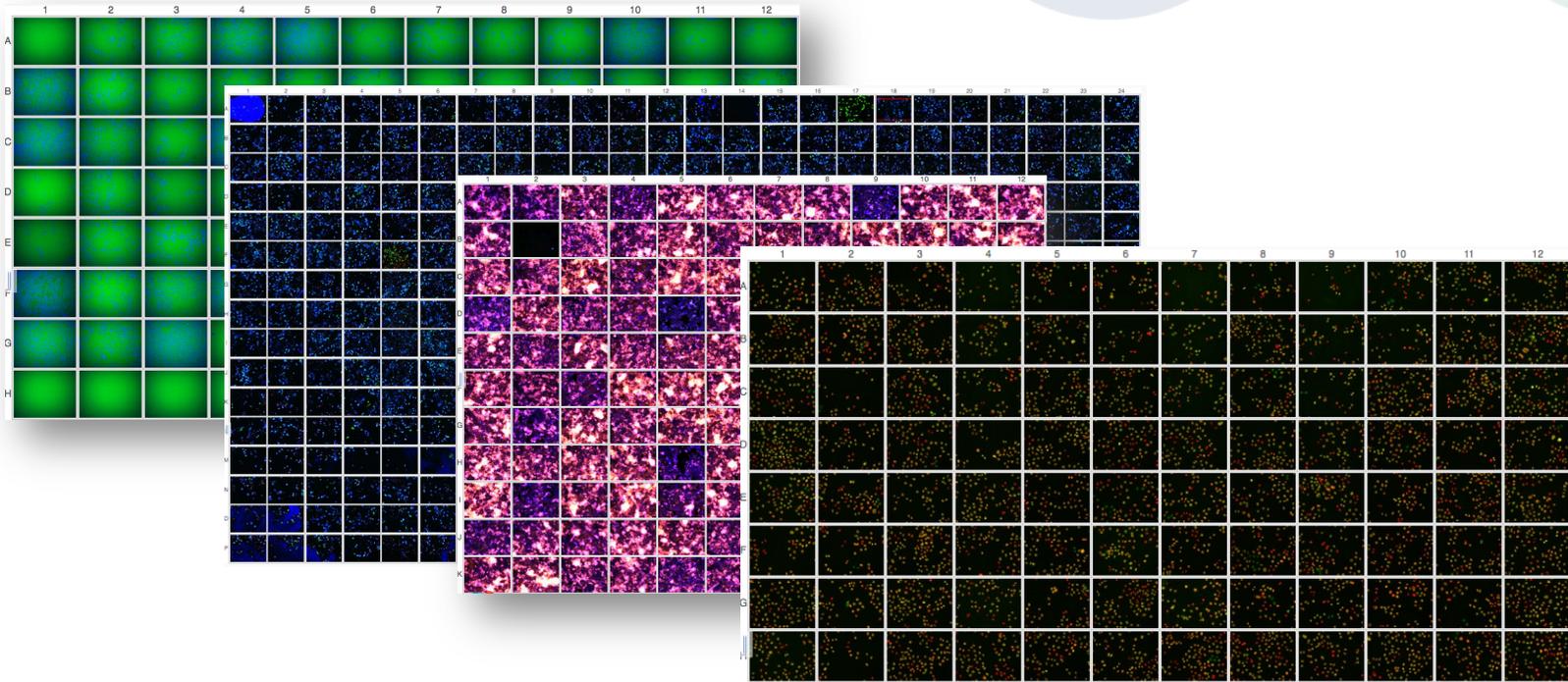
Experimental



Analytic



Combining existing data



JCB

LINCS

QBI

SSBD

SYN:GR:CE





Make it re-usable

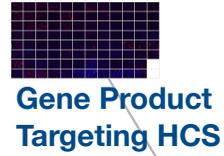
Building a Next Generation Image Repository: Molecular Annotation and Cloud-based Data Processing and Analysis

Reference	BB/M018423/1
Principal Investigator / Supervisor	Professor Jason Swedlow
Co-Investigators / Co-Supervisors	Dr Alvis Brazma , Dr Rafael Edgardo Carazo Salas
Institution	University of Dundee
Department	School of Life Sciences
Funding type	Research
Value (£)	1,788,152
	Current
	Research Grant
	01/01/2015
	30/06/2016
	18 months

In particular we will build links with established molecular and structural resources and work towards a seamless integration of these data, so that any scientist can easily browse, query and compute on genomic, structural and phenotypic data across several scales.

based on hardware infrastructure located at EMBL-EBI and integrated with its datasets to the world's scientific community. These resources will serve as the gateways and OMERO will be used to read, manage, serve, and link the data to the IDR. We will build custom user interfaces and workflows for the IDR, to ensure easy access and browsing to the datasets it holds. To enable computational re-analysis of the data, we will extend OMERO's distributed compute capacity and make use of EMBL-EBI's Embassy system to allow virtual access to IDR data. This virtual resource will provide a 'sandbox' for performing processing and reanalysis of data deposited in the IDR and provide a working example of a next generation data repository that stores and manages data, but also provides community services for scientific data.

IDR Vision



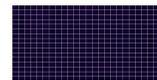
Gene Product Targeting HCS



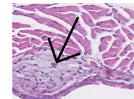
Genetic HCS



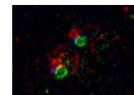
Geographic HCS



Chemical HCS



Histopathology



3D-SIM



Super-resolution

The screenshot shows the OMERO webclient interface. On the left is a file explorer with a tree view of experiments. The main area displays a grid of thumbnails labeled 'A' through 'H' and '1' through '7'. On the right, there is a metadata panel with tabs for 'General', 'Acquisition', and 'Preview'. The 'General' tab is active, showing details like 'Acquisition Date', 'Import Date', 'Dimensions (XY)', 'Pixels Type', 'Pixels Size (XYZ) (µm)', 'Z-sections/Timepoints', 'Channels', and 'ROI Count'. Below this are sections for 'Tags', 'Key-Value Pairs', and 'Tables'.

