

From experimental planning to data publication: **the ISA infrastructure** and case studies in toxicology

Alejandra González-Beltrán, Ph.D

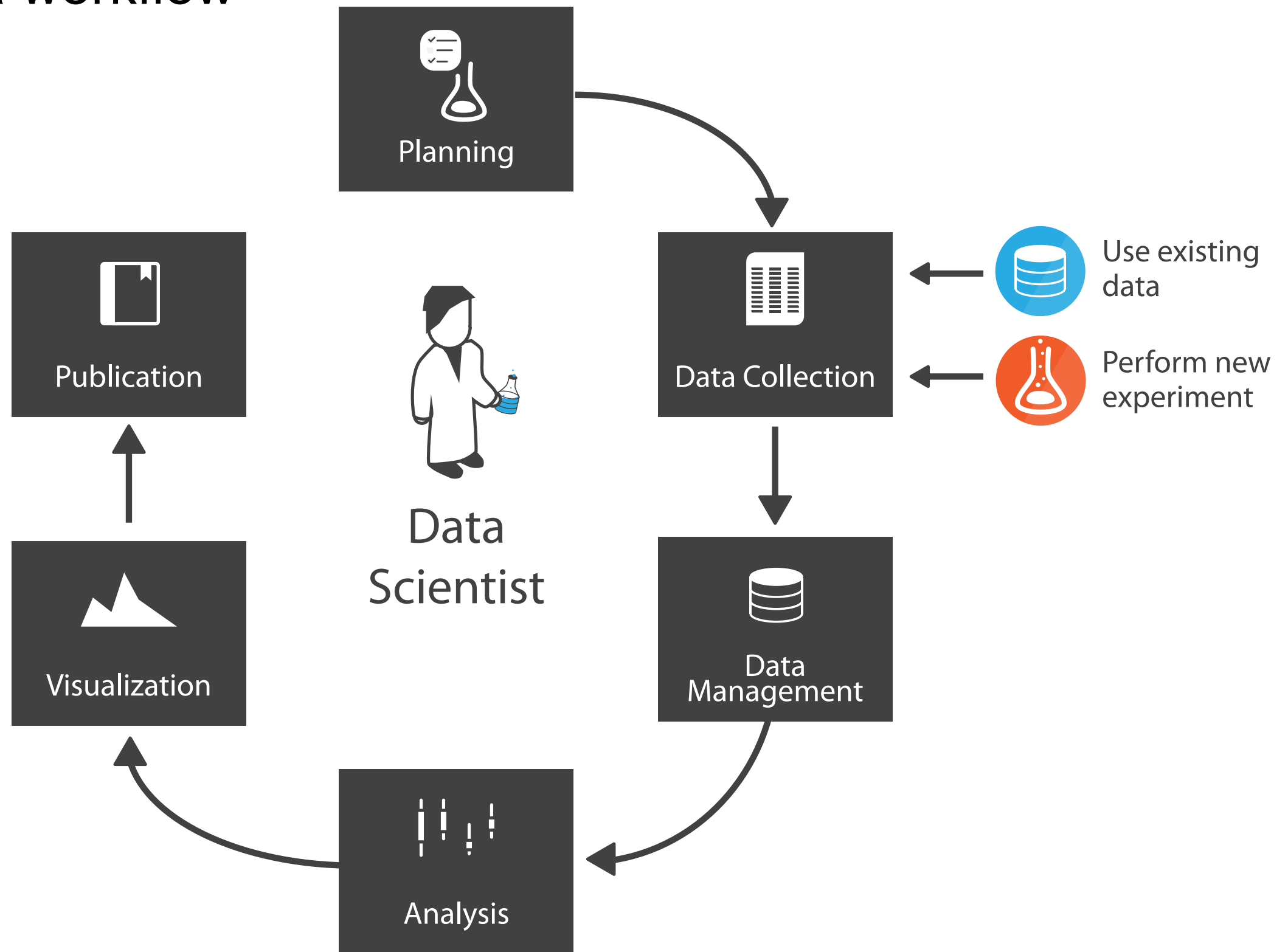
University of Oxford e-Research Centre, UK

alejandra.gonzalezbeltran@oerc.ox.ac.uk

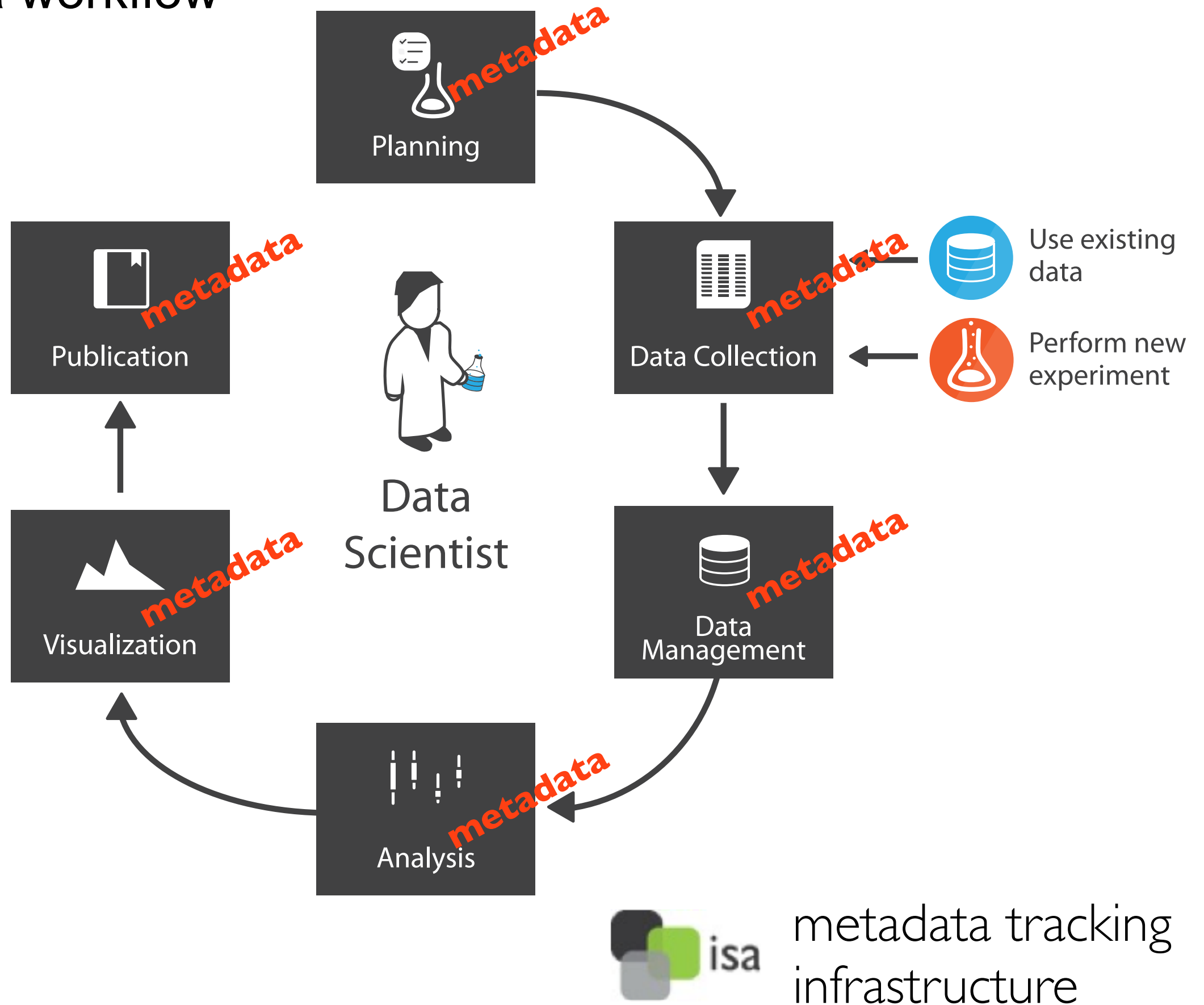
OpenTox Europe - Mainz, Germany - 30th September, 2013

