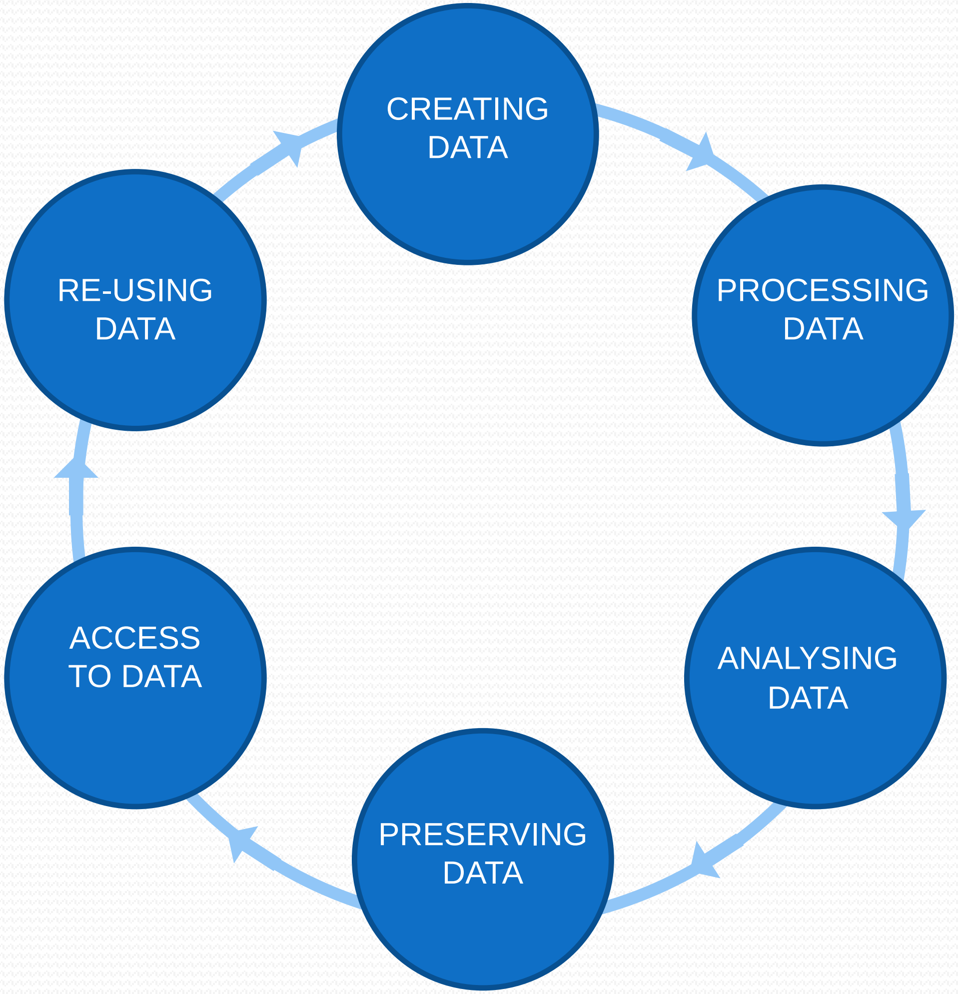
Tipping the balance

introducing data management on a centre-wide level

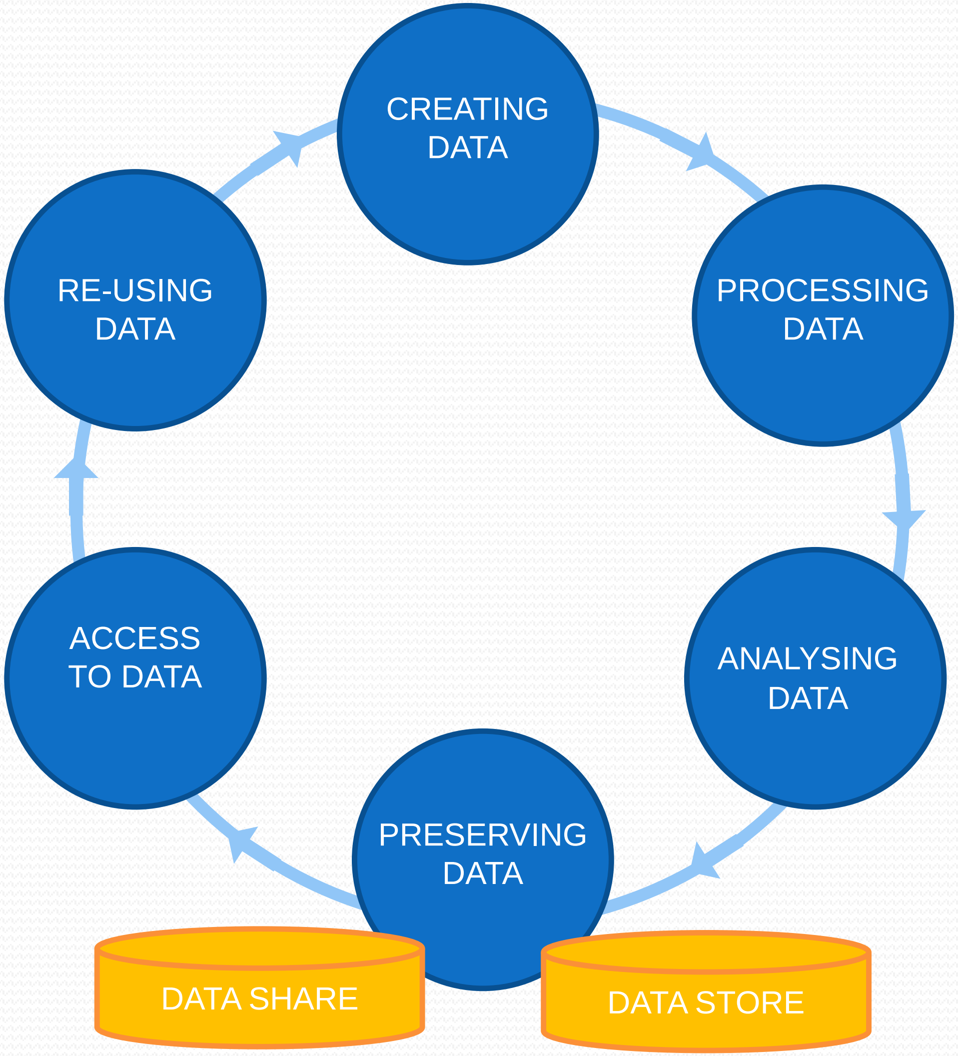
Tomasz Zieliński, Eilidh Troup, Andrew Millar

SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

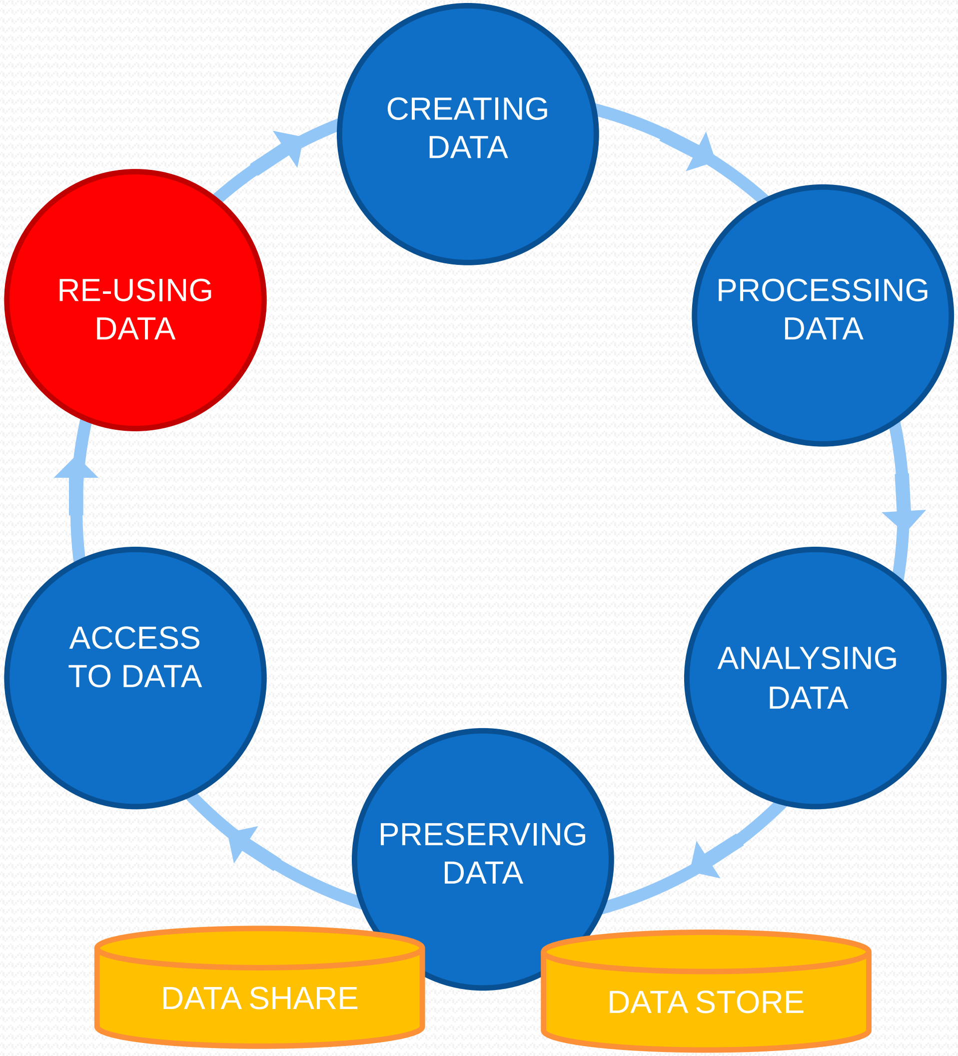
Data life-cycle



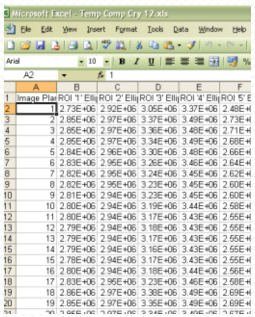
Data life-cycle



Data life-cycle



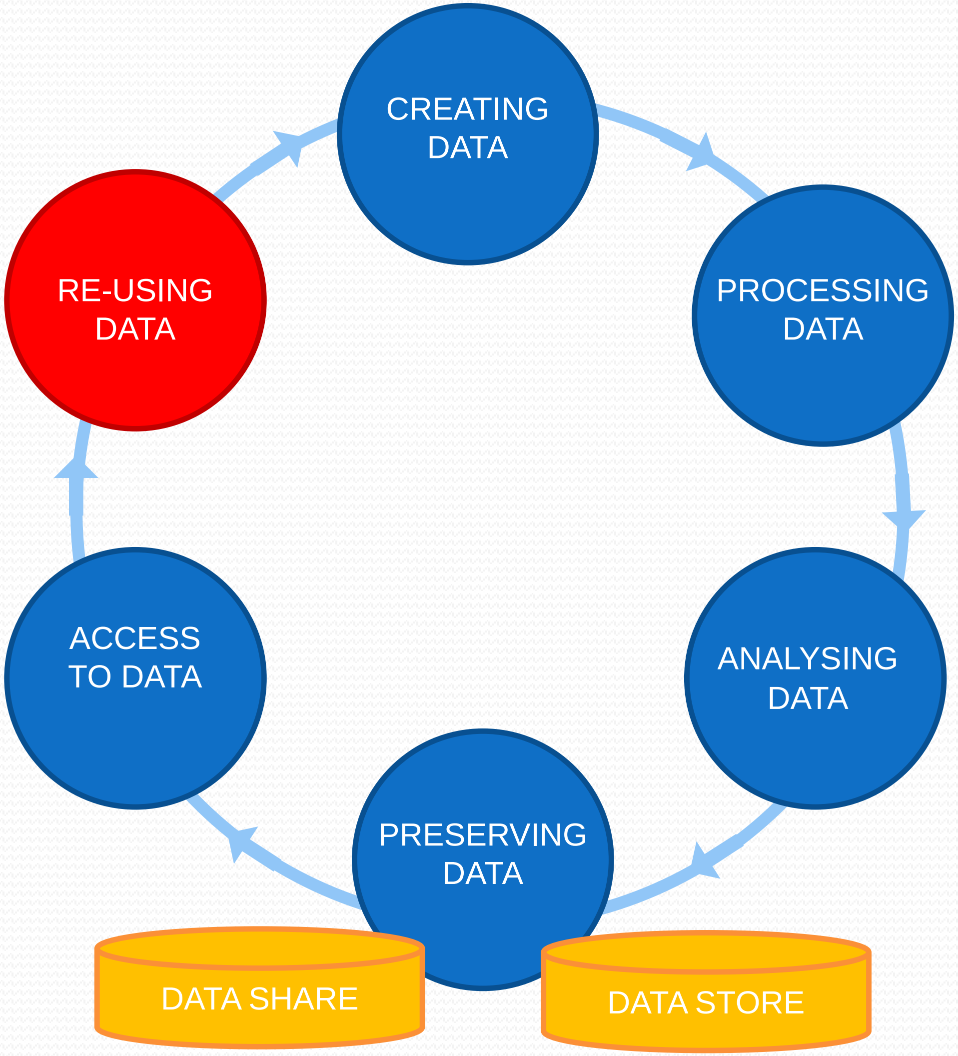
Data life-cycle


Microsoft Excel - Temp Comp Cry 12.xls

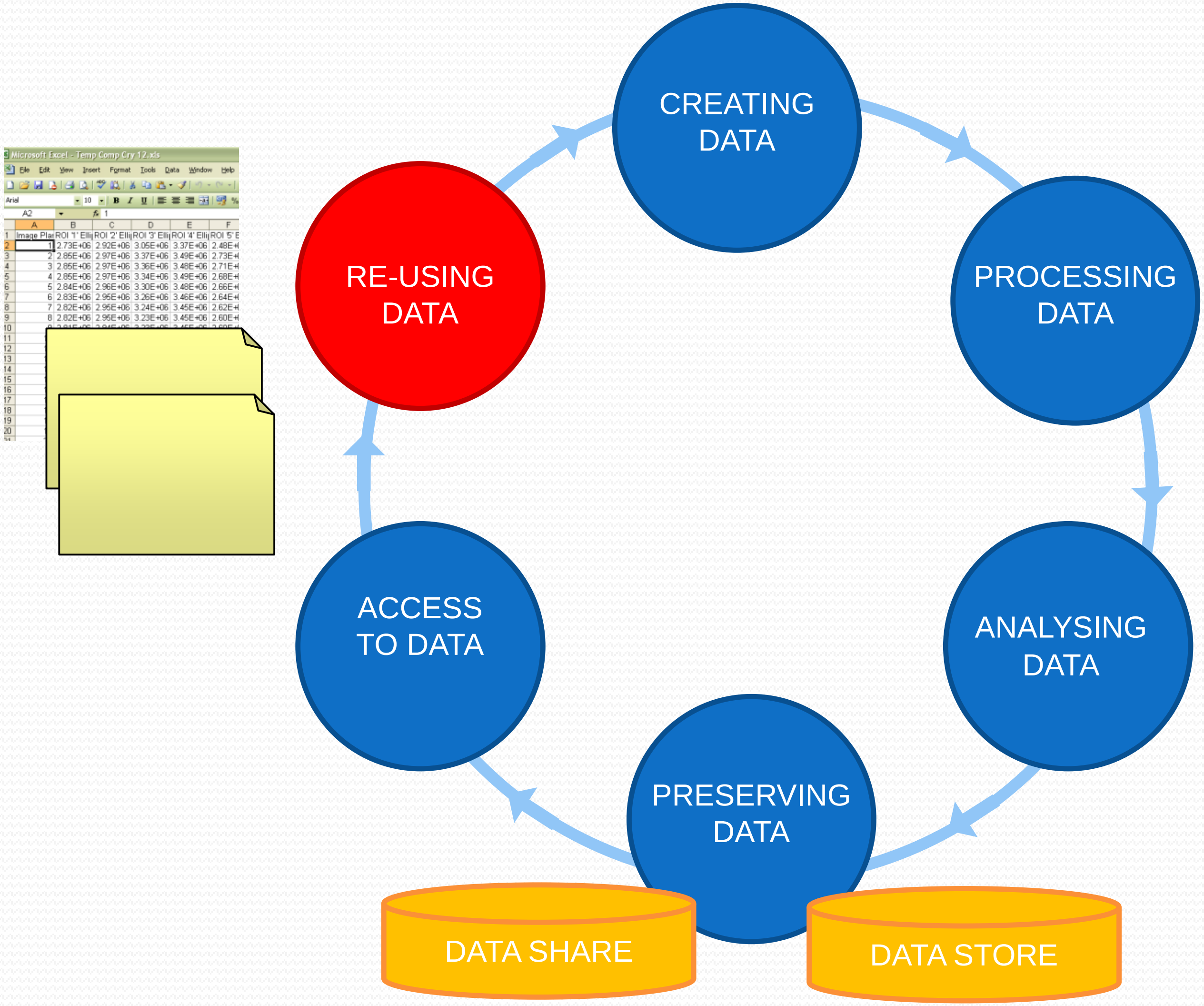
File Edit View Insert Format Tools Data Window Help

A2 1

	A	B	C	D	E	F
1	Image Plar	ROI '1' Elli	ROI '2' Elli	ROI '3' Elli	ROI '4' Elli	ROI '5' E
2	1	2.73E+06	2.92E+06	3.05E+06	3.37E+06	2.48E+
3	2	2.85E+06	2.97E+06	3.37E+06	3.49E+06	2.73E+
4	3	2.85E+06	2.97E+06	3.36E+06	3.48E+06	2.71E+
5	4	2.85E+06	2.97E+06	3.34E+06	3.49E+06	2.68E+
6	5	2.84E+06	2.96E+06	3.30E+06	3.48E+06	2.66E+
7	6	2.83E+06	2.95E+06	3.26E+06	3.46E+06	2.64E+
8	7	2.82E+06	2.95E+06	3.24E+06	3.45E+06	2.62E+
9	8	2.82E+06	2.95E+06	3.23E+06	3.45E+06	2.60E+
10	9	2.81E+06	2.94E+06	3.23E+06	3.45E+06	2.60E+
11	10	2.80E+06	2.94E+06	3.19E+06	3.44E+06	2.58E+
12	11	2.80E+06	2.94E+06	3.17E+06	3.43E+06	2.55E+
13	12	2.79E+06	2.94E+06	3.18E+06	3.43E+06	2.55E+
14	13	2.79E+06	2.94E+06	3.17E+06	3.43E+06	2.55E+
15	14	2.79E+06	2.94E+06	3.16E+06	3.43E+06	2.55E+
16	15	2.78E+06	2.94E+06	3.17E+06	3.43E+06	2.55E+
17	16	2.80E+06	2.94E+06	3.18E+06	3.44E+06	2.56E+
18	17	2.83E+06	2.95E+06	3.23E+06	3.46E+06	2.58E+
19	18	2.86E+06	2.97E+06	3.38E+06	3.49E+06	2.69E+
20	19	2.85E+06	2.97E+06	3.35E+06	3.49E+06	2.69E+