Integrated studies

Thumbnails

Experimental metadata

Ontological annotations

Feature vectors



Download local analysis



Cloud analysis



Cross-data browsing





Simon Li
@crucifixkiss



Following

Another 20TB of #BigData for the
@openmicroscopy @BBSRC @emlebi image
repository



RETWEET
1

LIKES
3

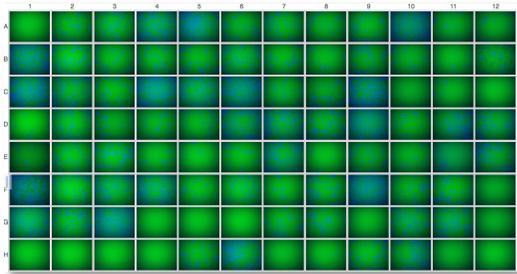


1:43 PM - 15 Sep 2015

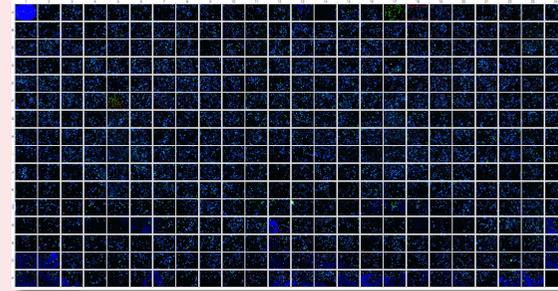
<http://j.mp/idr-disks>

COLLECTION & INTEGRATION

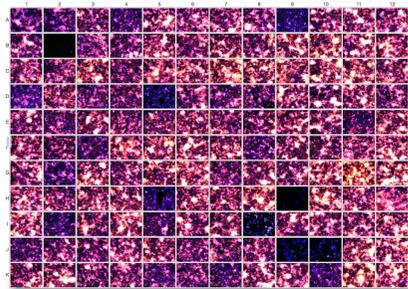
Data Integration



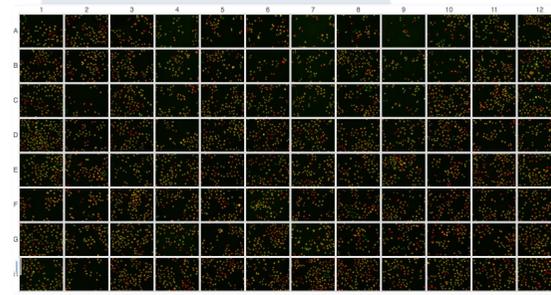
study 1



study 2

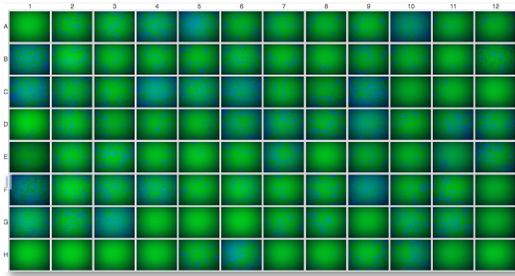


study 3

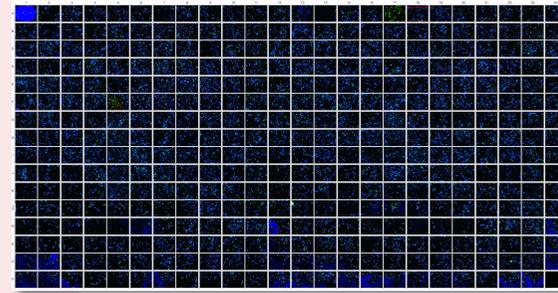


study 4

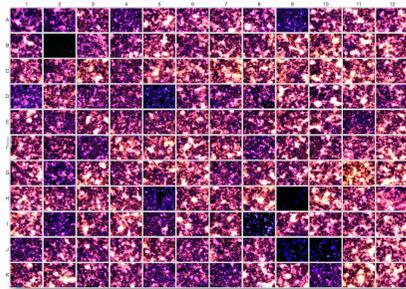
Data Integration



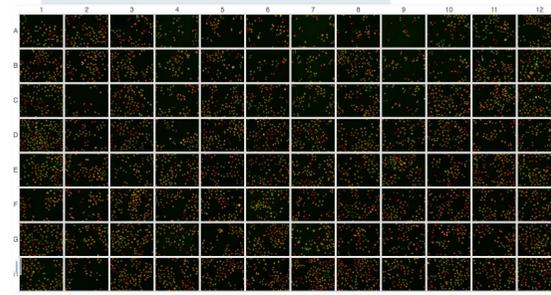
study 1



study 2

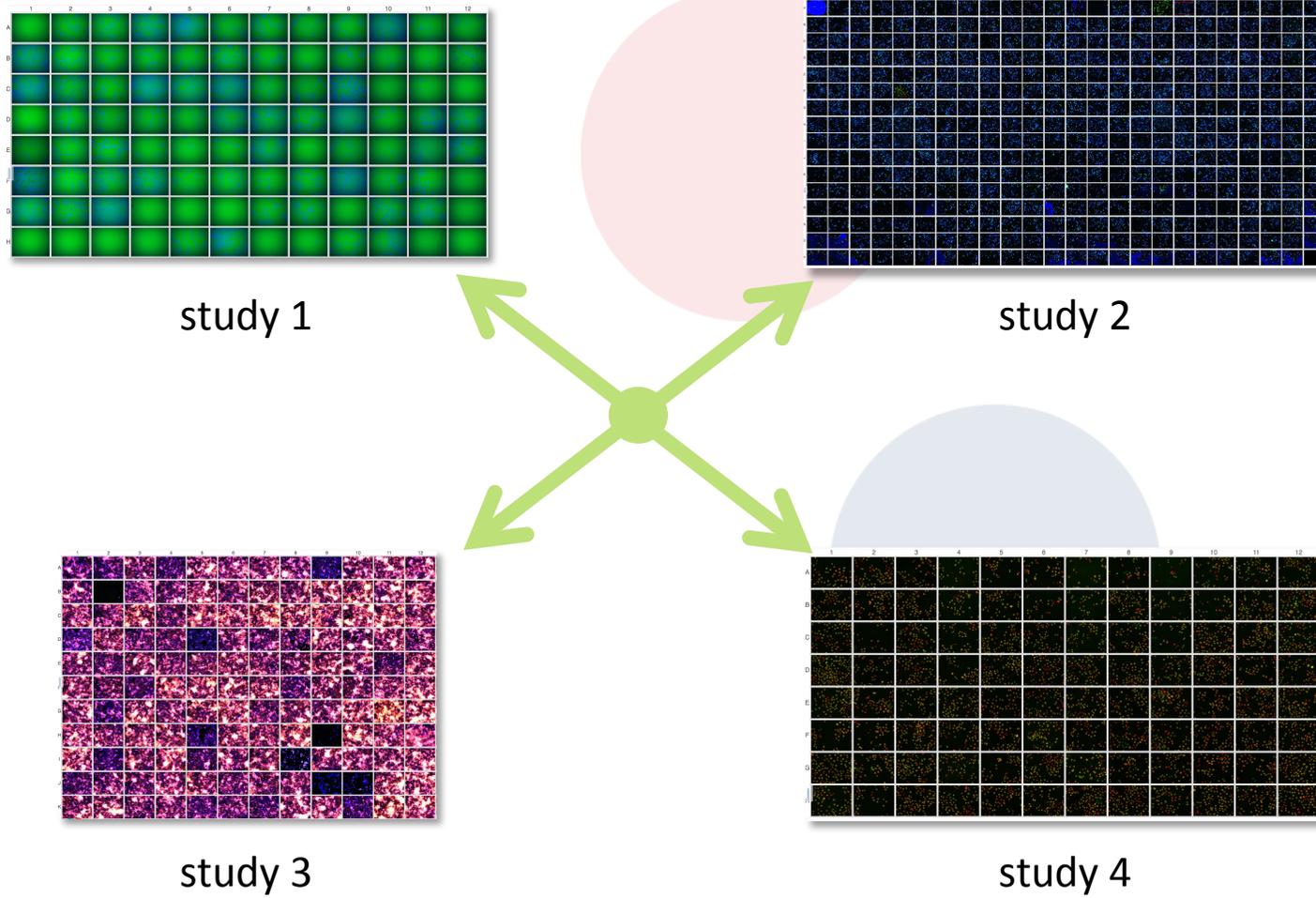


study 3



study 4

Data Integration



Study Metadata and Ontology Annotation

Your records and analyses



output →



study file



library or assay file



results file

experimental metadata

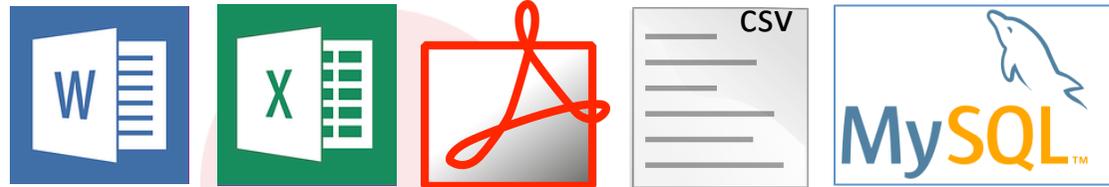
analytic metadata

Study Metadata and Ontology Annotation

Your records and analyses



output →



study file



library or assay file



results file



experimental metadata

analytic metadata

study type
imaging method
study description