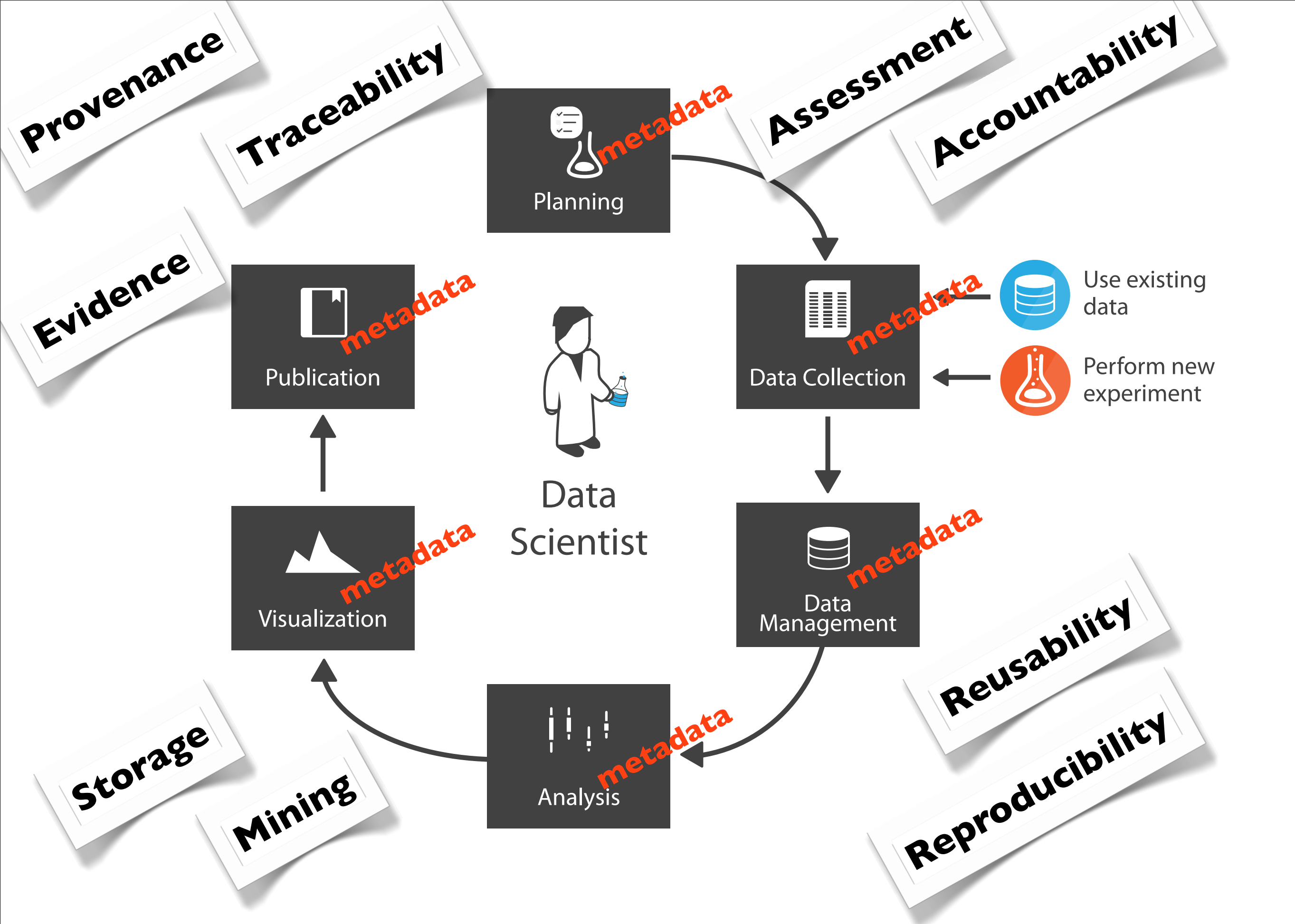


The data workflow



The data workflow



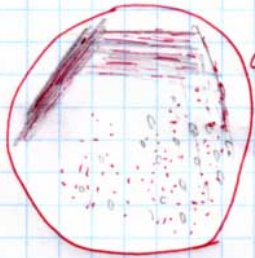


Purple Non-sulfurs

Jim Kim

11/16/06

the TA took the anaerobe jars w/ the PMS plates out of the lighted incubator, & opened it. the smell is foul-like a dead animal. My plate is pretty damp, & smells very bad, like vomit.



← plate

there are a variety of clear, white & yellow-ish colonies
- lots of tiny dark red colonies - too small to make out clearly - opaque, round & maybe raised

→ zone

1050X image of a red colony (wet mount)



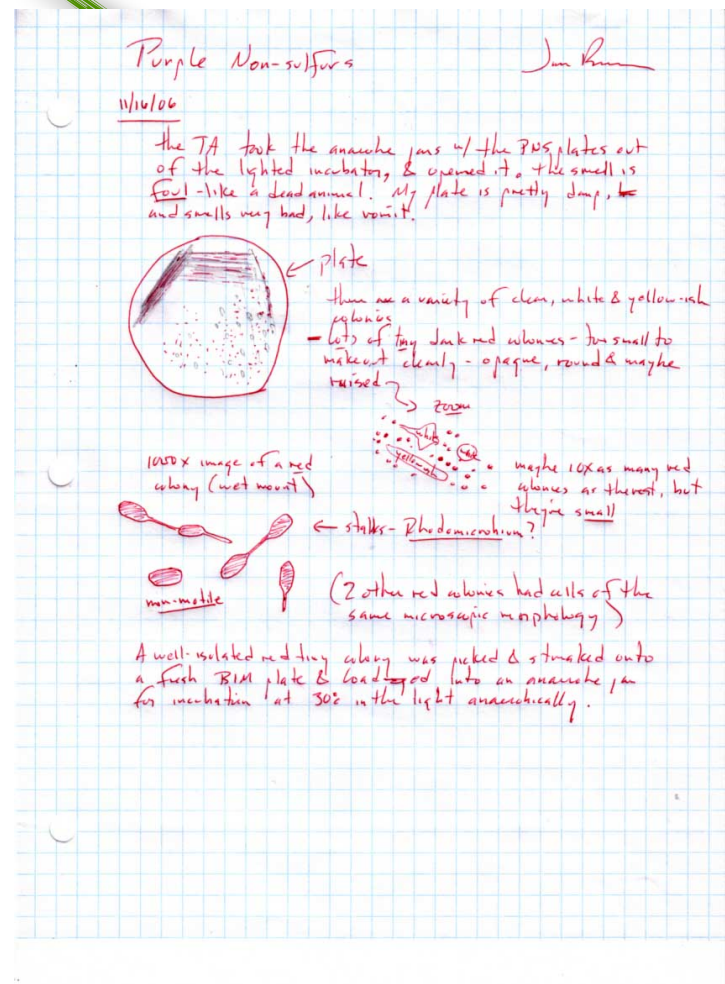
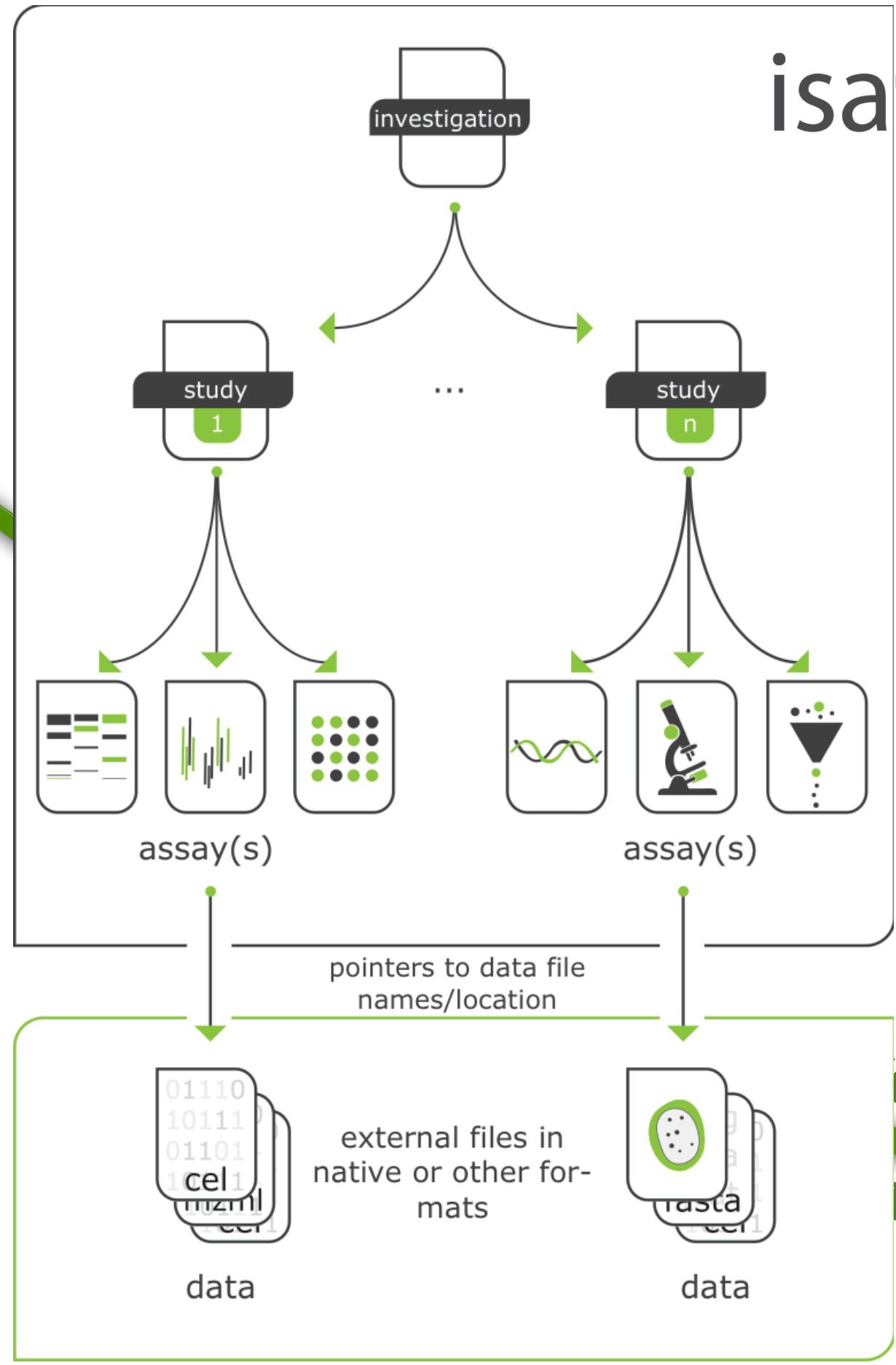
← stalks - Rhodospirillum?

non-motile

(2 other red colonies had cells of the same microscopic morphology)

A well-isolated red tiny colony was picked & streaked onto a fresh BIM plate & loaded into an anaerobe jar for incubation at 30°C in the light anaerobically.