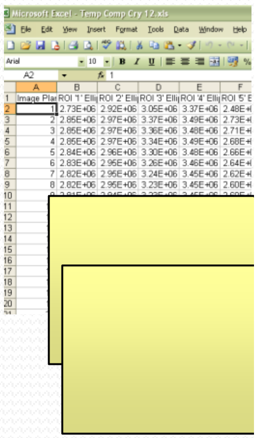


Data life-cycle



SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Data life-cycle

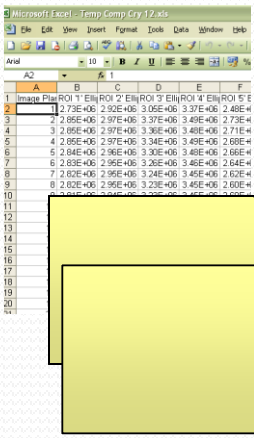


Data Creator:	Martin, Sarah F.; Noordally, Zeenat B.; van Ooijen, Gerben; Barrios-Llerena, Martin E.; Simpson, T. Ian; Millar, Andrew J.; Hindle, Matthew M.; Thierry Le Bihan
Date Available:	2014-06-24
Citation:	Martin, Sarah F.; Noordally, Zeenat B.; van Ooijen, Gerben; Barrios-Llerena, Martin E.; Simpson, T. Ian; Millar, Andrew J.; Hindle, Matthew M.; Thierry Le Bihan. (2014). The reduced kinome of Ostreococcus tauri: core eukaryotic signalling components in a tractable model species, [Dataset]. University of Edinburgh. SynthSys and School of Biological Sciences. http://dx.doi.org/10.7488/ds/72.
Dataset Description (abstract):	Background The current knowledge of eukaryote signalling originates from phenotypically diverse organisms. There is a pressing need to identify conserved signalling components among eukaryotes, which will lead to the transfer of knowledge across kingdoms. Two useful properties of a eukaryote model for signalling are (1) reduced signalling complexity, (2) conservation of signalling components. The alga Ostreococcus tauri is described as the smallest free-living eukaryote. With less than 8,000 genes, it represents a highly constrained genomic palette. Results Our survey revealed 133 protein kinases and 34 protein phosphatases (1.7% and 0.4% of the proteome). We conducted phospho-proteomic experiments and constructed domain structures and phylogenies for the catalytic protein-kinases. For each of the major kinases families we review the completeness and divergence of O. tauri representatives in comparison to the well-studied kinomes of the laboratory models Arabidopsis thaliana and Saccharomyces cerevisiae, and of Homo sapiens. Many kinase clades in O. tauri were reduced to a single member, in preference to the loss of family diversity, whereas TKL and ABC1 clades were expanded. We also identified kinases that have been lost in A. thaliana but retained in O. tauri. For three, contrasting eukaryotic pathways – TOR, MAPK, and the circadian clock – we established the subset of conserved components and demonstrate conserved sites of substrate phosphorylation and kinase motifs. Conclusions We conclude that O. tauri satisfies our two central requirements. Several of its kinases are more closely related to H. sapiens orthologs than S. cerevisiae is to H. sapiens. The greatly reduced kinome of O. tauri is therefore a suitable model for signalling in free-living eukaryotes.

DATA SHARE

DATA STORE

SynthSys
SYNTHETIC & SYSTEMS BIOLOGY



Data Creator:	Martin, Sarah F.; Noordally, Ian; Millar, Andrew J.; Hindle
Date Available:	2014-06-24
Citation:	Martin, Sarah F.; Noordally, Ian; Millar, Andrew J.; Hindle Ostreococcus tauri: core eu University of Edinburgh. Syn /ds/72.
Dataset Description (abstract):	Background The current kn organisms. There is a press eukaryotes, which will lead t eukaryote model for signallir components. The alga Ostre less than 8,000 genes, it rep revealed 133 protein kinases conducted phospho-proteom the catalytic protein-kinases and divergence of O. tauri re laboratory models Arabidops Many kinase clades in O. ta diversity, whereas TKL and lost in A. thaliana but retain MAPK, and the circadian cl demonstrate conserved site conclude that O. tauri satisf related to H. sapiens orthol O. tauri is therefore a suitab

DATA SHARE

Hindle *et al. BMC Genomics* 2014, **15**:640
http://www.biomedcentral.com/1471-2164/15/640

BMC Genomics

RESEARCH ARTICLE **Open Access**

The reduced kinome of *Ostreococcus tauri*: core eukaryotic signalling components in a tractable model species

Matthew M Hindle[1,4], Sarah F Martin[1,2], Zeenat B Noordally[1,2], Gerben van Ooijen[1,2], Martin E Barrios-Llerena[1,2], T Ian Simpson[3,4], Thierry Le Bihan[1,2] and Andrew J Millar[1,2*]

Abstract

Background: The current knowledge of eukaryote signalling originates from phenotypically diverse organisms. There is a pressing need to identify conserved signalling components among eukaryotes, which will lead to the transfer of knowledge across kingdoms. Two useful properties of a eukaryote model for signalling are (1) reduced signalling complexity, and (2) conservation of signalling components. The alga *Ostreococcus tauri* is described as the smallest free-living eukaryote. With less than 8,000 genes, it represents a highly constrained genomic palette.

Results: Our survey revealed 133 protein kinases and 34 protein phosphatases (1.7% and 0.4% of the proteome). We conducted phosphoproteomic experiments and constructed domain structures and phylogenies for the catalytic protein-kinases. For each of the major kinases families we review the completeness and divergence of *O. tauri* representatives in comparison to the well-studied kinomes of the laboratory models *Arabidopsis thaliana* and *Saccharomyces cerevisiae*, and of *Homo sapiens*. Many kinase clades in *O. tauri* were reduced to a single member, in preference to the loss of family diversity, whereas TKL and ABC1 clades were expanded. We also identified kinases that have been lost in *A. thaliana* but retained in *O. tauri*. For three, contrasting eukaryotic pathways – TOR, MAPK, and the circadian clock – we established the subset of conserved components and demonstrate conserved sites of substrate phosphorylation and kinase motifs.

Conclusions: We conclude that *O. tauri* satisfies our two central requirements. Several of its kinases are more closely related to *H. sapiens* orthologs than *S. cerevisiae* is to *H. sapiens*. The greatly reduced kinome of *O. tauri* is therefore a suitable model for signalling in free-living eukaryotes.

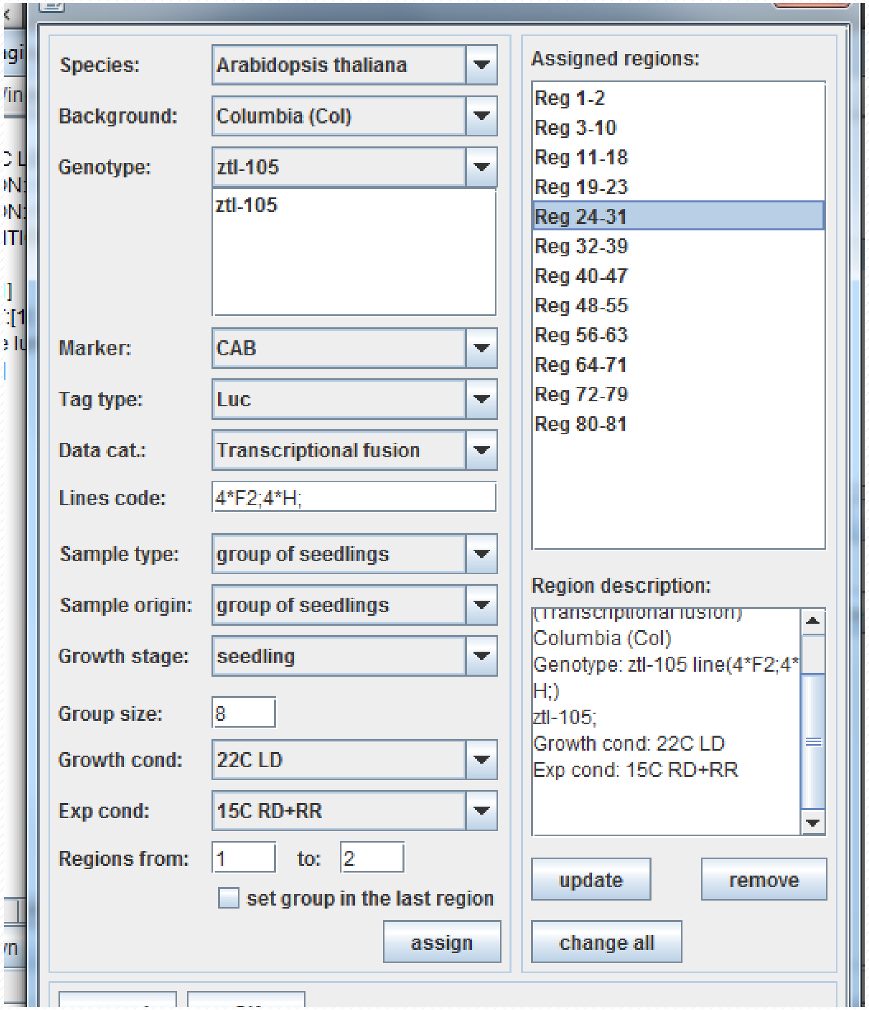
Keywords: Conserved eukaryote signalling, Protein kinase phylogeny, *Ostreococcus tauri*, Model kinome, Phosphorylation, TOR signalling, MAPK cascade, Circadian clock

Background

Protein kinases are a major component of the complex signalling networks that coordinate all fundamental cellular processes, including transcription, cell cycle and metabolism. Protein kinases and phosphatases elicit reversible phosphorylation, which enable the rapid cellular responses that are crucial for survival in a continually changing environment. Protein kinases activate and deactivate proteins by addition of the gamma-phosphate from ATP to serine (S), threonine (T), tyrosine (Y), aspartate (D) or histidine (H) amino acid residues [1]. Cascades of consecutive kinase-mediated phosphorylation events constitute the backbone of signalling pathways [2]. The complexity of the signalling networks scales with size. Part of this complexity is constrained by the number of genes encoding protein kinases, also known as the kinome. The number of encoded protein kinases in free-living eukaryotes ranges from as little as 126 kinases in *Saccharomyces cerevisiae* [3] to ~1000