contacts

reagent identifiers
gene identifiers
controls
quality control

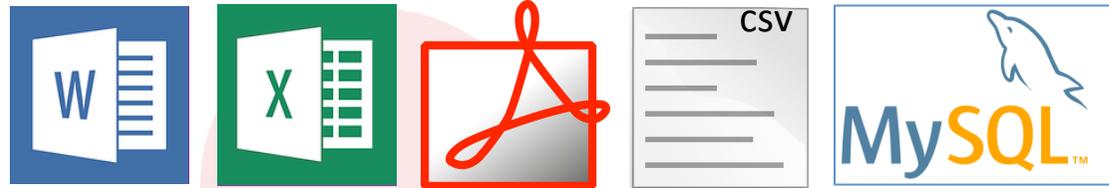
measurements
reproducibility
phenotypes

Study Metadata and Ontology Annotation

Your records and analyses



output



Cellular Phenotype Database



MAGE-TAB

isatab



study file



library or assay file



results file



experimental metadata

analytic metadata

study type
imaging method
study description
contacts

reagent identifiers
gene identifiers
controls
quality control

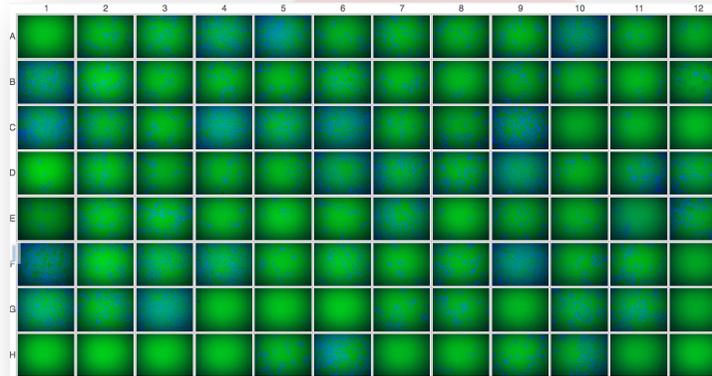
measurements
reproducibility
phenotypes

Experimental Metadata

High Content Screen

Screen Type

primary screen
secondary screen
validation screen



Screen Technology

Type

RNAi screen
gene deletion screen
protein screen
compound screen

Library Type

siRNA library
diploid homozygous deletion library
haploid deletion library
tag protein fusion library
 GFP protein fusion library
 YFP protein fusion library
 HA-Flag protein fusion library
compound library



Heterogeneity of Experimental Metadata

Controls



Negative control – no effect expected
e.g scramble, wild type cell line, DMSO, mock



Positive control – effect expected
- often not mentioned or listed in publication only



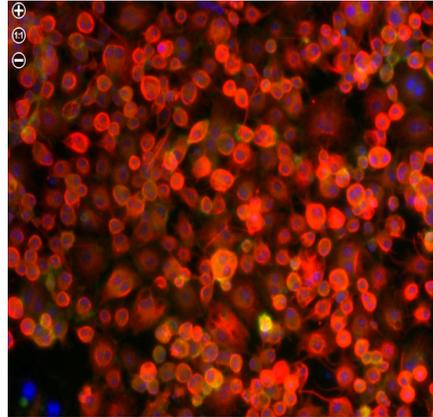
No reagent – no effect expected
- no treatment, often no annotation