structure
semantics



iinvestigation
study
assay

structure
 data files

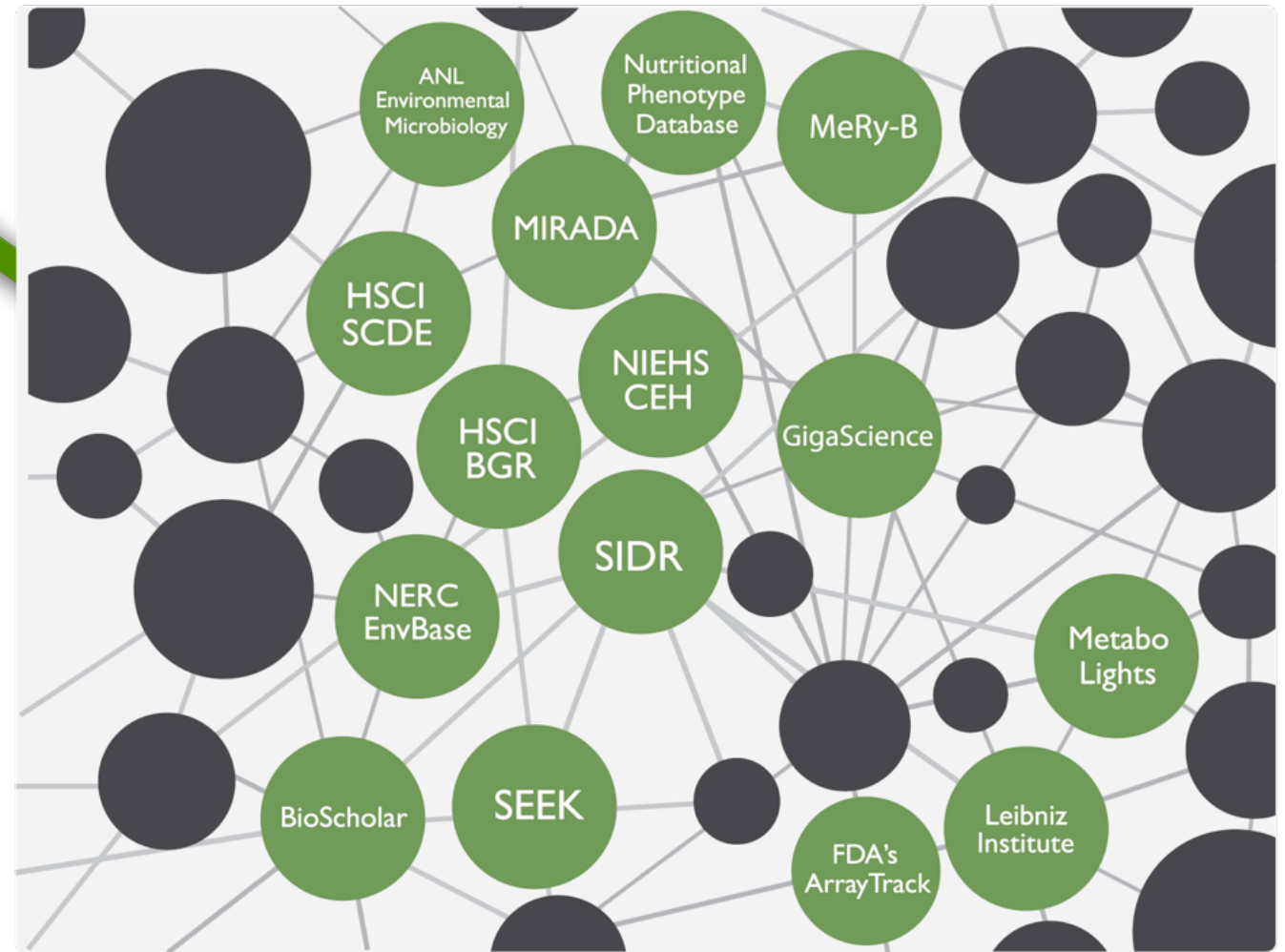
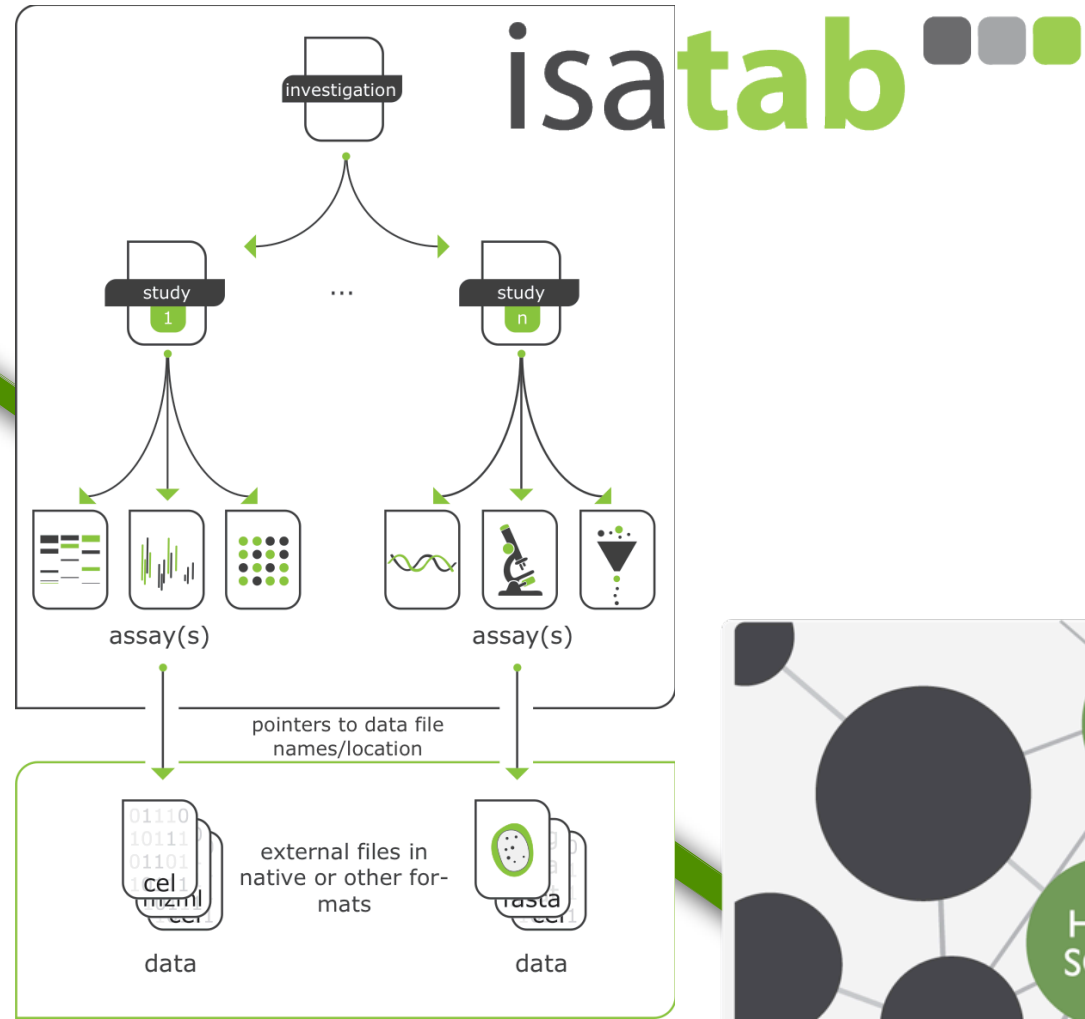
Purple Non-sulfur
Jan Rhee

Notes
 the TA took the anaerobic jars w/ the 30% plates out of the lighted incubator, & examined it w/ the small is food - like a dead animal. My plate is pretty damp, but not really very bad, like vomit.

← plate
 - there are a variety of clear, white & yellowish colonies
 - lots of big dark red colonies - too small to make out clearly - opaque, round & maybe raised?

1000x image of a red colony (next board)
 ← stiffer - Rhodospirillum?
 (2 other red colonies had cells of the same microscopic morphology)

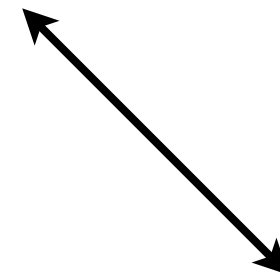
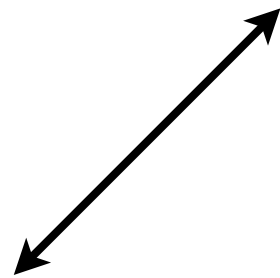
A well-isolated red colony was picked & streaked onto a fresh 30% plate & incubated into an anaerobic jar for incubation at 30°C in the light anaerobically.



The isa infrastructure

isatab

generic format for experimental
description and data exchange



isacommons
isacommons.org

community engagement



isatools

open source software tools



Towards interoperable bioscience data: Presenting the ISA Commons, authored by more than 50 collaborators at over 30 scientific organizations around the globe.

Sansone et al, 2012
Nature Genetics



ISA software suite:
Overview of ISA-Tab and first set of tools

Rocca-Serra et al, 2010
Bioinformatics

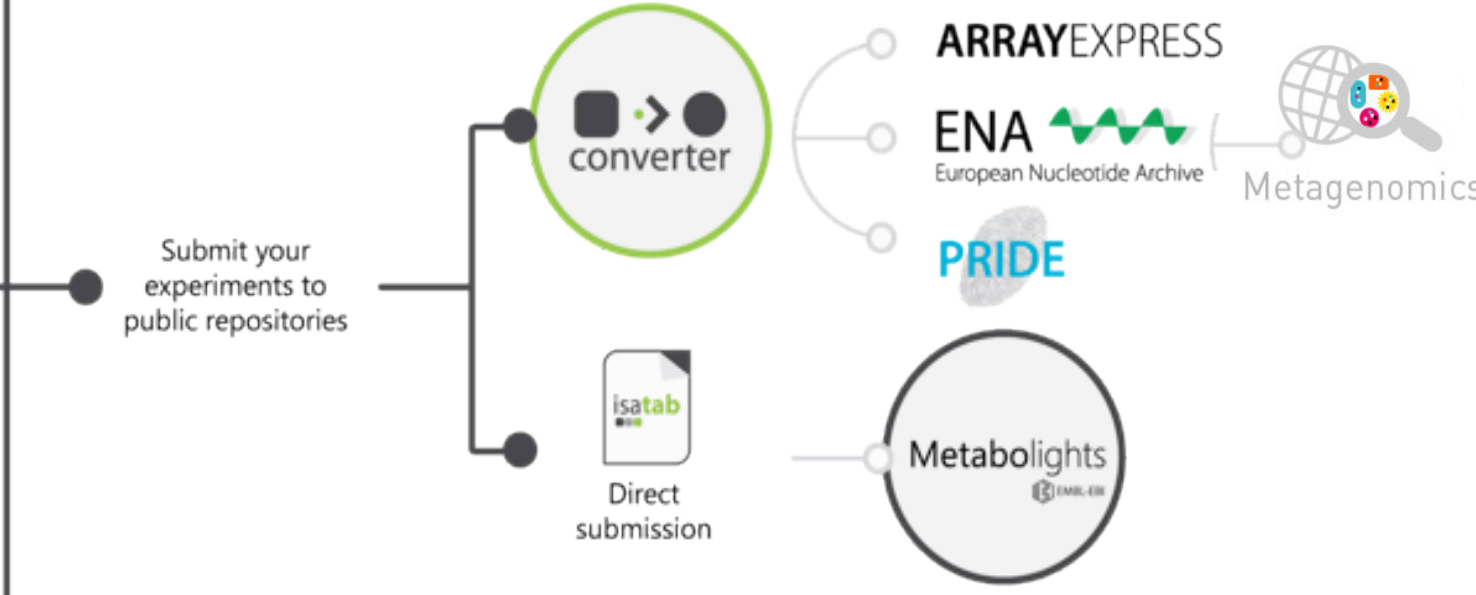
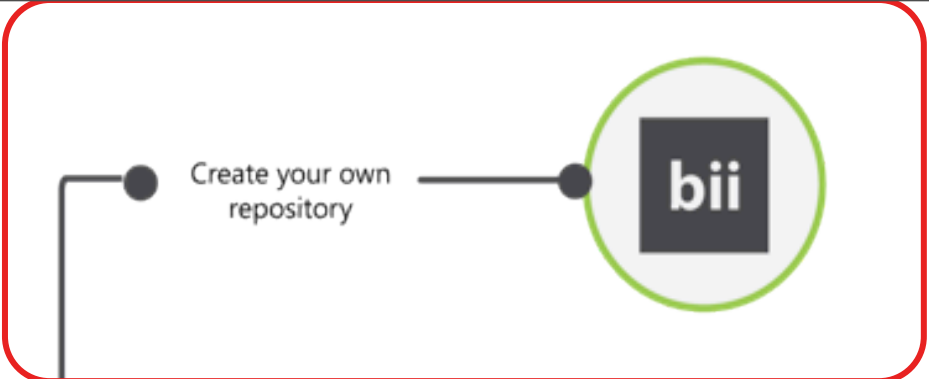