* Correspondence: andrew.millar@ed.ac.uk
[1]SynthSys and School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JD, UK

Getting metadata

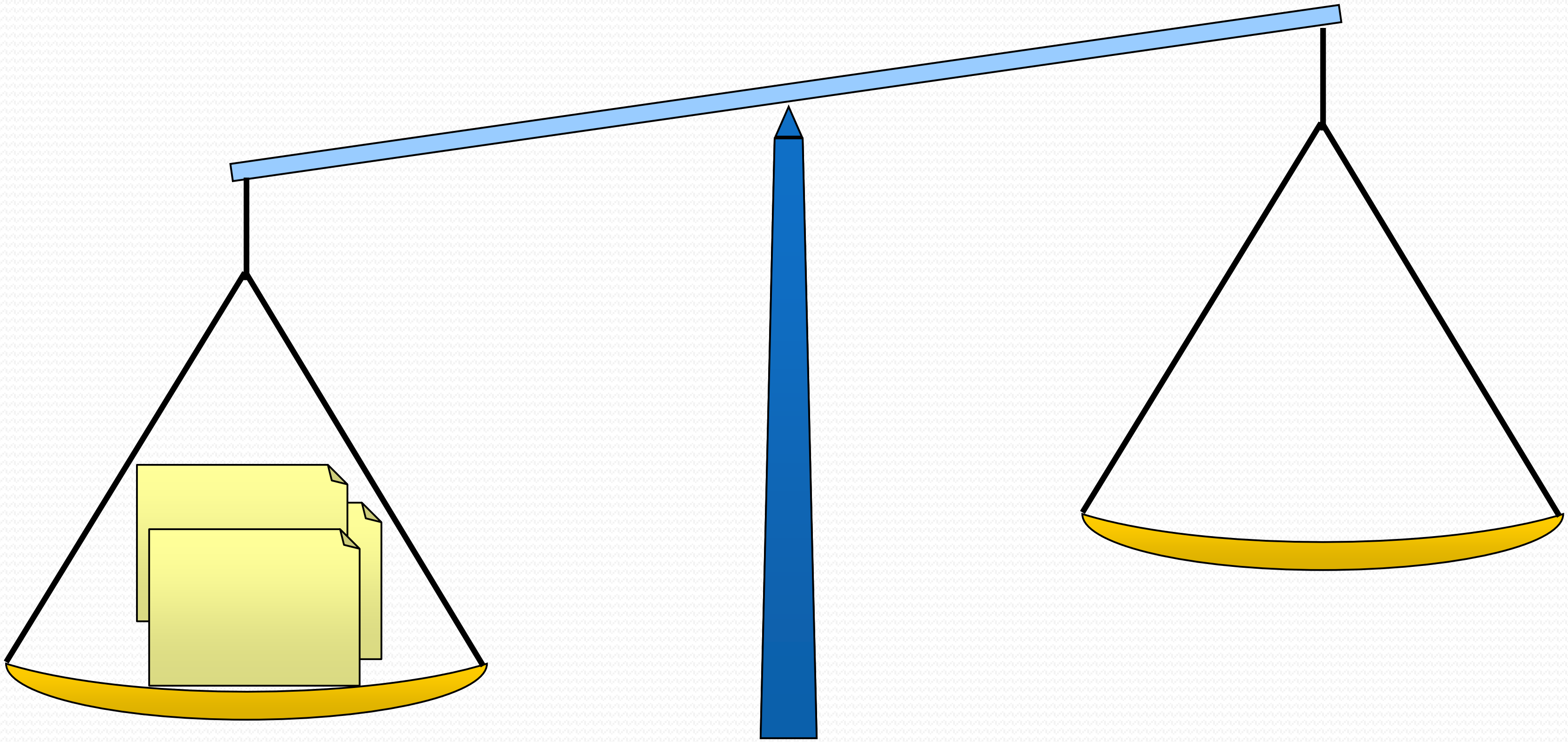


SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Getting metadata

Getting meta data

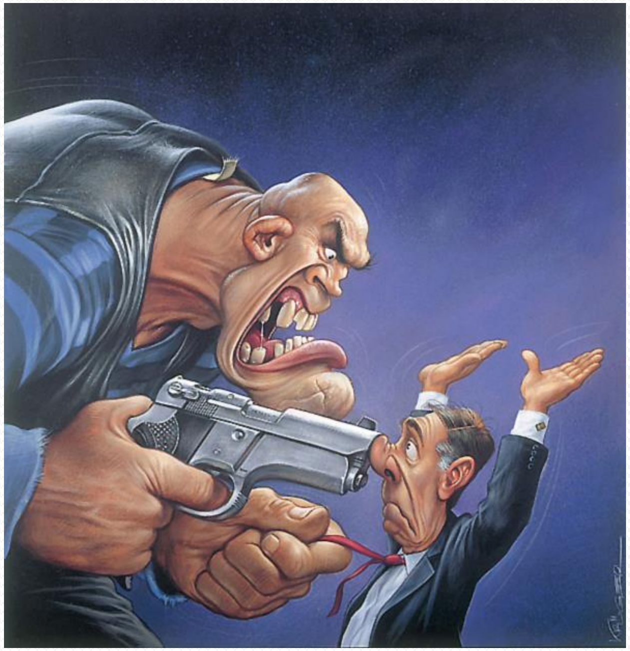
- No one likes describing data for repositories



SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

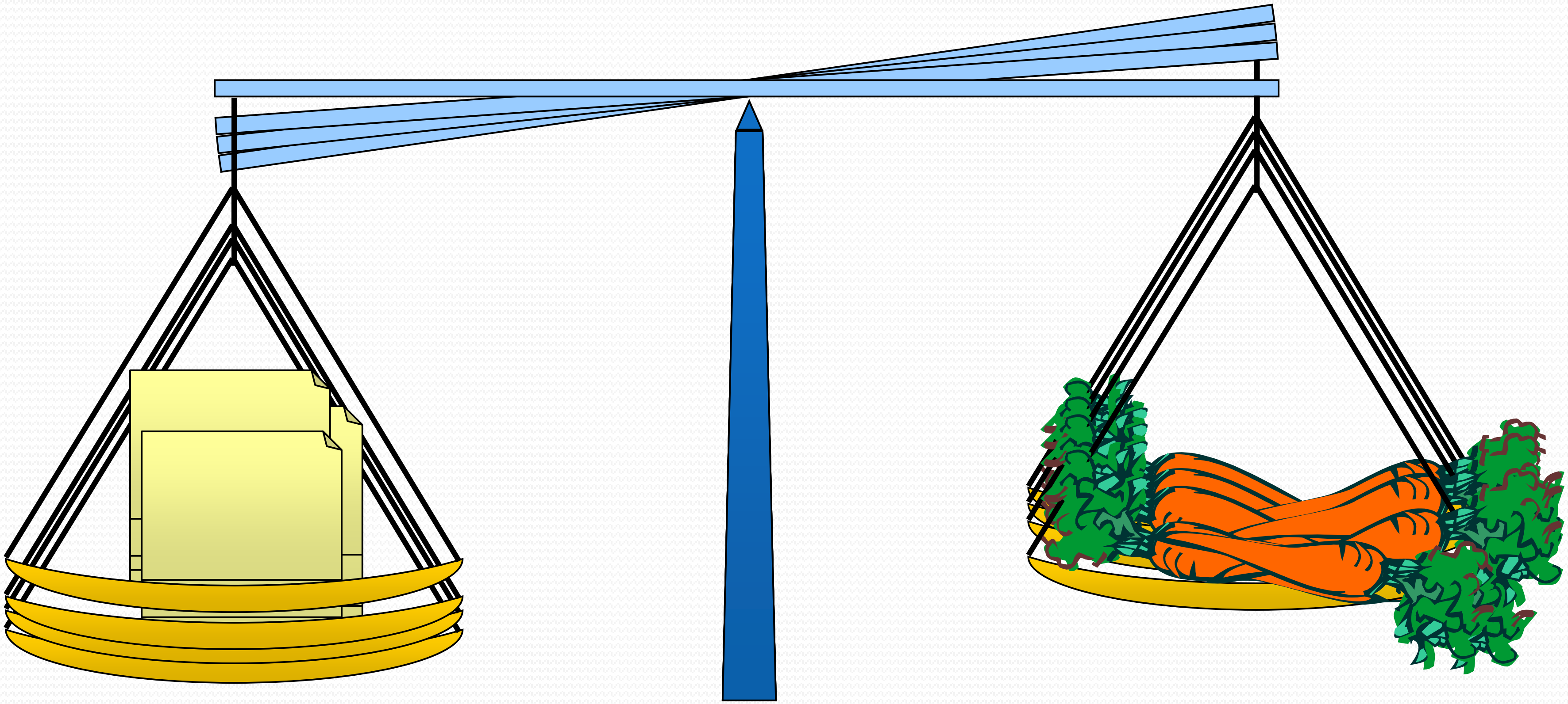
Getting metadata

- No one likes describing data for repositories

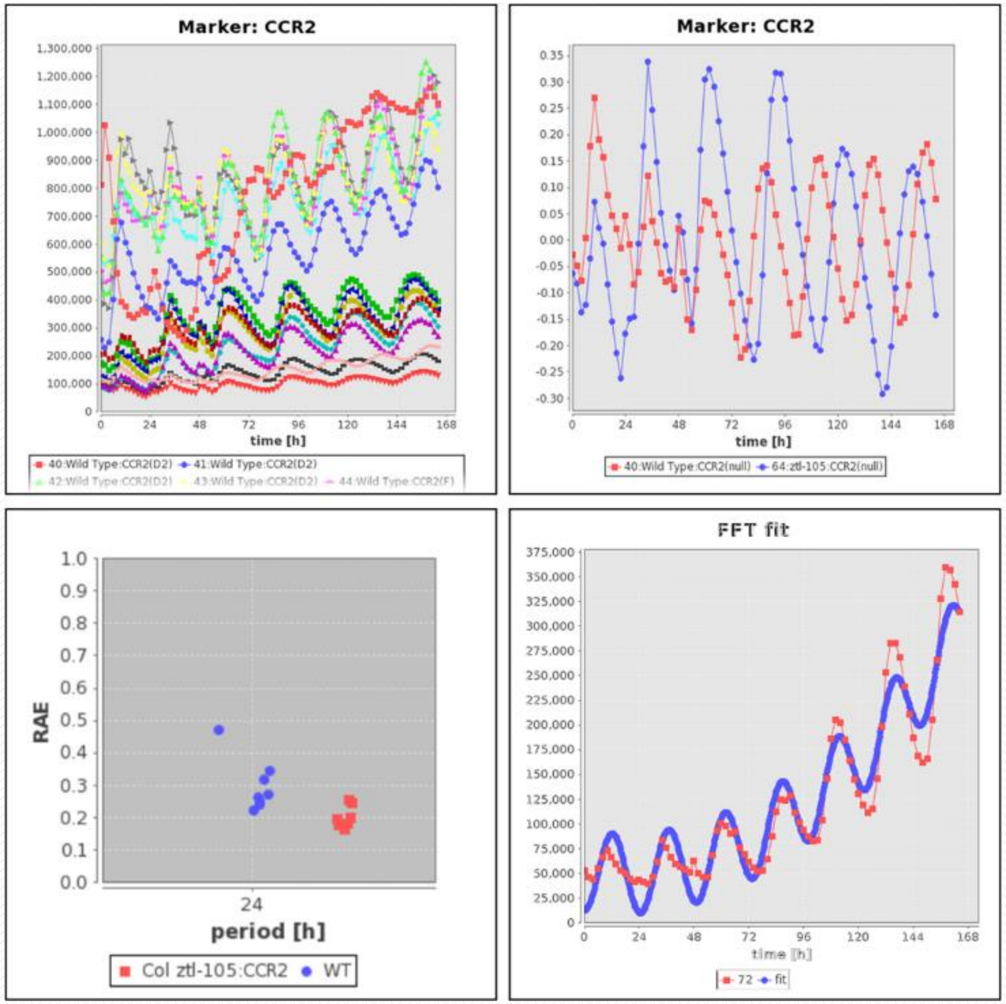


Getting metadata

- No one likes describing data for repositories

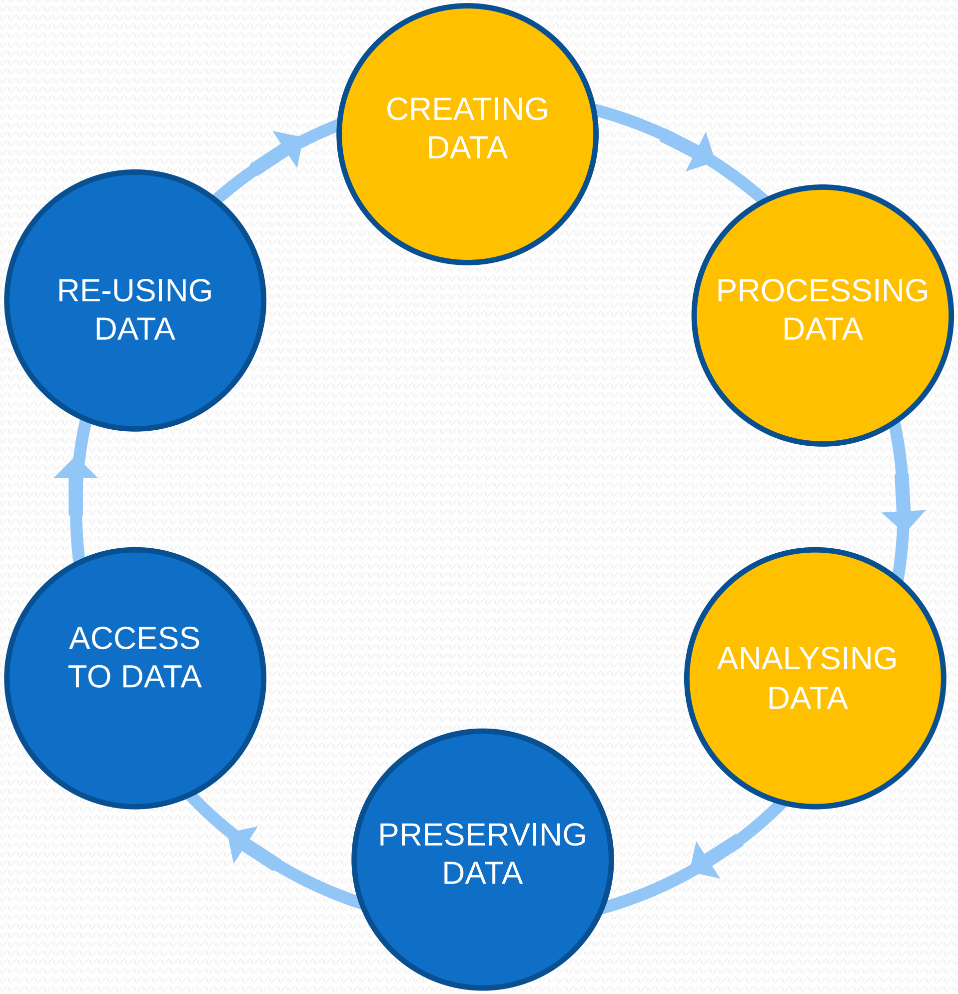


Tipping the balance



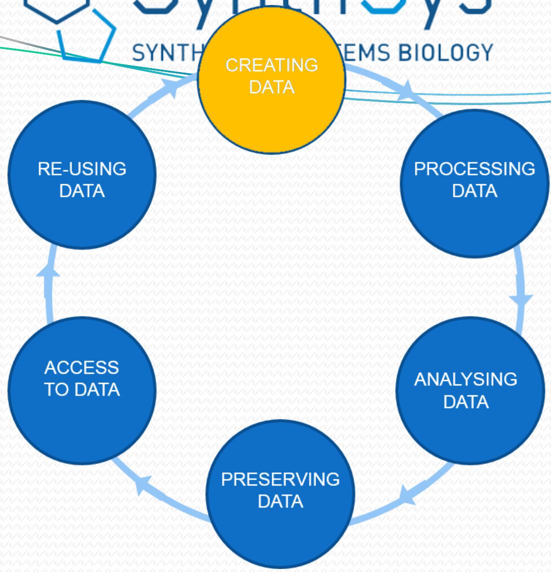
SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Data managment as part of the workflow



General repositories: SEEK and OpenBIS

- fine grained metadata,
- logical structure, data relationships
- customizable
- hooks for data processing