Empty well
- not listed, listed but no annotation, images/no images

Heterogeneity of Analytic Metadata

cell shape
round or
non-
adherent



circular

orb

low eccentricity

spheroid

Rohn et al, 2011
siRNA knockdown of
Act5C

round cell phenotype
CMPO_0000118

**89% of phenotypes mapped
to ontology term**

**20 ontology terms appear in more
than one study**



<http://www.ebi.ac.uk/cmipo/>

Heterogeneity of Analytic Metadata

Level of annotation of results and phenotypes

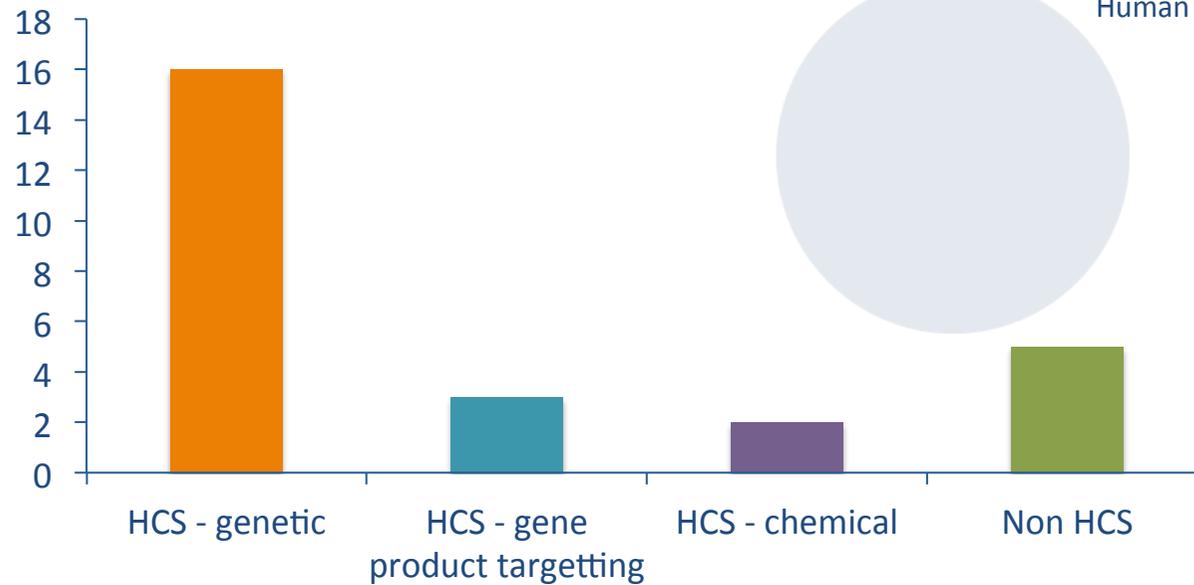
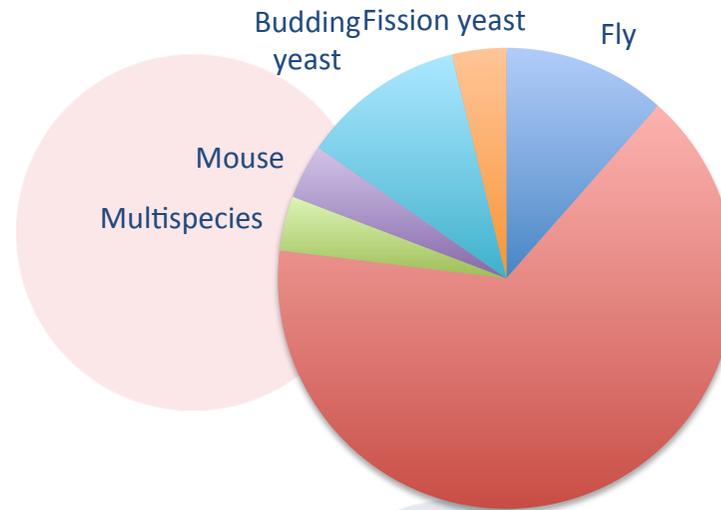
Gene	Phenotype
INTS1	shorter prophase, longer prophase