store and browse, locally or publicly

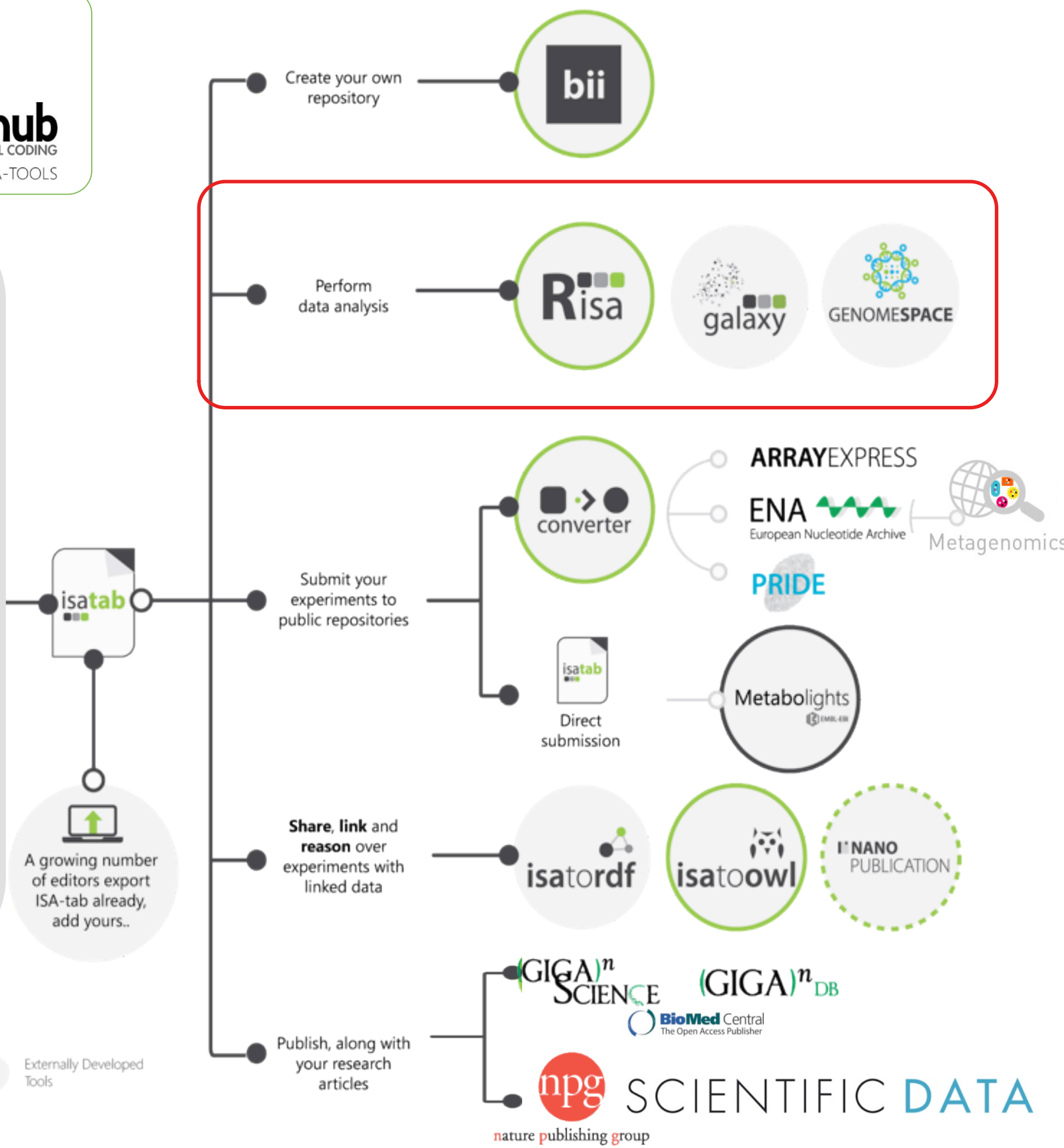
create your own repository to search and browse the experimental description and associated data making it close or open.

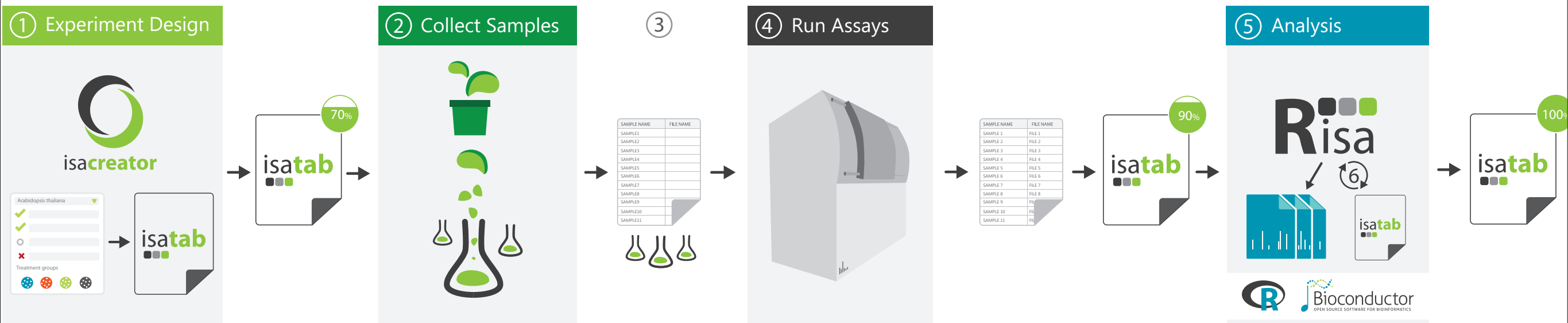


A growing number of editors export ISA-tab already, add yours..



analyse with existing tools
upload experimental description and associated data to a growing number of well-known analysis systems, ISA connects with.



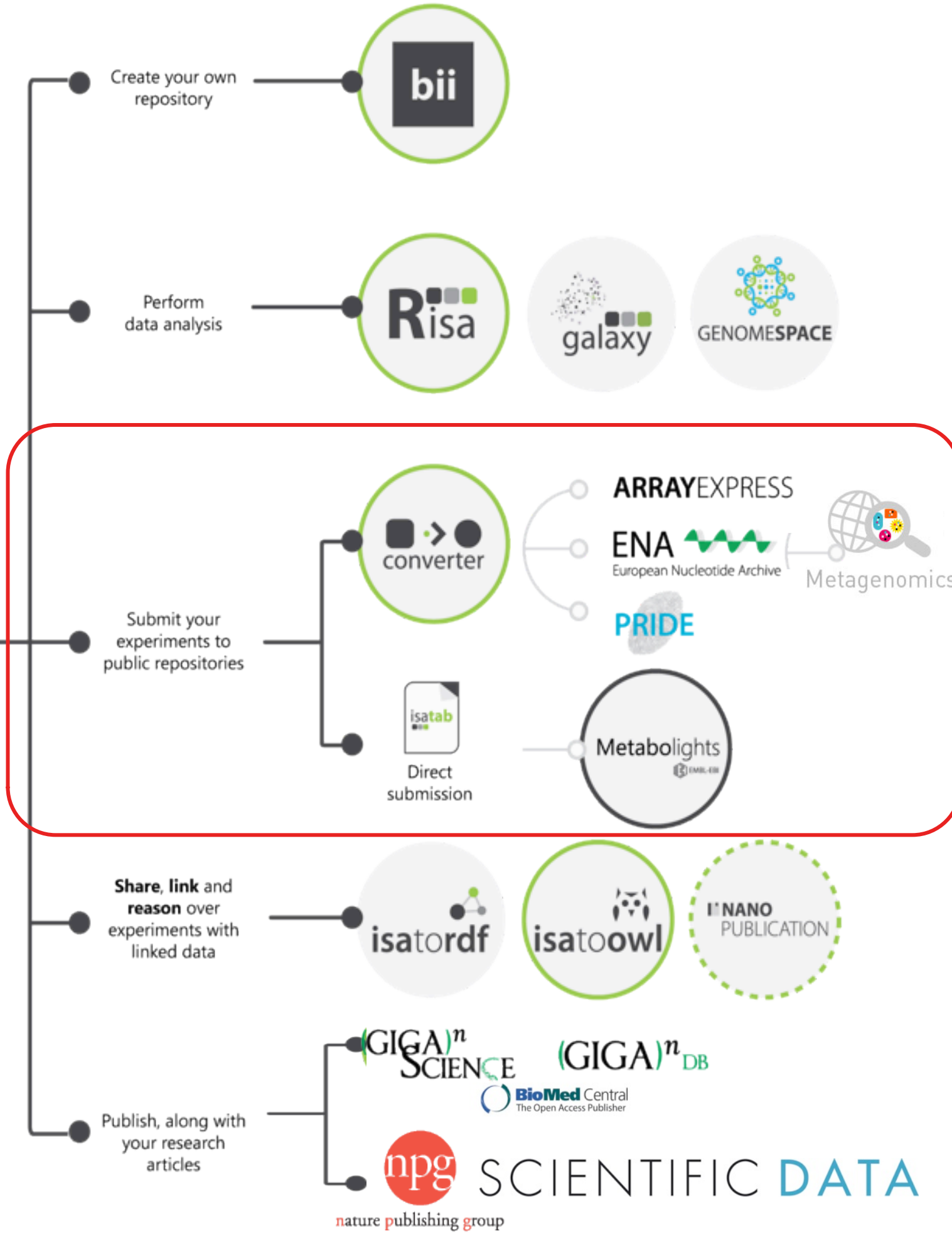
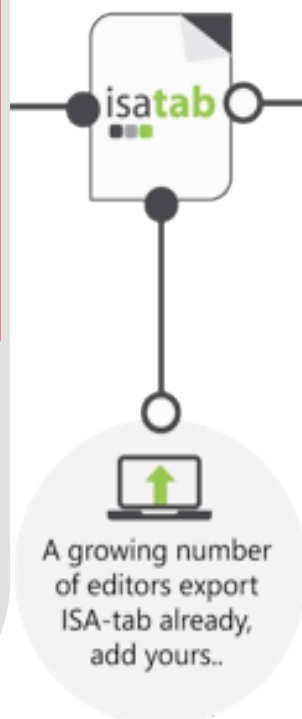


Parses ISA-Tab datasets into R objects, allowing to update them and save them after analysis.