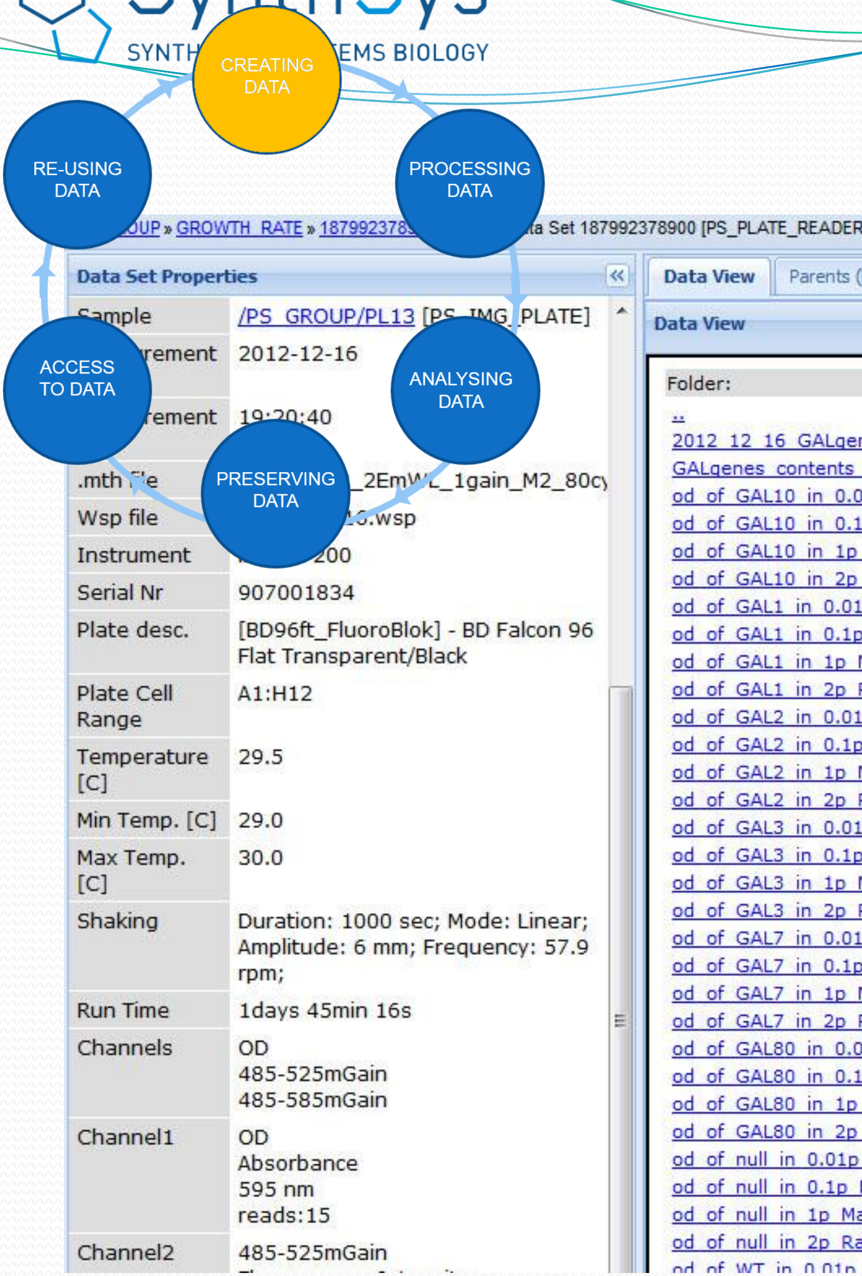


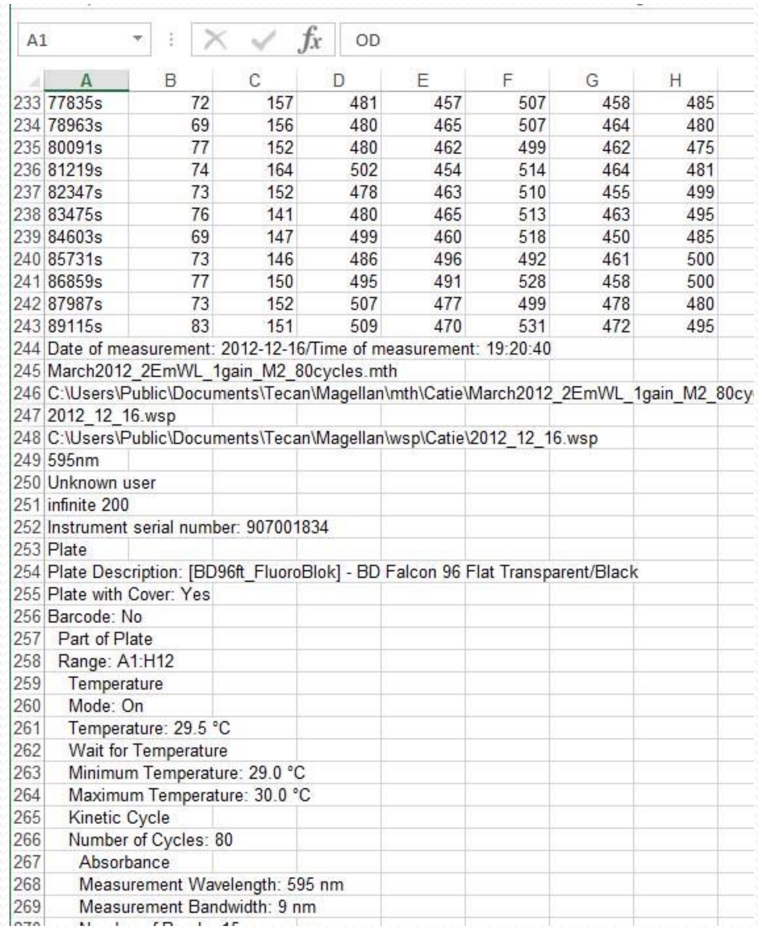
Workflow

A1 | OD

	A	B	C	D	E	F	G	H
233	77835s	72	157	481	457	507	458	485
234	78963s	69	156	480	465	507	464	480
235	80091s	77	152	480	462	499	462	475
236	81219s	74	164	502	454	514	464	481
237	82347s	73	152	478	463	510	455	499
238	83475s	76	141	480	465	513	463	495
239	84603s	69	147	499	460	518	450	485
240	85731s	73	146	486	496	492	461	500
241	86859s	77	150	495	491	528	458	500
242	87987s	73	152	507	477	499	478	480
243	89115s	83	151	509	470	531	472	495
244	Date of measurement: 2012-12-16/Time of measurement: 19:20:40							
245	March2012_2EmWL_1gain_M2_80cycles.mth							
246	C:\Users\Public\Documents\Tecan\Magellan\mth\Catie\March2012_2EmWL_1gain_M2_80cy							
247	2012_12_16.wsp							
248	C:\Users\Public\Documents\Tecan\Magellan\wsp\Catie\2012_12_16.wsp							
249	595nm							
250	Unknown user							
251	infinite 200							
252	Instrument serial number: 907001834							
253	Plate							
254	Plate Description: [BD96ft_FluoroBlok] - BD Falcon 96 Flat Transparent/Black							
255	Plate with Cover: Yes							
256	Barcode: No							
257	Part of Plate							
258	Range: A1:H12							
259	Temperature							
260	Mode: On							
261	Temperature: 29.5 °C							
262	Wait for Temperature							
263	Minimum Temperature: 29.0 °C							
264	Maximum Temperature: 30.0 °C							
265	Kinetic Cycle							
266	Number of Cycles: 80							
267	Absorbance							
268	Measurement Wavelength: 595 nm							
269	Measurement Bandwidth: 9 nm							

Workflow

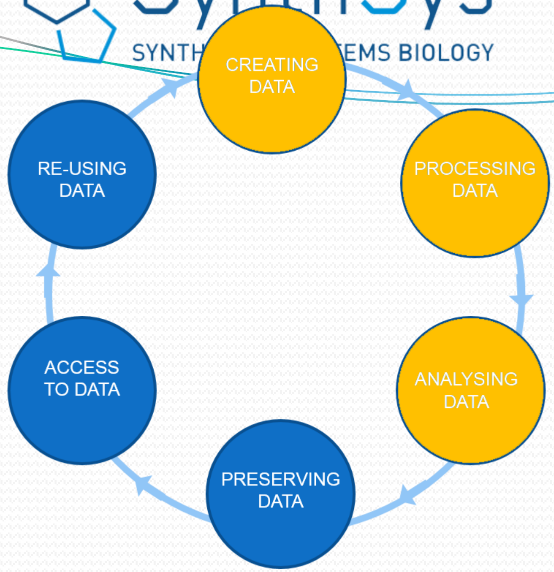




A1 | OD

	A	B	C	D	E	F	G	H
233	77835s	72	157	481	457	507	458	485
234	78963s	69	156	480	465	507	464	480
235	80091s	77	152	480	462	499	462	475
236	81219s	74	164	502	454	514	464	481
237	82347s	73	152	478	463	510	455	499
238	83475s	76	141	480	465	513	463	495
239	84603s	69	147	499	460	518	450	485
240	85731s	73	146	486	496	492	461	500
241	86859s	77	150	495	491	528	458	500
242	87987s	73	152	507	477	499	478	480
243	89115s	83	151	509	470	531	472	495
244	Date of measurement: 2012-12-16/Time of measurement: 19:20:40							
245	March2012_2EmWL_1gain_M2_80cycles.mth							
246	C:\Users\Public\Documents\Tecan\Magellan\mth\Catie\March2012_2EmWL_1gain_M2_80cy							
247	2012_12_16.wsp							
248	C:\Users\Public\Documents\Tecan\Magellan\wsp\Catie\2012_12_16.wsp							
249	595nm							
250	Unknown user							
251	infinite 200							
252	Instrument serial number: 907001834							
253	Plate							
254	Plate Description: [BD96ft_FluoroBlok] - BD Falcon 96 Flat Transparent/Black							
255	Plate with Cover: Yes							
256	Barcode: No							
257	Part of Plate							
258	Range: A1:H12							
259	Temperature							
260	Mode: On							
261	Temperature: 29.5 °C							
262	Wait for Temperature							
263	Minimum Temperature: 29.0 °C							
264	Maximum Temperature: 30.0 °C							
265	Kinetic Cycle							
266	Number of Cycles: 80							
267	Absorbance							
268	Measurement Wavelength: 595 nm							
269	Measurement Bandwidth: 9 nm							

SynthSys



Workflow

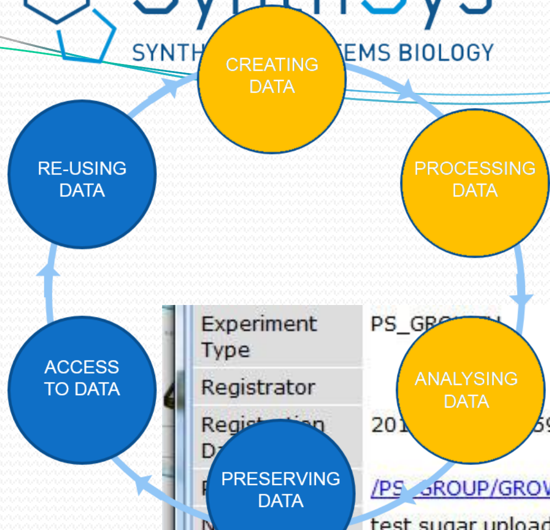
A1 | OD

	A	B	C	D	E	F	G	H
233	77835s	72	157	481	457	507	458	485
234	78963s	69	156	480	465	507	464	480
235	80091s	77	152	480	462	499	462	475
236	81219s	74	164	502	454	514	464	481
237	82347s	73	152	478	463	510	455	499
238	83475s	76	141	480	465	513	463	495
239	84603s	69	147	499	460	518	450	485
240	85731s	73	146	486	496	492	461	500
241	86859s	77	150	495	491	528	458	500
242	87987s	73	152	507	477	499	478	480
243	89115s	83	151	509	470	531	472	495
244	Date of measurement: 2012-12-16/Time of measurement: 19:20:40							
245	March2012_2EmWL_1gain_M2_80cycles.mth							
246	C:\Users\Public\Documents\Tecan\Magellan\mth\Catie\March2012_2EmWL_1gain_M2_80cy							
247	2012_12_16.wsp							

C	D	E	F	G
2	3	4	5	6
null in 2% Raf	WT in 2% Raf	WT in 2% Raf	GAL1 in 2% Raf	GAL2 in 2% Raf
null in 2% Raf	WT in 2% Raf	WT in 2% Raf	GAL1 in 2% Raf	GAL2 in 2% Raf
null in 1% Gal	WT in 1% Gal	WT in 1% Gal	GAL1 in 1% Gal	GAL2 in 1% Gal
null in 1% Gal	WT in 1% Gal	WT in 1% Gal	GAL1 in 1% Gal	GAL2 in 1% Gal
null in 0.1% Gal	WT in 0.1% Gal	WT in 0.1% Gal	GAL1 in 0.1% Gal	GAL2 in 0.1% Gal
null in 0.1% Gal	WT in 0.1% Gal	WT in 0.1% Gal	contaminated	GAL2 in 0.1% Gal
contaminated	WT in 0.01% Gal	WT in 0.01% Gal	GAL1 in 0.01% Gal	GAL2 in 0.01% Gal
null in 0.01% Gal	WT in 0.01% Gal	WT in 0.01% Gal	GAL1 in 0.01% Gal	GAL2 in 0.01% Gal