Gene	siRNA	Phenotype
INTS1	s25212	shorter prophase
INTS1	s25213	longer prophase

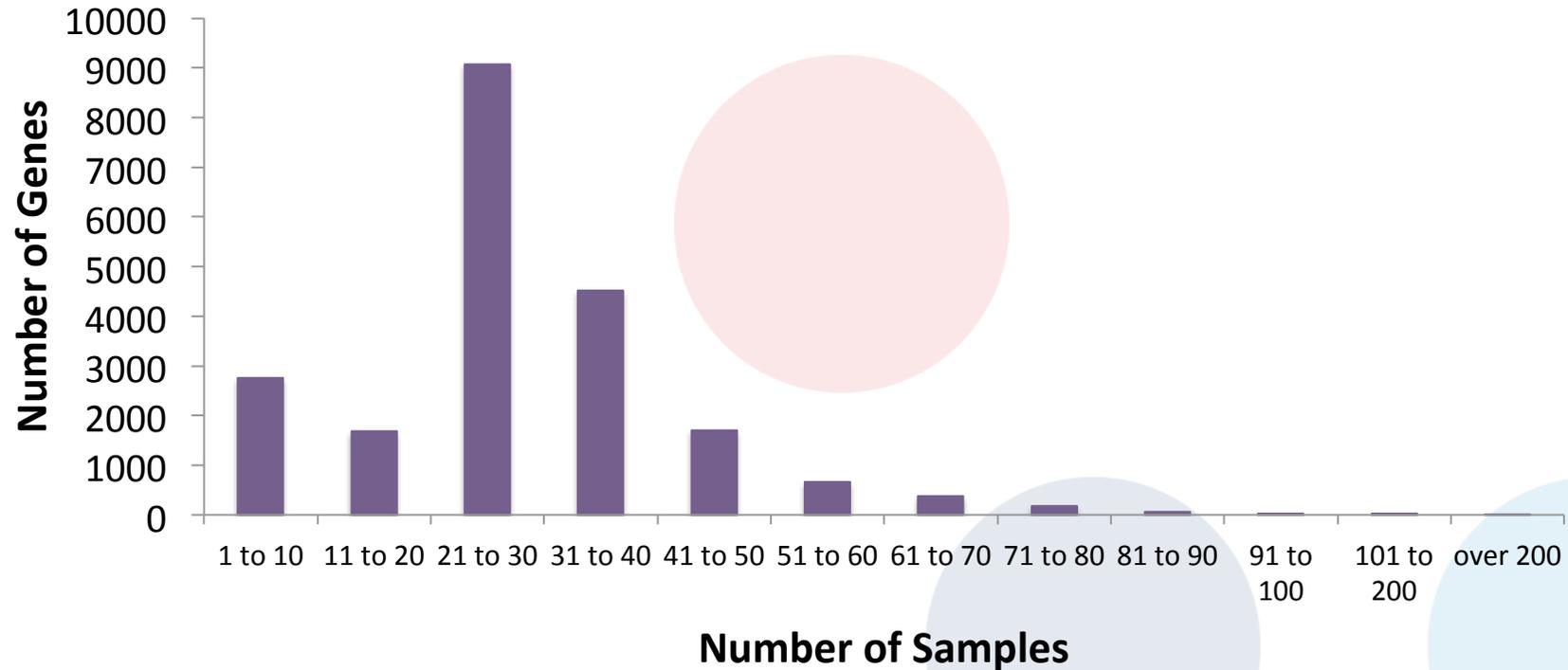
<https://github.com/IDR/idr-metadata>

Data collected so far ..

20 studies
24 screens or experiments



Wells per human gene (or ortholog) across screens



Each human gene or its ortholog

- in an average of 4 screens
- linked to an average of 28 different samples (excluding controls)

PUBLISHED & RE-USABLE?

IDR Reality

The screenshot displays the OMERO webclient interface. On the left is a file explorer showing a list of studies under 'All members'. The central area contains a grid of microscopy thumbnails labeled A through H and 1 through 7. On the right, a metadata panel is visible with sections for 'Well Details', 'Key-Value Pairs', and 'Tables'. The 'Well Details' section includes fields for Acquisition Date, Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ) in micrometers, Z-sections/Timepoints, Channels, and ROI Count. The 'Key-Value Pairs' section lists gene identifiers and phenotypes. The 'Tables' section shows information like Plate, Well Number, and Well.

Integrated studies

Experimental metadata

Ontological annotations

Thumbnails

Feature vectors



IDR Reality

The screenshot displays the OMERO webclient interface. On the left, a file explorer shows a list of studies under 'All members'. A blue arrow points from this list to a central grid of microscopy images. A green checkmark is placed next to the 'Integrated studies' label. Overlaid on the right is a 'Well Details' panel with the following information:

- Acquisition Date: 2012-07-31 12:09:15
- Import Date: 2015-11-23 16:28:47
- Dimensions (XY): 1376 x 1040
- Pixels Type: uint16
- Pixels Size (XYZ) (µm): 0.11 x 0.11
- Z-sections/Timepoints: 16 x 1
- Channels: Exp1Cam1, Exp2Cam1
- ROI Count: 65

Below the well details is a 'Gene Identifier' table:

Gene Identifier	SPAC3H1.13
Gene Identifier URL	http://www.pombase.org/spo
Strain Name	ppk13
Gene Symbol	ppk13

Further down, a 'Phenotype 1 Term' table is visible:

Phenotype 1 Term Accession	http://purl.obolibrary.org/obo/
Phenotype 1 Term Name	b microtubule cytoskeleton mo
Phenotype 1 Term Accession	CMPO_0000369
Phenotype 1 Term Accession	http://www.ebi.ac.uk/cmipo/C

At the bottom of the panel, a 'Tables' section shows:

INFO	
Plate:	2551
Well Number:	41
Well:	590652

Integrated studies

Experimental metadata

Ontological annotations

Feature vectors

<http://j.mp/idr0001>



IDR Reality

Integrated studies

Experimental metadata

Biological annotations

Feature vectors

General Acquisition Preview

TARA_HCS1_H5_G100001472_G100001473--2013_09_28_19_45_25_chamber--U00--V01

Plate ID: 151
Owner: Demo User

Creation Date: 2015-10-01 00:21:04

ANNOTATIONS Show all

SAMPLE_LABEL TARA_G100001472
SAMPLE_BARCODE G100001472
SAMPLE_LABEL_Replicates n/a
SAMPLE_LABEL_Complementary_Roscoff Roscoff:2558
SAMPLE_BARCODE_URI_Wetlab_Logsheet http://store.pangaea.de/Projects/TARA-OCEANS/
STATION_LABEL TARA_052
EVENT_LABEL_Complementary nan
EVENT_DEVICE_LABEL PUMP,High Volume Peristaltic Pump
http://store.pangaea.de/Projects/TARA-OCEANS/
(Dimier)
COMMENT_on_Logsheet=n/a.COMMENT_by_E
P
S
nan
0
PROT
ORGANISM

TARA OCEANS Tara_UTC YYYY MM DD HH MM ### EVENT_PUMP_01

Start 2011 09 10 17 57 Station
End 2011 09 11 01 24 129

	LAT	DD	MM.MMM	LON	DDD	MM.MMM	PUMP#	DAY / NIGHT
Start	N 06		40.257		153	03.945	1	DAY
End	06		51.574		153	06.659		

OPERATORS DEPTH_Intended (m) CABLE_Length (m) Angle (deg) Speed (m/s)

JP SURFACE 1.0

OPERATION	START TIME (HH:MM)	END TIME (HH:MM)	PUMP RATE (Hz)	COMMENTS
Rincing Pump:	17:57	18:02	60	
Filling 200L (B&V):	18:03	18:10	60	
Filling 200L GPSS (PROT&G):	18:11	18:36	20	
Filling 200L (B&V):				
Flow through GPSS (when 5µm net is not avail.) (indicate pauses)	18:52 21:36 23:33	21:27 23:10 01:24	20 20 20	Upper 5µm net probably broken

<http://j.mp/idr0015>



IDR Reality

The screenshot displays the OMERO webclient interface. On the left is a file explorer showing a list of studies under 'All members'. The central area contains a grid of microscopy thumbnails labeled A through H and 1 through 7. On the right, a metadata panel is visible with sections for 'Well Details', 'Key-Value Pairs', and 'Tables'. The 'Well Details' section includes fields for Acquisition Date, Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ) in micrometers, Z-sections/Timepoints, Channels, and ROI Count. The 'Key-Value Pairs' section lists gene identifiers and phenotypes. The 'Tables' section shows information like Plate, Well Number, and Well.