Bridges the ISA-Tab metadata to analysis pipelines of specific assay types, by building objects for use in other R packages downstream: currently considering mass spectrometry (*xmcs* package, *xcmsSet*) and DNA microarray (*Biobase* package, *ExpressionSet*)

Suggests packages in BioConductor that might be relevant for an assay type, according to the BioCViews annotations.

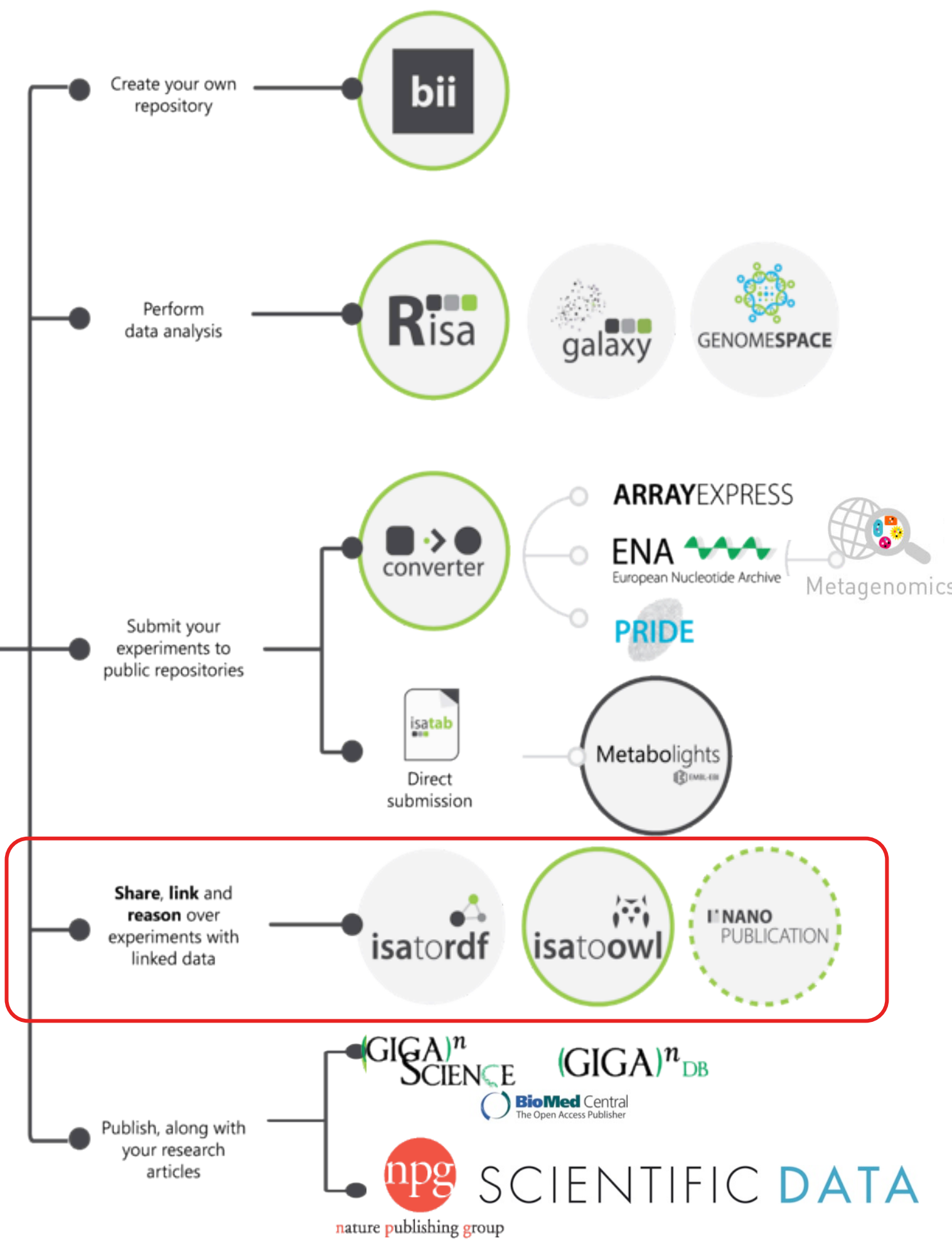
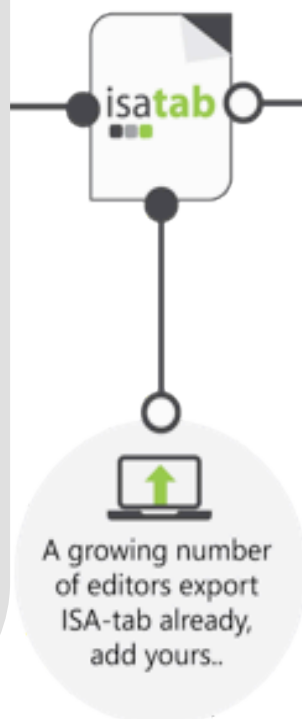
submit to public repositories
 when required, reformat the experiments for submission to supported public repositories or directly export to those already using ISA-Tab.





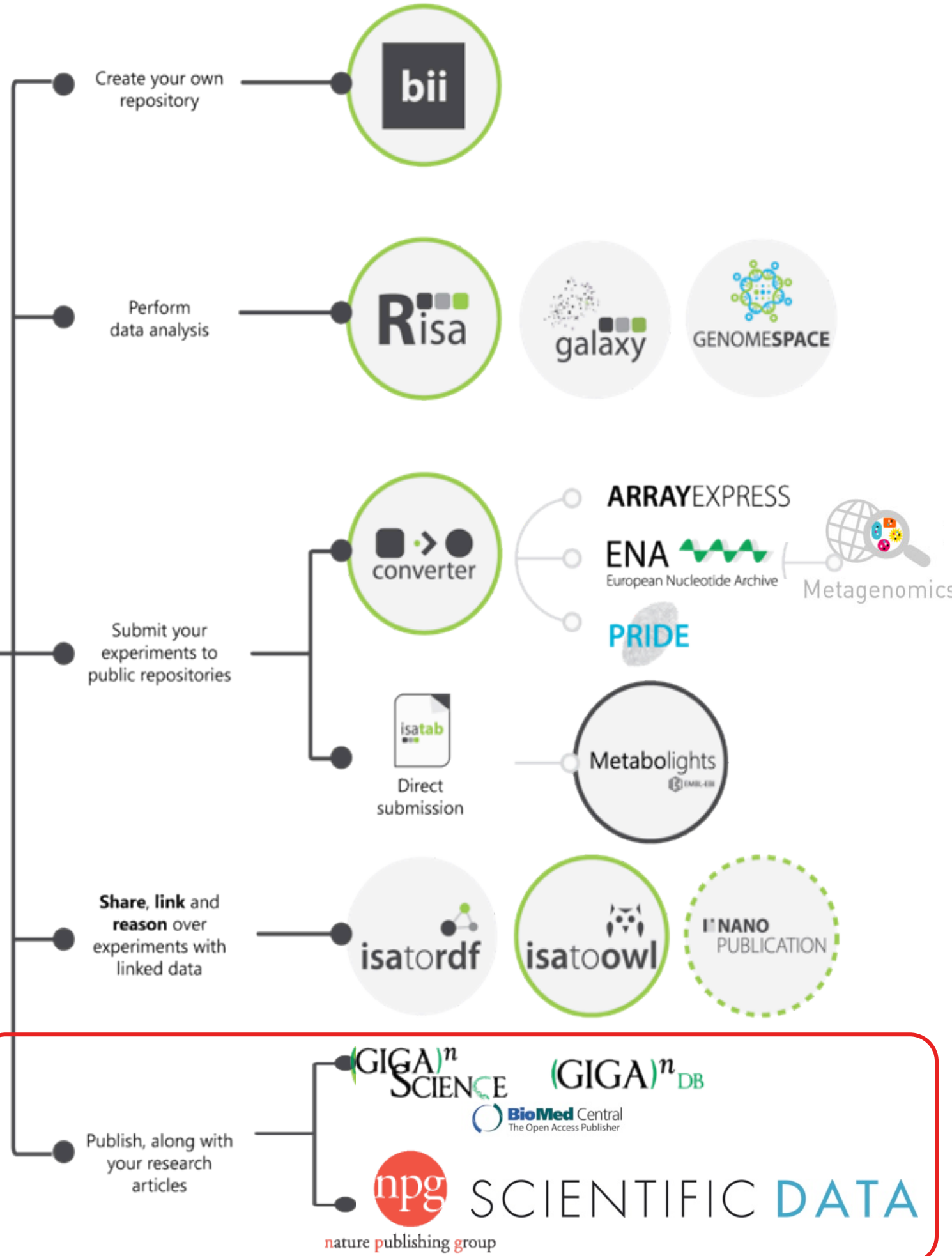
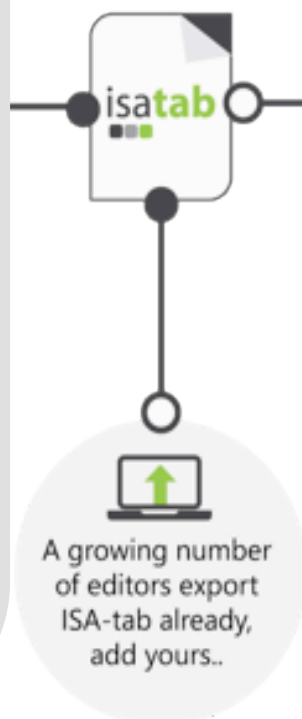
release, reason and nanopublish

explore how to reason over your experiments, open them to the linked data universe, or publish nano-statements of your discoveries.









publish data along your article

directly export your experiments to the new generation of data journals, accepting submission in ISA-Tab.