SynthSys
SYNTH… …EMS BIOLOGY

Workflow



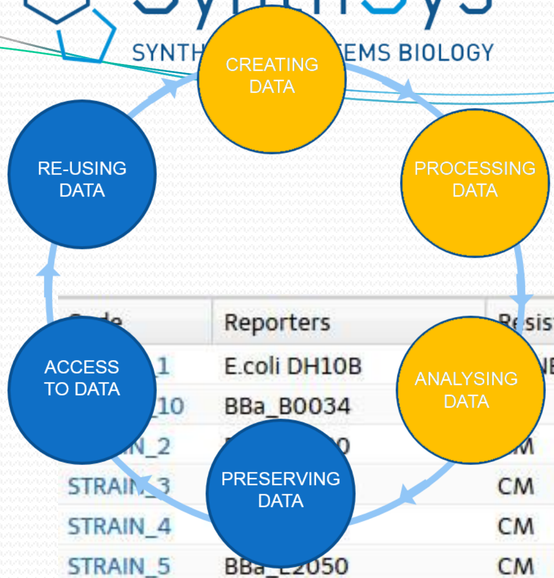
Experiment Type	PS_GR[illegible]
Registrator	
Regi[illegible] D[illegible]	201[illegible]59:08
[illegible]	/PS_GROUP/GROWTH_RATE
N[illegible]	test sugar upload
Desc[illegible]	Testing uploading a sugar file.
Aim	Measure expression of GAL genes in maltose
Authors	Ivan
Submitter	testps
Date	2013-09-12
Measurement date	2012-12-16
Strains	WT GAL1 GAL2 GAL3 GAL7 GAL10 GAL80
Sugars	RAF MAL
Comments	DropBox ERROR: Missing strains in the repository: GAL2,GAL3,GAL7,GAL10,GAL80 Please update the repository and add the correct links to the strains Missing sugars in the repository: MAL

A1 | fx | OD

	A	B	C	D	E	F	G	H
233	77835s	72	157	481	457	507	458	485
234	78963s	69	156	480	465	507	464	480
235	80091s	77	152	480	462	499	462	475
236	81219s	74	164	502	454	514	464	481
237	82347s	73	152	478	463	510	455	499
238	83475s	76	141	480	465	513	463	495
239	84603s	69	147	499	460	518	450	485
240	85731s	73	146	486	496	492	461	500
241	86859s	77	150	495	491	528	458	500
242	87987s	73	152	507	477	499	478	480
243	89115s	83	151	509	470	531	472	495
244	Date of measurement: 2012-12-16/Time of measurement: 19:20:40							
245	March2012_2EmWL_1gain_M2_80cycles.mth							
246	C:\Users\Public\Documents\Tecan\Magellan\mth\Catie\March2012_2EmWL_1gain_M2_80cy							
247	2012_12_16.wsp							

D	E	F	G
3	4	5	6
WT in 2% Raf	WT in 2% Raf	GAL1 in 2% Raf	GAL2 in 2% Raf
WT in 2% Raf	WT in 2% Raf	GAL1 in 2% Raf	GAL2 in 2% Raf
WT in 1% Gal	WT in 1% Gal	GAL1 in 1% Gal	GAL2 in 1% Gal
WT in 1% Gal	WT in 1% Gal	GAL1 in 1% Gal	GAL2 in 1% Gal
WT in 0.1% Gal	WT in 0.1% Gal	GAL1 in 0.1% Gal	GAL2 in 0.1% Gal
WT in 0.1% Gal	WT in 0.1% Gal	contaminated	GAL2 in 0.1% Gal
WT in 0.01% Gal	WT in 0.01% Gal	GAL1 in 0.01% Gal	GAL2 in 0.01% Gal
WT in 0.01% Gal	WT in 0.01% Gal	GAL1 in 0.01% Gal	GAL2 in 0.01% Gal