Integrated studies

Experimental metadata

Ontological annotations

Thumbnails

Feature vectors



IDR Reality

The image displays two screenshots of the OMERO web interface. The left screenshot shows a 'Demo data' list with 13 studies and a grid of 13 thumbnails. The right screenshot shows an 'All members' list with 20 studies. The word 'Thumbnails' is written in blue text below the grid.

Thumbnails

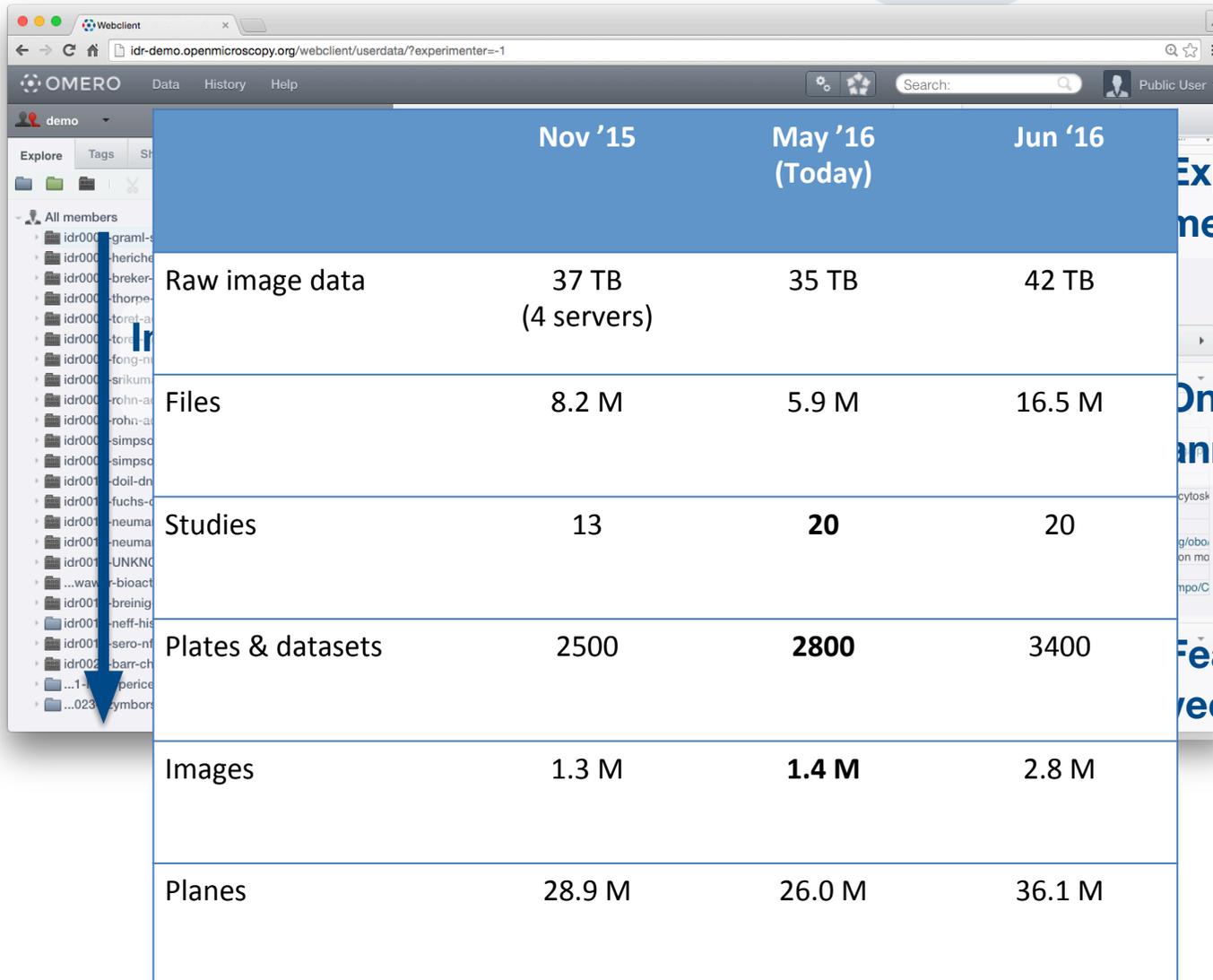
Experimental data

Biological variations

13 studies across 4 OMERO 5.1 servers → 20 studies in 1 OMERO 5.2 server



IDR Reality



	Nov '15	May '16 (Today)	Jun '16
Raw image data	37 TB (4 servers)	35 TB	42 TB
Files	8.2 M	5.9 M	16.5 M
Studies	13	20	20
Plates & datasets	2500	2800	3400
Images	1.3 M	1.4 M	2.8 M
Planes	28.9 M	26.0 M	36.1 M

Experimental
metadata

Ontological
annotations

Feature
vectors



IDR Reality

The screenshot displays the OMERO webclient interface. On the left is a file explorer showing a list of studies under 'All members'. The central area contains a grid of microscopy thumbnails labeled A through H and 1 through 7. On the right, a metadata panel is visible with sections for 'Well Details', 'Key-Value Pairs', and 'Tables'. The 'Well Details' section includes fields for Acquisition Date, Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ) in micrometers, Z-sections/Timepoints, Channels, and ROI Count. The 'Key-Value Pairs' section lists fields such as Gene Identifier, Strain Name, Gene Symbol, and Phenotype 1. The 'Tables' section shows a table with columns for INFO, Plate, Well Number, and Well.

Integrated studies

Experimental metadata

Ontological annotations

Thumbnails

Feature vectors



IDR Reality

The screenshot shows the OMERO webclient interface. On the left is a file explorer with a list of folders and files, including names like 'idr000-graml-sysgro/screenA 192'. In the center is a grid of microscopy thumbnails labeled A through H and 1 through 7. On the right is a metadata panel with sections for 'Well Details', 'Tags', 'Key-Value Pairs', and 'Tables'. The 'Well Details' section includes fields for Acquisition Date, Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ) in micrometers, Z-sections/Timepoints, Channels, and ROI Count. The 'Key-Value Pairs' section lists gene identifiers and phenotype terms. The 'Tables' section shows information like Plate, Well Number, and Well.