Data Publication with isat**ab**^{'''}

Helping you publish, discover,
and reuse research data

 Credit Credit, through a citable publication, for depositing & sharing your data	 Reuse Complete, curated & standardized descriptions enable the reuse of your data	 Quality Rigorous community based peer review
 Discovery Find datasets relevant to your research	 Open Promotes & endorses open science principles & available to all through a Creative Commons license	 Service In-house curation, rapid peer review & publication of your data descriptions

 Sign up for Scientific Data e-alert

 Scientific Data Updates

Introducing Scientific Data Updates
April 4, 2013

If not now then when – my view from within
April 3, 2013

[Press Release] NPG to launch Scientific Data to help scientists publish and reuse research data
April 3, 2013

Follow Us



nature publishing group 

www.nature.com/scientificdata

<http://www.nature.com/scientificdata/>

- New open-access, online-only publication for descriptions of scientifically valuable datasets
- Only content type: **Data Descriptor, narrative + structured parts**
- Initially focused on the **life, environmental** and **biomedical** sciences
- Data Descriptor will be complementary to traditional research journals and data repositories
- Designed to foster data sharing and reuse, and ultimately to accelerate scientific discovery

Data Publication with **isatab** 

Helping you publish, discover, and reuse research data

Sign up for Scientific Data e-alert

Scientific Data Updates

Introducing Scientific Data Updates
April 4, 2013

If not now then when – my view from within
April 3, 2013

[Press Release] NPG to launch Scientific Data to help scientists publish and reuse research data
April 3, 2013

Follow Us



Credit
Credit, through a citable publication, for depositing & sharing your data

Reuse
Complete, curated & standardized descriptions enable the reuse of your data

Quality
Rigorous community based peer review

Discovery
Find datasets relevant to your research

Open
Promotes & endorses open science principles & available to all through a Creative Commons license

Service
In-house curation, rapid peer review & publication of your data descriptions



- New open-access, online-only publication for descriptions of scientifically valuable datasets
- Only content type: **Data Descriptor, narrative + structured parts**
- Initially focused on the **life, environmental** and **biomedical** sciences
- Data Descriptor will be complementary to traditional research journals and data repositories
- Designed to foster data sharing and reuse, and ultimately to accelerate scientific discovery

www.nature.com/scientificdata

<http://www.nature.com/scientificdata/>

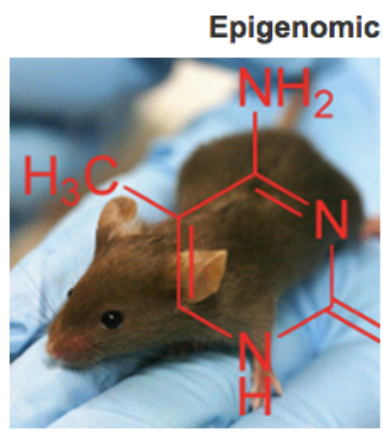
Data Publication with isatab



Resources for the MeDUSA (Methylated DNA Utility for Sequence Analysis) MeDIP-seq computational analysis pipeline for the identification of differentially methylated regions, and associated methylome data from 18 wild-type and mutant mouse ES, NP and MEF cells.

Wilson, G; Dharmi, P; Saito, Y; Cortazar, D; Kunz, C; Schär, P; Beck, S (2012): Resources for the MeDUSA (Methylated DNA Utility for Sequence Analysis) MeDIP-seq computational analysis pipeline for the identification of differentially methylated regions, and associated methylome data from 18 wild-type and mutant mouse ES, NP and MEF cells. GigaScience. <http://dx.doi.org/10.5524/100035>

Here we present 18 genome-wide DNA methylation profiles of wild type and Thymine DNA glycosylase (*Tdg*) knockout cells, which serve as an excellent murine methylome resource. The 18 samples represent 6 biological cohorts: 6 samples were derived from mouse embryonic stem cells (3 *Tdg*^{+/+}, 3 *Tdg*^{-/-}), 6 samples were from mouse neural precursor cells (3 *Tdg*^{+/+}, 3 *Tdg*^{-/-}) and 6 samples were obtained from mouse embryonic fibroblasts (3 *Tdg*^{+/+}, 3 *Tdg*^{-/-}).



DOI	File Name	Sample ID	File Type
10.5524/100035	i_Investigation.txt		ISA-Tab
10.5524/100035	s_GIGAScience-ISA1.txt		ISA-Tab
10.5524/100035	a_GIGAScience-ISA1.txt		ISA-Tab

<http://gigasciencejournal.com>

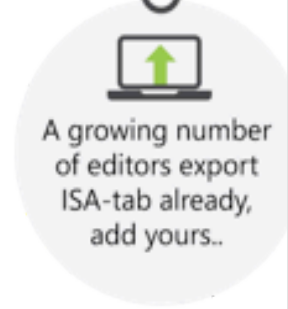
Describe & curate your experiment with geographically distributed collaborators



Create templates to fit the type of experiments to be described



Curate your experiment using a desktop-based, platform independent tool.



collect and curate, following standards

Describe the experimental steps using community-defined minimum reporting requirements and ontologies, where possible.

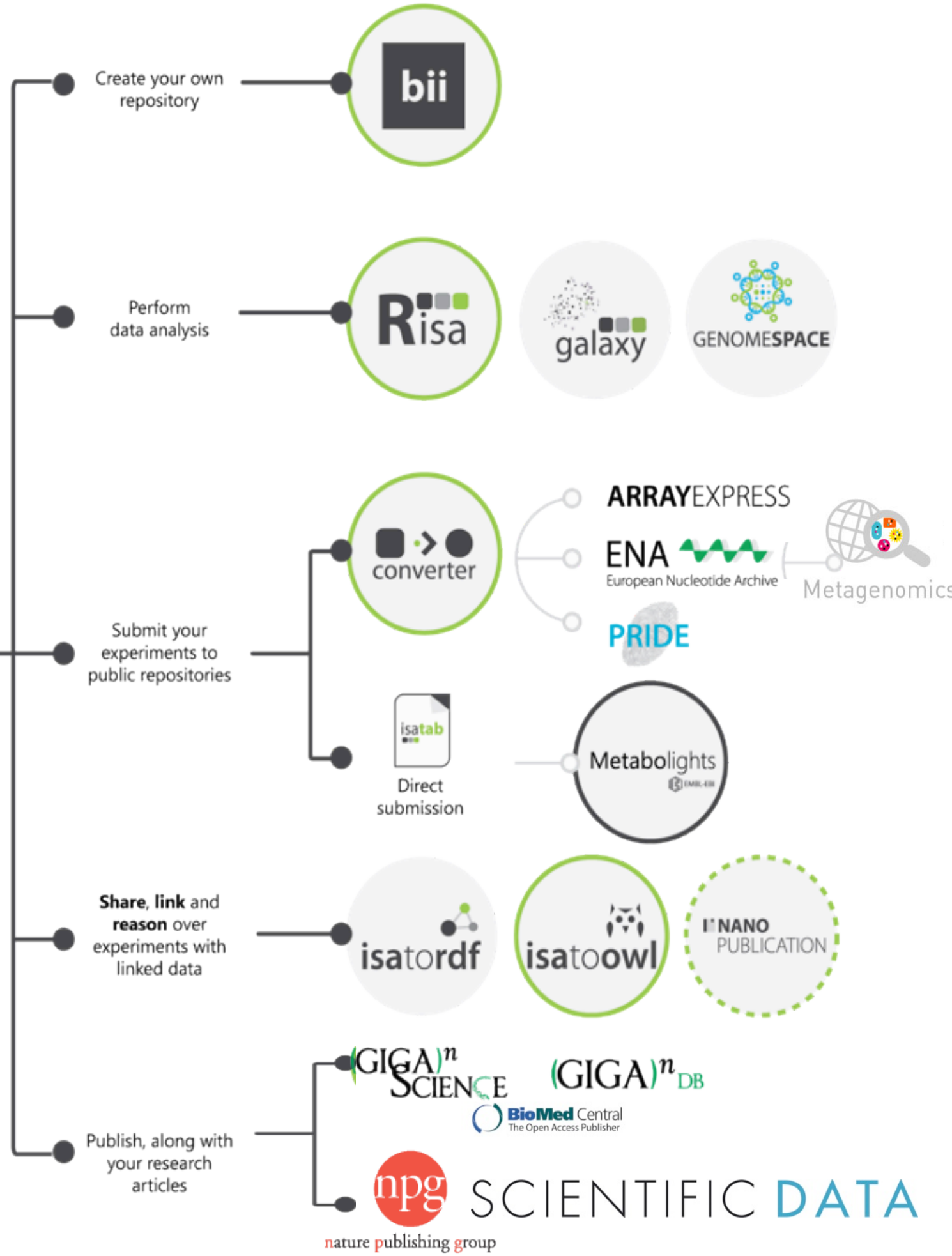
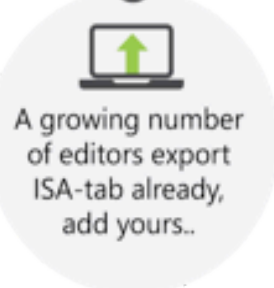
Describe & curate your experiment with geographically distributed collaborators



Create templates to fit the type of experiments to be described

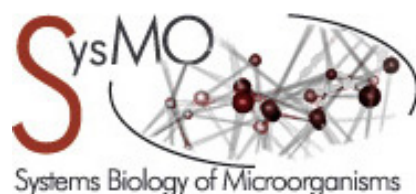


Curate your experiment using a desktop-based, platform independent tool.



A growing **ecosystem of over 30 public and internal resources** using the **ISA metadata tracking framework (ISA-Tab and/or format)** to facilitate **standards-compliant collection, curation, management and reuse** of investigations in an increasingly diverse set of life science domains, including:

- environmental health
- environmental genomics
- metabolomics
- metagenomics
- nanotechnology
- proteomics
- stem cell discovery
- system biology
- transcriptomics
- toxicogenomics
- also by communities working to build a library of cellular signatures



Toxicity data









THE LD₅₀ OF TOXICITY DATA IS
2 KILOGRAMS PER KILOGRAM.