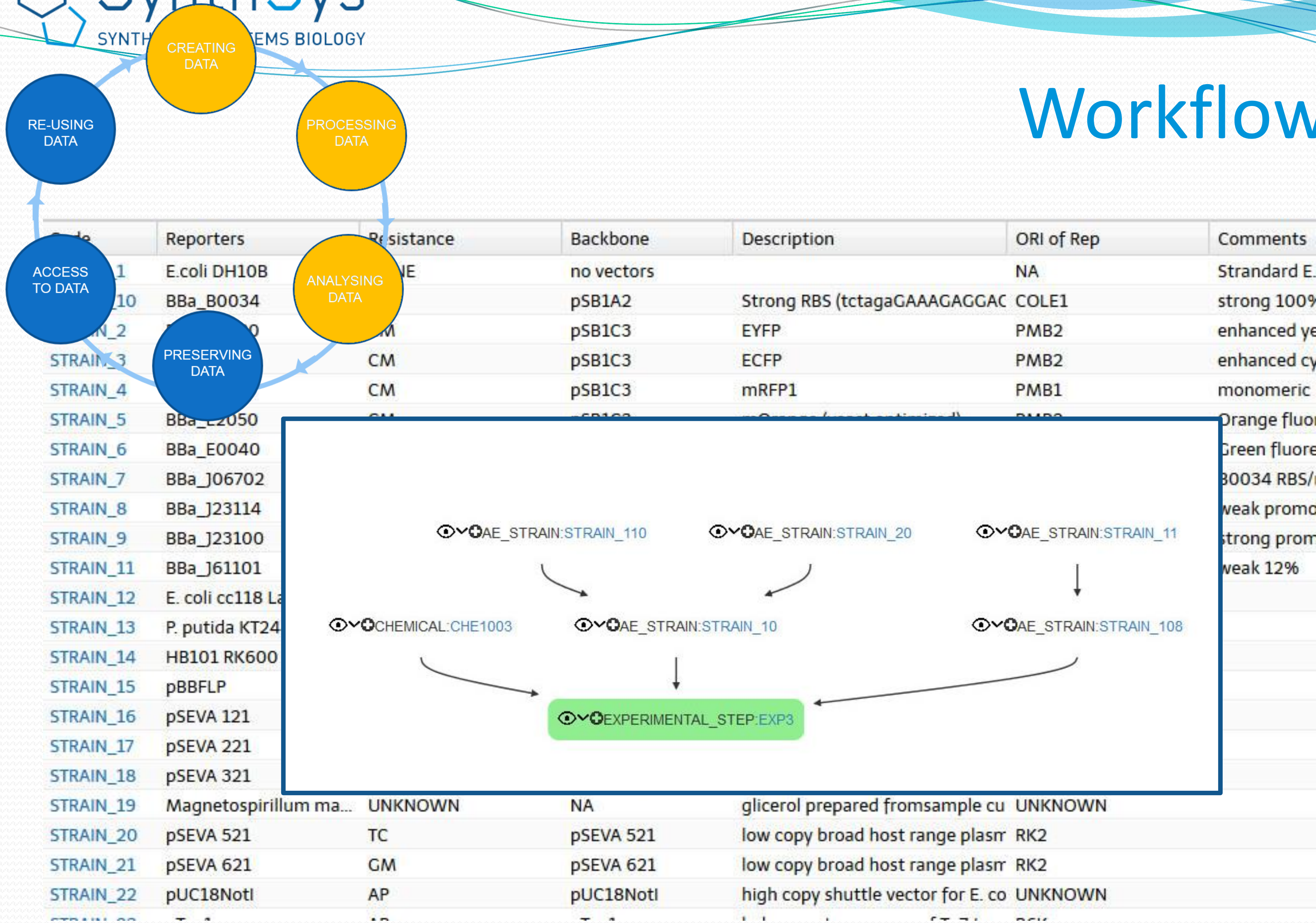
Workflow



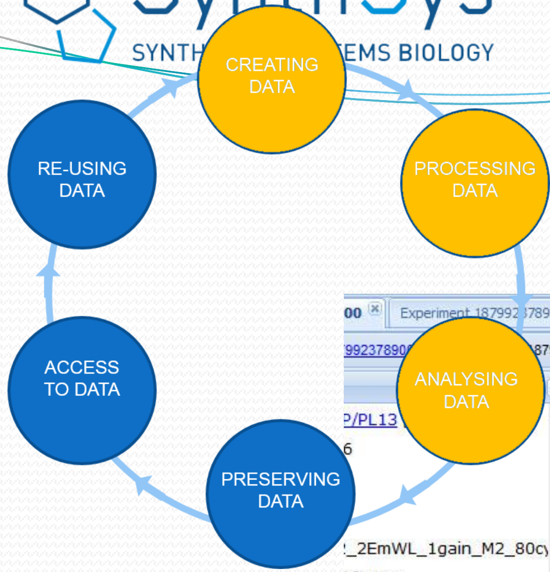
Code	Reporters	Resistance	Backbone	Description	ORI of Rep	Comments
[illegible]	E.coli DH10B	[illegible]	no vectors		NA	Strandard E.
[illegible]	BBa_B0034	[illegible]	pSB1A2	Strong RBS (tctagaGAAAGAGGAC	COLE1	strong 100%
[illegible]	[illegible]	[illegible]	pSB1C3	EYFP	PMB2	enhanced ye
STRAIN_3	[illegible]	CM	pSB1C3	ECFP	PMB2	enhanced cy
STRAIN_4	[illegible]	CM	pSB1C3	mRFP1	PMB1	monomeric
STRAIN_5	BBa_E2050	CM	pSB1C3	mOrange (yeast optimized)	PMB2	Orange fluo
STRAIN_6	BBa_E0040	AP	pSB1A2	GFPmut3b	COLE1	Green fluore
STRAIN_7	BBa_J06702	AP	pSB1A2	mCherry (yeast version also)	COLE1	B0034 RBS/
STRAIN_8	BBa_J23114	AP	pJ61002	Weak promoter (tttatggctagctca	COLE1	weak promo
STRAIN_9	BBa_J23100	AP	pJ61002	Strong promoter (ttgacggctagct	COLE1	strong prom
STRAIN_11	BBa_J61101	AP	pSB1A2	weak RBS (tctagaGAAAGACAGGA	COLE1	weak 12%
STRAIN_12	E. coli cc118 Lambda ...	NONE	no vectors	cloning strain to maintain RK6 pl	NA	
STRAIN_13	P. putida KT2440	NONE	no vectors	standard P. putida strain without	NA	
STRAIN_14	HB101 RK600	CM	no vectors	E. coli HB101 helper strain to mo	NA	
STRAIN_15	pBBFLP	TC	pBBFLP	Plasmid source Flipase	UNKNOWN	
STRAIN_16	pSEVA 121	AP	pSEVA 121	low copy broad host range plasm	RK2	
STRAIN_17	pSEVA 221	KM	pSEVA 221	low copy broad host range plasm	RK2	
STRAIN_18	pSEVA 321	CM	pSEVA 321	low copy broad host range plasm	RK2	
STRAIN_19	Magnetospirillum ma...	UNKNOWN	NA	glicerol prepared fromsample cu	UNKNOWN	
STRAIN_20	pSEVA 521	TC	pSEVA 521	low copy broad host range plasm	RK2	
STRAIN_21	pSEVA 621	GM	pSEVA 621	low copy broad host range plasm	RK2	
STRAIN_22	pUC18NotI	AP	pUC18NotI	high copy shuttle vector for E. co	UNKNOWN	

SynthSys

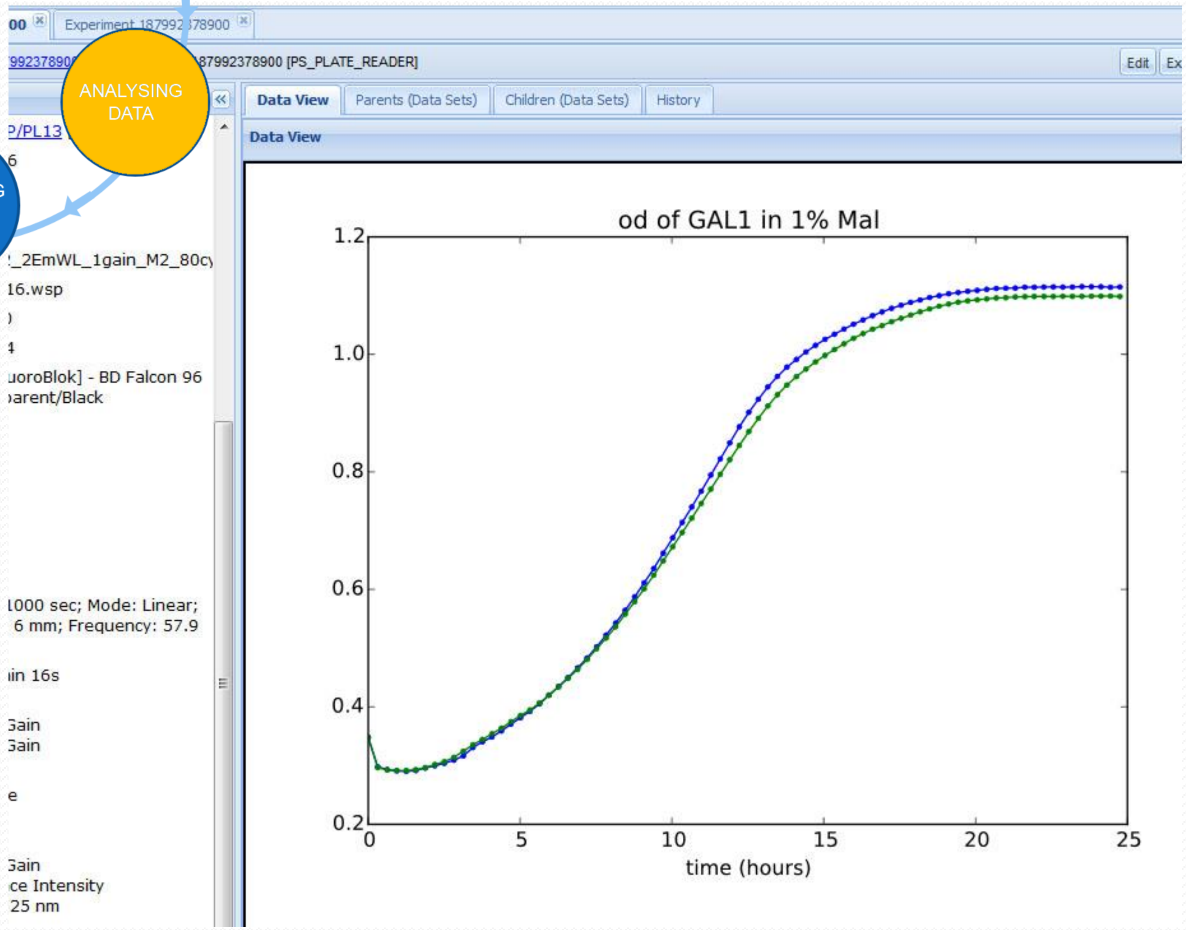
Workflow



Workflow



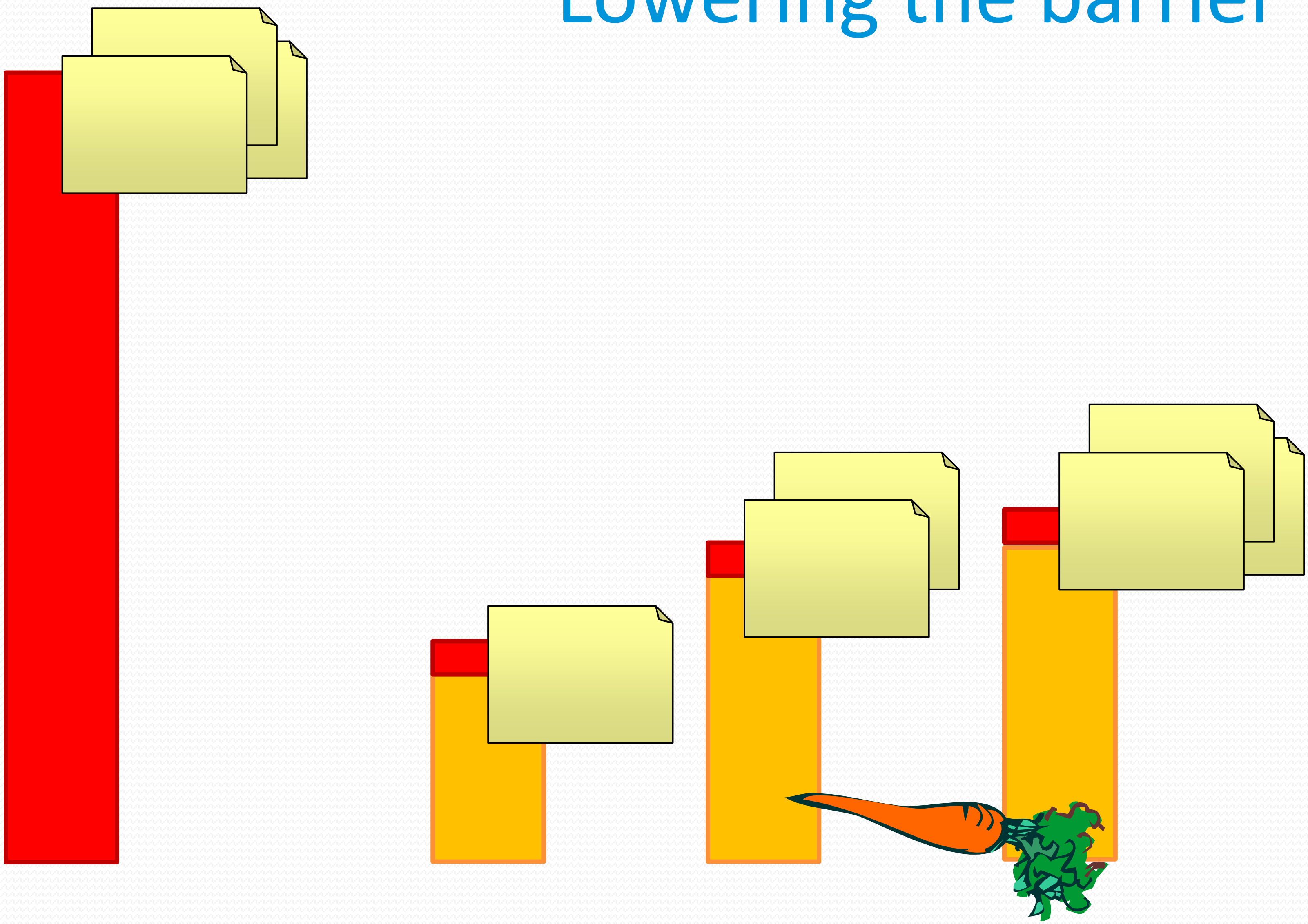
CREATING DATA
PROCESSING DATA
ANALYSING DATA
PRESERVING DATA
ACCESS TO DATA
RE-USING DATA



Experiment 187992378900
87992378900 [PS_PLATE_READER]
Edit
Data View
Parents (Data Sets)
Children (Data Sets)
History
Data View
od of GAL1 in 1% Mal
1.2
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
time (hours)

SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Lowering the barrier



SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Successful data management?

- is part of the research workflow not an extra burden
- solves user problems not creates new ones
- provides extra value for the user
- focuses on both data consumers and producers

The missing link