Integrated studies

Experimental metadata

Ontological annotations

Thumbnails

Feature vectors



Download local analysis



Cloud analysis



Cross-data browsing



THE FUTURE

This week: how we did it?

IDR / idr-metadata

Curated metadata for the Image Data Repository — Edit

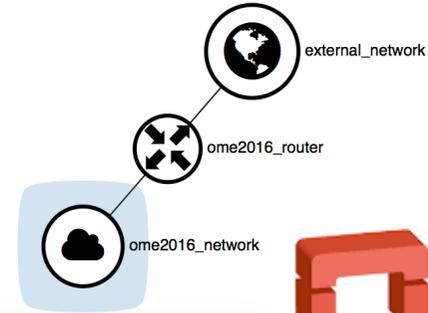
595 commits | 1 branch | 2 releases | 4 contributors

Branch: master | New pull request

Create new file | Upload files | Find file | Clone or download

ac744 Merge pull request #92 from joshmoore/idr0020 | Latest commit 1f239fc a day ago

- idr0001-graml-sysgro Merge pull request #77 from joshmoore/sysgro-rdef | a day ago
- idr0002-heriche-condensation Add other bulk.yml files using parent `columns` | 20 days ago
- idr0003-breker-plasticity Remove bogus breker plate lines
- idr0004-thorpe-rad52 Add other bulk.yml files using parent `columns`
- idr0005-toret-adhesion Merge pull request #75 from manics/renderdefs
- idr0006-fong-nuclearbodies Merge pull request #75 from manics/renderdefs
- idr0007-srikumar-sumo Add other bulk.yml files using parent `columns`
- idr0008-rohn-actinome Add other bulk.yml files using parent `columns`
- idr0009-simpson-secretion Add other bulk.yml files using parent `columns`
- idr0010-doil-dnamage Merge pull request #80 from eleanorwilliams/idr



ac744 / idr-metadata

idr-metadata / scripts / featuresAnalysis / .ipynb_checkpoints / CondensationPCAnalysis-checkpoint.ipynb

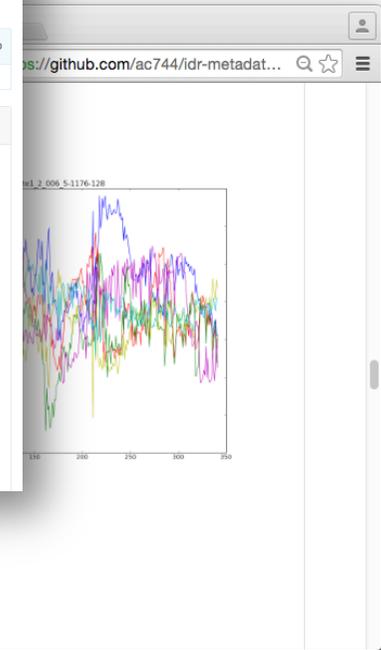
1 contributor

```

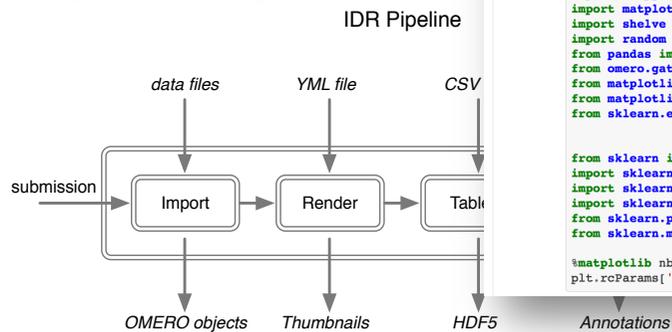
In [1]: import os
import numpy as np
import matplotlib.pyplot as plt
import shelve
import random
from pandas import Series, DataFrame, read_csv, merge, concat, read_hdf
from omero.gateway import BlitzGateway
from matplotlib.widgets import Slider
from matplotlib import gridspec
from sklearn.externals import joblib

from sklearn import random_projection
import sklearn.neighbors as nn
import sklearn.manifold as man
import sklearn.decomposition as dec
from sklearn.preprocessing import scale, robust_scale
from sklearn.metrics.pairwise import euclidean_distances

%matplotlib nbagg
plt.rcParams['image.cmap'] = 'gray'
  
```



<http://j.mp/idr-git>



O. Tange (2011): **GNU Parallel - The Command-Line Power**
 login: The USENIX Magazine, February 2011:42-47.



IDR Future

The screenshot shows the OMERO webclient interface. On the left, a list of studies is displayed under the heading "Integrated studies". The main area shows a grid of microscopy thumbnails labeled "Thumbnails". On the right, a metadata panel is visible, containing sections for "Well Details", "Key-Value Pairs", and "Tables".

Integrated studies

Thumbnails

Well Details

Acquisition Date:	2012-07-31 12:09:15
Import Date:	2015-11-23 16:28:47
Dimensions (XY):	1376 x 1040
Pixels Type:	uint16
Pixels Size (XYZ) (µm):	0.11 x 0.11
Z-sections/Timepoints:	16 x 1
Channels:	Exp1Cam1, Exp2Cam1
ROI Count:	65

Key-Value Pairs

Added by: Demo User
openmicroscopy.org/omero/bu...

Gene Identifier	SPAC3H1.13
Gene Identifier URL	http://www.pombase.org/SPAC3H1.13
Strain Name	ppk13
Gene Symbol	ppk13
Phenotype 1	abnormal microtubule cytoskeleton
Phenotype 1 Term Name a	abnormal
Phenotype 1 Term Accessio	PATO_0000460
Phenotype 1 Term Accessio	http://purl.obolibrary.org/obo/PATO_0000460
Phenotype 1 Term Name b	microtubule cytoskeleton mo
Phenotype 1 Term Accessio	CMPO_0000369
Phenotype 1 Term Accessio	http://www.ebi.ac.uk/cmipo/C

Tables

INFO	
Plate:	2551
Well Number:	41
Well:	590652

Experimental metadata

Ontological annotations

Feature vectors



Download local analysis



Cloud analysis



Cross-data browsing



IDR Future

The screenshot shows the OMERO webclient interface. On the left is a file explorer with a list of folders and files. The main area displays a grid of microscopy thumbnails. On the right, there is a metadata panel with sections for 'Well Details', 'Tags', 'Key-Value Pairs', and 'Tables'. A large blue arrow points from the 'Integrated studies' text to the file explorer.