<http://xkcd.com/1260/>

InnoMed PredTox Project

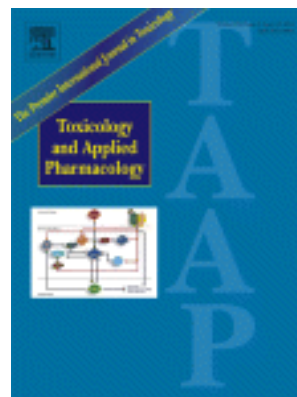
Goal: earlier pre-clinical safety evaluation by combining results from 'omics technologies and conventional toxicology methods



	 liver	 kidney	 blood serum	 plasma	 urine
 protein expression profiling by mass spectrometry	✓	✓	✓		✓
 transcription profiling by dna microarray	✓	✓	✓	✓	
 metabolite profiling by mass spectrometry	✓	✓	✓		✓
 metabolite profiling by nmr spectroscopy	✓	✓	✓		✓
 histology	✓	✓	✓	✓	✓
 clinical chemistry			✓	✓	✓
 hematology			✓	✓	

Suter et al 2011. EU Framework 6 Project: Predictive Toxicology (PredTox)—overview and outcome.

Boitier et al 2011. A comparative integrated transcript analysis and functional characterization of differential mechanisms for induction of liver hypertrophy in the rat



2-week systemic rat study using male Wistar rats (N=15 per dose group)



blood collection before sacrifice ●
sample collection after sacrifice ●

- ▼ BII-S-8
 - ▼ s_FP001RO-all-samples.txt
 - 🧪 a_FP001RO-hematology
 - 🧪 a_FP001RO-kidney-histop
 - 🧪 a_FP001RO-liver-histopa
 - 📊 a_FP001RO-SELDI-MS-p
 - 📊 a_FP001RO-DNACHIP-tr
 - 📊 a_FP001RO-2DPAGE-MA
 - 🧪 a_FP001RO-serum-bioch
 - 📊 a_FP001RO-NMR-metab
 - 🧪 a_FP001RO-urine-bioche
 - 📊 a_FP001RO-LCMS-metal

study description

Study Identifier

BII-S-8

Study Title

Study FP001RO: Evaluation of the Acute Toxicity, Gene Expression, Protein Express

Study Description

Repeated dose 14-day toxicity study in adult male rats (*Rattus norvegicus*), Wistar strain, using chemical compound FP001RO, administered daily orally, sponsored by Hoffmann-La Roche AG, funded by European Union Framework Program 6 - Innovative Medicines

Study Grant Number [c]

Study Funding Agency [c]







Study Submission Date

28/04/2006

Study Public Release Date

STUDY ASSAYS

+ add new assay(s)

 VIEW X OBI:Hematology a_FP001RO-hematology.txt	 VIEW X OBI:Histology a_FP001RO-kidney-histopat...	 VIEW X OBI:Histology a_FP001RO-liver-histopath...	 VIEW X OBI:protein expression p... OBI:mass spectrometry SELDI-TOF MS a_FP001RO-SELDI-MS-prot...	 VIEW X OBI:transcription profiling OBI:DNA microarray Affymetrix GeneChip a_FP001RO-DNACHIP-trans...	 VIEW VII OBI:protein expressi OBI:mass spectromet 2D-PAGE / MALDI-T a_FP001RO-2DPAGE-M
--	---	---	--	--	--

STUDY DESIGN DESCRIPTORS

+ add a new design column

Field Name

Study Design Type

STUDY PUBLICATIONS

+ add a new publication column 🔍 search for publication

Field Name	publication	publication
Study PubMed ID	21315101	20162557
Study Publication DOI	10.1016/j.taap...	10.1002/pmic...
Study Publication Author List	Boitier E, Amb...	Collins BC, Sp...
Study Publication Title	A comparative...	Use of SELDI ...
Study Publication Status	published	published

information

study

a study contains information about: samples; treatments applied; and associated assays.



+ add a new row column

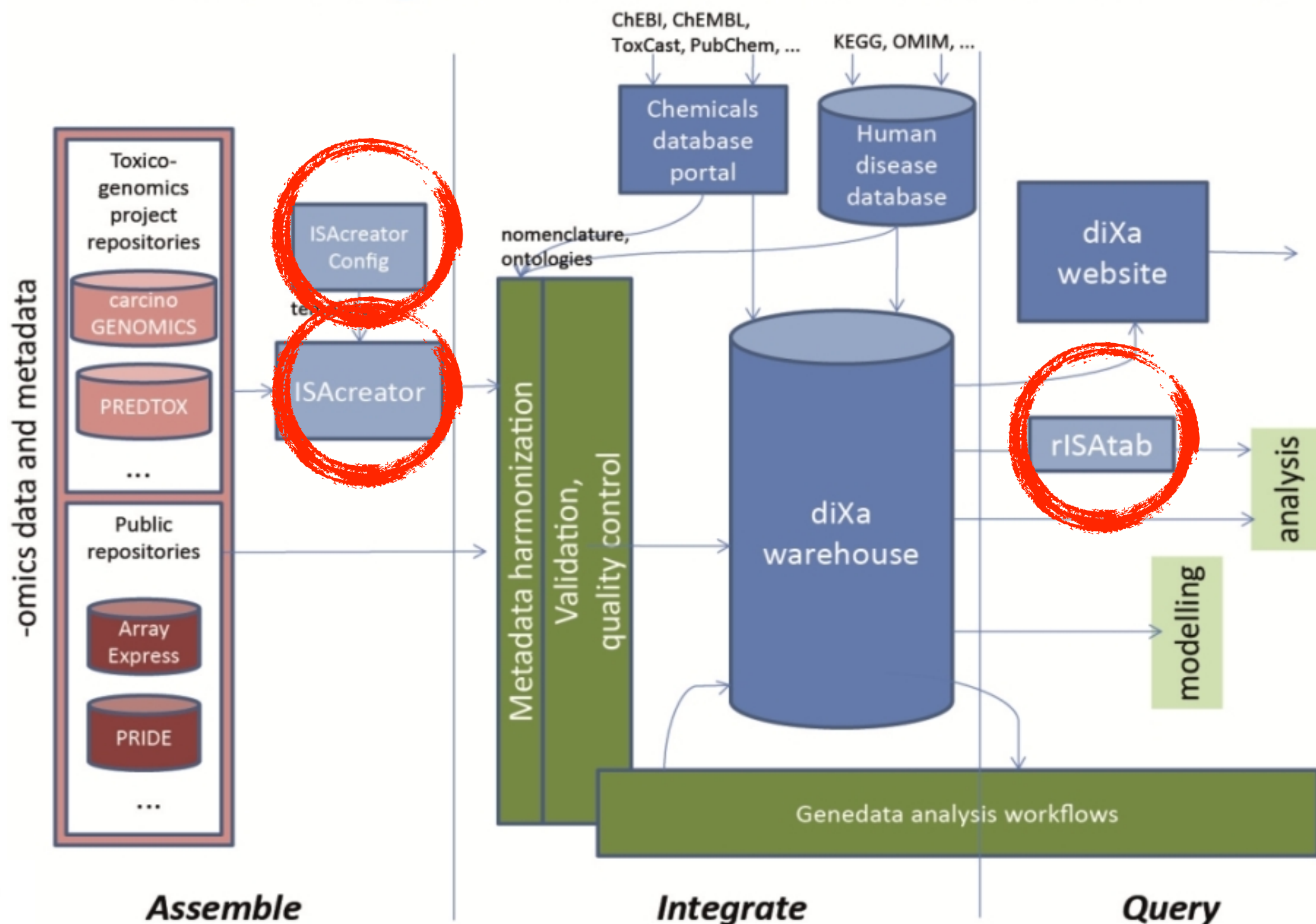
Field Name	● row	● row	● row	● row	● row	● row	● row	● row
Source Name	FP001RO_A31	FP001RO_A32	FP001RO_A33	FP001RO_A34	FP001RO_A35	FP001RO_A36	FP001RO_A37	FP001RO_A38
organism [CH]	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...	NEWT:Rattus n...
strain [CH]	Wistar	Wistar	Wistar	Wistar	Wistar	Wistar	Wistar	Wistar
provider [CH]	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...	RCC Ltd, CH-4...
age [CH]	8 to 10	8 to 10	8 to 10	8 to 10	8 to 10	8 to 10	8 to 10	8 to 10
Unit	UO:week	UO:week	UO:week	UO:week	UO:week	UO:week	UO:week	UO:week
development stage [CH]	adult	adult	adult	adult	adult	adult	adult	adult
body weight at begining of study [CH]		250.6	249.78	256.44	244.12	248.89	251.59	249.44
Unit	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram
body weight at sacrifice [CH]	331.25	322.32	331.23	341.69	319.1	330.33	321.28	327.3
Unit	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram
liver weight at sacrifice [CH]	8.97	8.2	9.04	9.02	8.51	10.56	8.63	8.99
Unit	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram
kidney weight at sacrifice [CH]	2.27	1.92	2.13	2.07	2.11	2.28	2.18	2.09
Unit	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram	UO:gram
Protocol REF	animal care an...	animal care an...	animal care an...	animal care an...	animal care an...	animal care an...	animal care an...	animal care an...
housing density - cage type - beddin...	1 animal per c...	1 animal per c...	1 animal per c...	1 animal per c...	1 animal per c...	1 animal per c...	1 animal per c...	1 animal per c...
diet name - diet form [PV]	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...	Provimi Kliba, ...
compound name - formula [PV]	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	FP001RO (C23...	FP001RO (C23...	FP001RO (C23...
quantitative dose [PV]	no dose	no dose	no dose	no dose	no dose	20	20	20
Unit						mg/kg/day	mg/kg/day	mg/kg/day
vehicle [PV]	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin	nacl/gelatin
substance administration [PV]	oral gavage	oral gavage	oral gavage	oral gavage	oral gavage	oral gavage	oral gavage	oral gavage
substance formulation [PV]	microsuspensi...	microsuspensi...	microsuspensi...	microsuspensi...	microsuspensi...	microsuspensi...	microsuspensi...	microsuspensi...
Performer	lsuter	lsuter	lsuter	lsuter	lsuter	lsuter	lsuter	lsuter
Date	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...	2006-10-19T1...
Protocol REF	necropsy	necropsy	necropsy	necropsy	necropsy	necropsy	necropsy	necropsy
sacrifice event [PV]	planned sacrifice	planned sacrifice	planned sacrifice	planned sacrifice	planned sacrifice	planned sacrifice	planned sacrifice	planned sacrifice
Sample Name	FP001RO_A31...	FP001RO_A32...	FP001RO_A33...	FP001RO_A34...	FP001RO_A35...	FP001RO_A36...	FP001RO_A37...	FP001RO_A38...
organism part [CH]	FMA:urine	FMA:urine	FMA:urine	FMA:urine	FMA:urine	FMA:urine	FMA:urine	FMA:urine
compound [FV]	vehicle	vehicle	vehicle	vehicle	vehicle	FP001RO	FP001RO	FP001RO
dose [FV]	no dose	no dose	no dose	no dose	no dose	low dose	low dose	low dose
duration of exposure [FV]	15	15	15	15	15	15	15	15
Unit	UO:day	UO:day	UO:day	UO:day	UO:day	UO:day	UO:day	UO:day

BII-S-8

Title:	Study FP001RO: Evaluation of the Acute Toxicity, Gene Expression, Protein Expression, Metabolite Production, Clinical Chemistry and Pathology Profile Following an Oral Administration of Compound R2717 to Rats
Organism(s):	Rattus norvegicus
Description:	Repeated dose 14-day toxicity study in adult male rats (Rattus norvegicus), Wistar strain, using chemical compound FP001RO, administered daily orally, sponsored by Hoffmann-La Roche AG, funded by European Union Framework Program 6 - Innovative Medicines Initiatives - Integrated Project Predictive Toxicology (EU FP6 InnoMed PredTox: LSHB-CT-2005-518170), The study will examine the gene, protein and metabolite profiles, along with traditional toxicological endpoint information in the liver, kidney, blood and urine of rats 1-14 days after exposure. By design, low doses lie between No Observed Adverse Effect Level (NOAEL) and Lowest Observed Adverse Effect Level (LOAEL).
Design(s):	factorial design
Experimental factor(s):	<div style="background-color: #f0f0f0; padding: 5px;"> <p>dose 3 recorded ▼</p> <p>duration of exposure 3 recorded ▼</p> <p>compound 2 recorded ▼</p> </div>
Publication(s):	<p>Collins BC, Sposny A, McCarthy D, Brandenburg A, Woodbury R, Pennington SR, Gautier JC, Hewitt P, Gallagher WM Use of SELDI MS to discover and identify potential biomarkers of toxicity in InnoMed PredTox: a multi-site, multi-compound study. CiteXplore:20162557</p> <p>Boitier E, Amberg A, Barbie V, Blichenberg A, Brandenburg A, Gmuender H, Gruhler A, McCarthy D, Meyer K, Riefke B, Raschke M, Schoonen W, Sieber M, Suter L, Thomas CE, Sajot N. A comparative integrated transcript analysis and functional characterization of differential mechanisms for induction of liver hypertrophy in the rat. CiteXplore:21315101</p>
Sample attribute(s):	<div style="background-color: #f0f0f0; padding: 5px;"> <p>RNA concentration 1 recorded ▼</p> <p>organism part 7 recorded ▼</p> <p>liver weight at sacrifice 41 recorded ▼</p> <p>kidney weight at sacrifice 30 recorded ▼</p> <p>organism 1 recorded ▼</p> </div>

Data Infrastructure for Chemical Safety

Full walk-through from data sources to data consumers





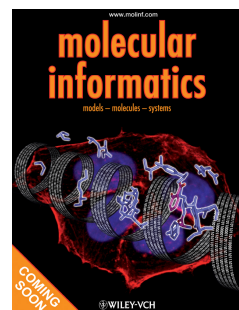
Towards the Replacement of *in vivo* Repeated Dose Systemic Toxicity Testing



Supporting integrated data analysis & servicing of alternative testing methods in toxicology

Safety Evaluation Ultimately Replacing Animal Testing- I (SEURAT-1): looking at improving safety assessment without the need for animal experiments

ToxBank: cross-cluster infrastructure project



Kohonen et al. 2013 The ToxBank Data Warehouse: a research cluster of 7 EU FP7 Health systems toxicology and toxicogenomics projects.

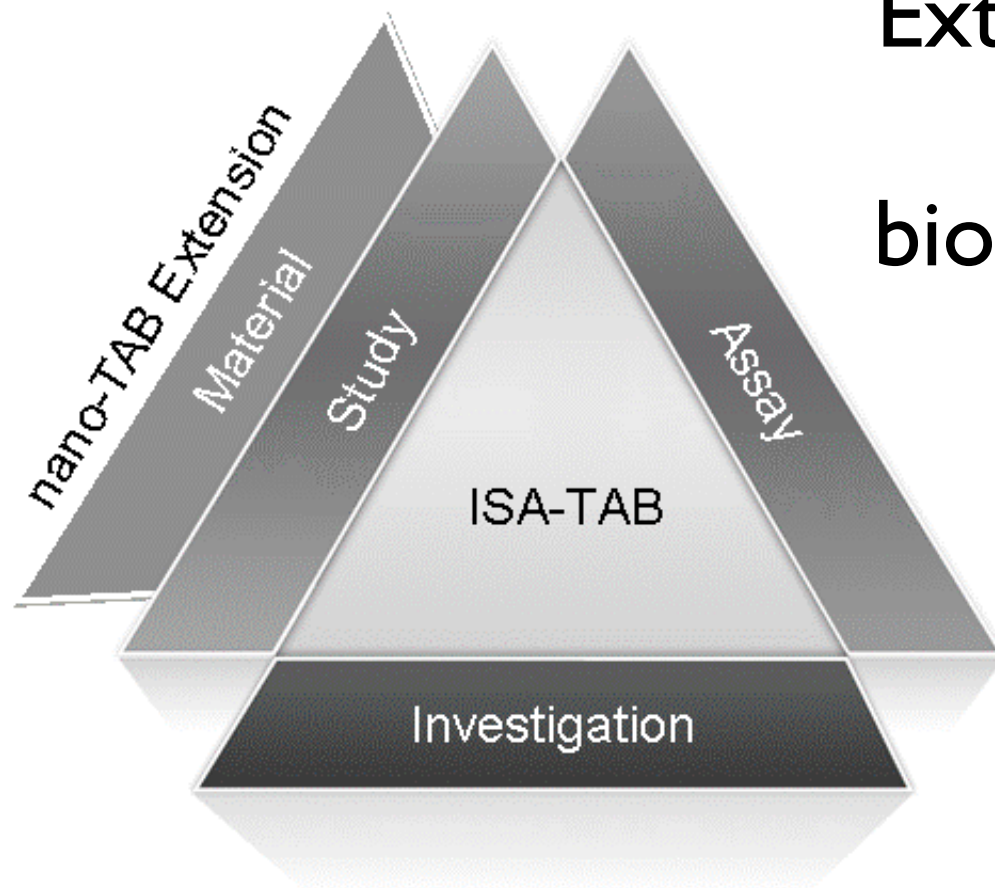
<http://toxbank.net>



ISA-TAB-Nano



Extension of ISA-TAB format to represent nano-materials, small molecules and biological specimens along with their assay characterisation data



Work Item: ASTM WK28974 - New Specification for A Standard File Format For the Submission and Exchange of Data on Nanomaterials and Characterizations



Baker et al. 2013 **Standardizing data**

Thomas et al. 2013 **ISA-TAB-Nano: A specification for sharing nanomaterial research data in spreadsheet-based format**

isatoowl

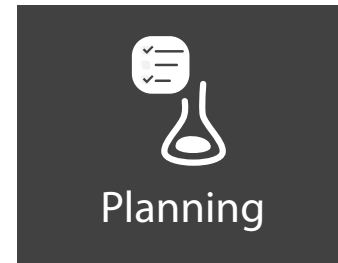


nature publishing group

SCIENTIFIC DATA

(GIGA)ⁿ SCIENCE

(GIGA)ⁿ DB



isatordf

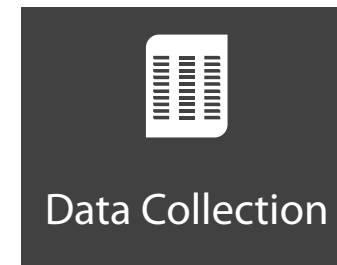
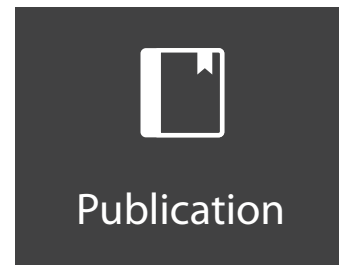
isa-tab

MetaboLights

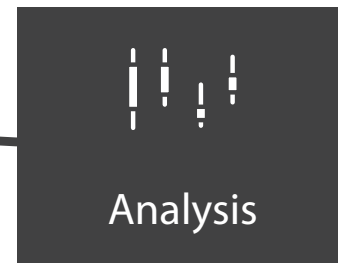
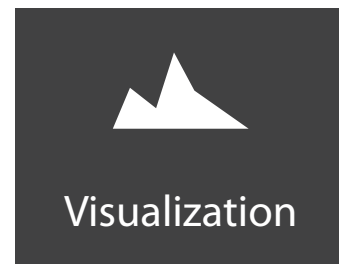


converter

mage-tab | pride-ml | sra-xml | others in progress
ArrayExpress | PRIDE | ENA



OntoMaton



30 GENOMESPACE



Susanna Sansone

Team Leader

University of Oxford

LinkedIn

Twitter



Philippe Rocca-Serra

Technical Co-ordinator

University of Oxford

GitHub

LinkedIn

Twitter



Eamonn Maguire 31

Lead Software Engineer

University of Oxford

Figshare

GitHub

LinkedIn

Twitter



Alejandra Gonzalez-Beltran

Senior Software Engineer

University of Oxford

GitHub

LinkedIn

Twitter

funding



2012-2015, latest funding stream for metabolomics standards from the EU COSMOS.

2012-2015, travel funds for a UK-China partnership on harmonization of data curation from the BBSRC, a collaboration with TGAC and BGI/GigaScience

2012-2015, more funding for ISA project from the BBSRC MGportal project, a collaboration with the EMBL-EBI.

2011-2014, the ISA project has received another new funding stream from BBSRC and NERC.

2010-2013, the ISA project has received new funding stream from BBSRC and NERC.

2007-2010, the ISA project has been mainly supported by funds from the EU IP CarcinoGenomics, with contribution from EU NoE NuGO, NERC-NEBC, BBSRC and EMBL-EBI.





Questions?

You can email us...
isatools@googlegroups.com

View our website
<http://www.isa-tools.org>

View our Git repo & contribute
<http://github.com/ISA-tools>

View our blog
<http://isatools.wordpress.com>

Follow us on Twitter
[@isatools](https://twitter.com/isatools)