Index	Field#1	1	2	3	4	5	6	7
A								
B								
C								
D								
E								
F					FS			
G								
H								

Well Details	Value
Acquisition Date:	2012-07-31 12:09:15
Import Date:	2015-11-23 16:28:47
Dimensions (XY):	1376 x 1040
Pixels Type:	uint16
Pixels Size (XYZ) (µm):	0.11 x 0.11
Z-sections/Timepoints:	16 x 1
Channels:	Exp1Cam1, Exp2Cam1
ROI Count:	65

Key-Value Pairs	Value
Gene Identifier	SPAC3H1.13
Gene Identifier URL	http://www.pombase.org/SPAC3H1.13
Strain Name	ppk13
Gene Symbol	ppk13
Phenotype 1	abnormal microtubule cytoskeleton
Phenotype 1 Term Name a	abnormal
Phenotype 1 Term Accession	PATO_0000460
Phenotype 1 Term Accession	http://purl.obolibrary.org/obo/PATO_0000460
Phenotype 1 Term Name b	microtubule cytoskeleton mo
Phenotype 1 Term Accession	CMPO_0000369
Phenotype 1 Term Accession	http://www.ebi.ac.uk/cmipo/C

Tables	Value
INFO	
Plate:	2551
Well Number:	41
Well:	590652

Integrated studies

Experimental metadata

Ontological annotations

Thumbnails

Feature vectors



Download local analysis



Cloud analysis



Cross-data browsing



Thanks to the team



Jason Swedlow



Josh Moore



Simon Li



Eleanor Williams



Richard Ferguson



Simone Leo



Ola Tarkowska



Alvis Brazma



Ugis Sarkans



Simon Jupp



Tony Burdett



Rafael Carazo-Salas



Balint Antal

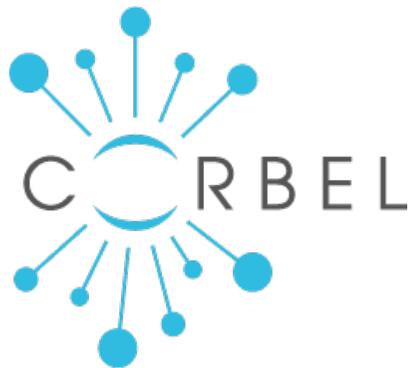


Anatole Chessel

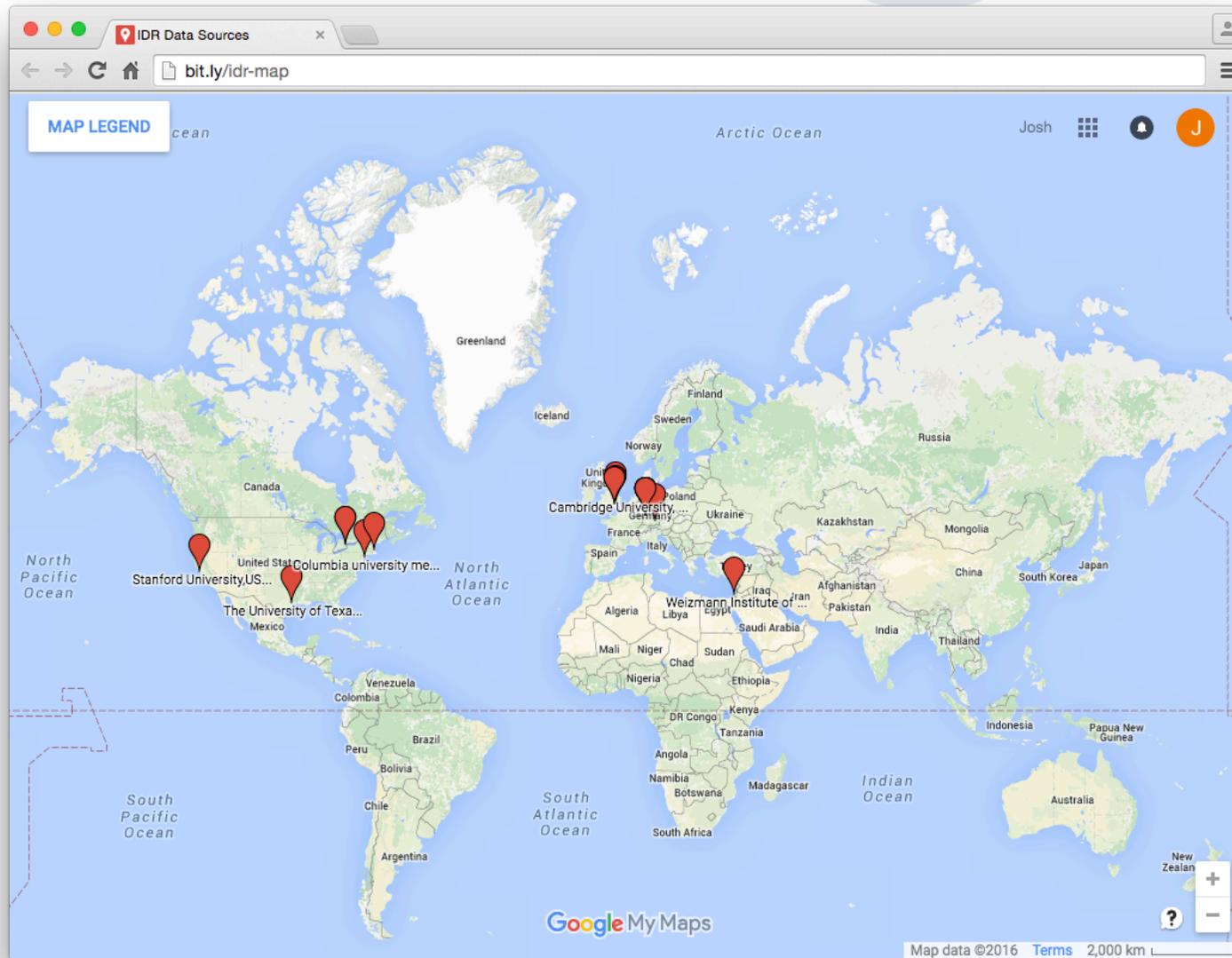


Thanks to funders

wellcometrust



Thanks for the data



<http://j.mp/idr-map>