

Biocuration 2010 The Conference of the International Society for Biocuration
October 11-14, 2010 Odaiba, Tokyo, Japan Fourth International Biocuration Conference

MetaCuration Standards and Minimum Information about a Bioinformatics Investigation (MIABi)

TAN Tin Wee

Asia Pacific Bioinformatics Network
(APBioNet)



ENT OF
MISTRY
medical sciences



The Asia Pacific Bioinformatics Network



Pet Peeve 1



The connection has timed out

The server at www.somedatabase.org is taking too long to respond.

- The site could be temporarily unavailable or too busy. Try again in a few moments.
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

or

Try Again

“404 Not Found!”

- Disappearing Databases and Web sites Services
- Instant access to Information:
“When you see a citation, you should be able to read a pdf file of the paper in seconds”

Pet Peeve 2

“Now how did he get that conclusion?”

*“Nice graphic, but how the h*ll did he plot that?”*

- Insufficient information and data for reconstructing in silico experiments
- Full Disclosure in materials and methods, supplementary materials and all – Scientific Reproducibility

Pet Peeve 3

“Compilation Error”
“Missing Library File”
“Loads of gibberrish”

- Cannot find software or test it, installation errors, missing dependencies
- Bioinformatics software should be downloadable and installable and work instantly

Pet Peeve 4

陳定煒

Tan Tin Wee



*“Is this guy
TW Tan or TT Wee?
When is he Tin Wee Tan?
Even the Chinese can’t tell!”*

Geof Bilder story

Cross Ref Contributor ID

STM Innov London 2009)

http://river-valley.tv/media/conferences/stm-innovations-2009/0102-Geoffrey_Bilder/

- Name Order, Name Variants, Name Changes
- Name Collisions
- Problem with Citations and H-Index
- Multiple Scopus IDs; Hopefully one ResearcherID
- Any number of OpenID, ORCID ?

WIBNI

“Wouldn't it be Nice If”

- Reproduce everything the paper did instantly?
- State the author name and have all his papers correct citations etc all pop up?
- Review a paper and “play” back the authors’ workflow and poke and prod the system at any step
- Take the software and plug in my data and pop up my own nice result?

WIBFI projects

“Wouldn't It Be Fun If”

- 1984 Punchcards to do Chou Fasman secondary structure plots
- 1991/1992 Biocomputing, WAIS, Gopher and Web: Amos Bairoch story, Prosite on WAIS (Time before Gopher and the Web browser)
- 1994: Internet for Hearing Impaired and Visually Handicapped
- 1996: 3D Photocopying: Internet2 and Rapid Prototyping:
- 1998: Internationalized Domain Names (IDNs)
- 2001: S* Alliance: global bioinformatics teaching
- 2005: Integrated Workflow and Manufacturing Systems for BioSurveillance
- 2011: Synthesizing Left Handed Molecules

Disambiguation: Unique identifiers

- Author identifiers
equivalencing ResearcherIDs, ScopusIDs etc
- Unique identifiers in Scientific Terminology
- Accreditation; Accountability; Non-Repudiability

Nature Precedings : doi:10.1038/npre.2010.5078.1 : Posted 21 Oct 2010

The screenshot shows a web browser window with the URL <http://aid.apbionet.org/>. The page title is "Unified Author ID Portal". Below the title, there is a section for "Author Name/Information Search Facility". It contains instructions: "Search using an author name or other public publication information about the author and return all possible matches to select from. E.g. Enter 'Li X' or 'Tan TW' or 'John Jacobs' as author names, 'Curtin University' as an affiliation, etc. (enter at least one search keyword)". There are four input fields: "Name or Part of Name", "Search Keyword 1", "Search Keyword 2", and "Search Keyword 3". A "Search" button is located below these fields. Below this is a section for "Multiple ID Resolution Search Facility" with instructions: "Search by entering two author identifiers to determine if they are the same person. E.g. Enter Scopus ID and ResearcherID and determine if the individuals with the same name are the same person." There are two dropdown menus, both currently set to "Researcher ID", and a "Search" button.

APBioNet's Prototype
For Disambiguating
Author IDs

In <http://aid.apbionet.org>
- Self-Assertion Mode or
Hybrid Assertion Mode
- Distributed system based
On Internet Domain Name
System

Reachability: Digital Object Deposition

<http://docid.apbionet.org>

ly1Bio 2010 Con... BioMed Central Blog : ... Panton Principles Open Knowledge Defi... Science Commons » P... MIABD-V3.doc - Goog..
FutureSchedule ... MIABD-V3.doc - ... The Conference... nbt.1411.pdf (a... isatab - Google ... malaysia suhaimi... DocID

The Asia Pacific Bioinformatics Network

Endorsed By APAN TEL AIOING iSC

DOCUMENT DROP

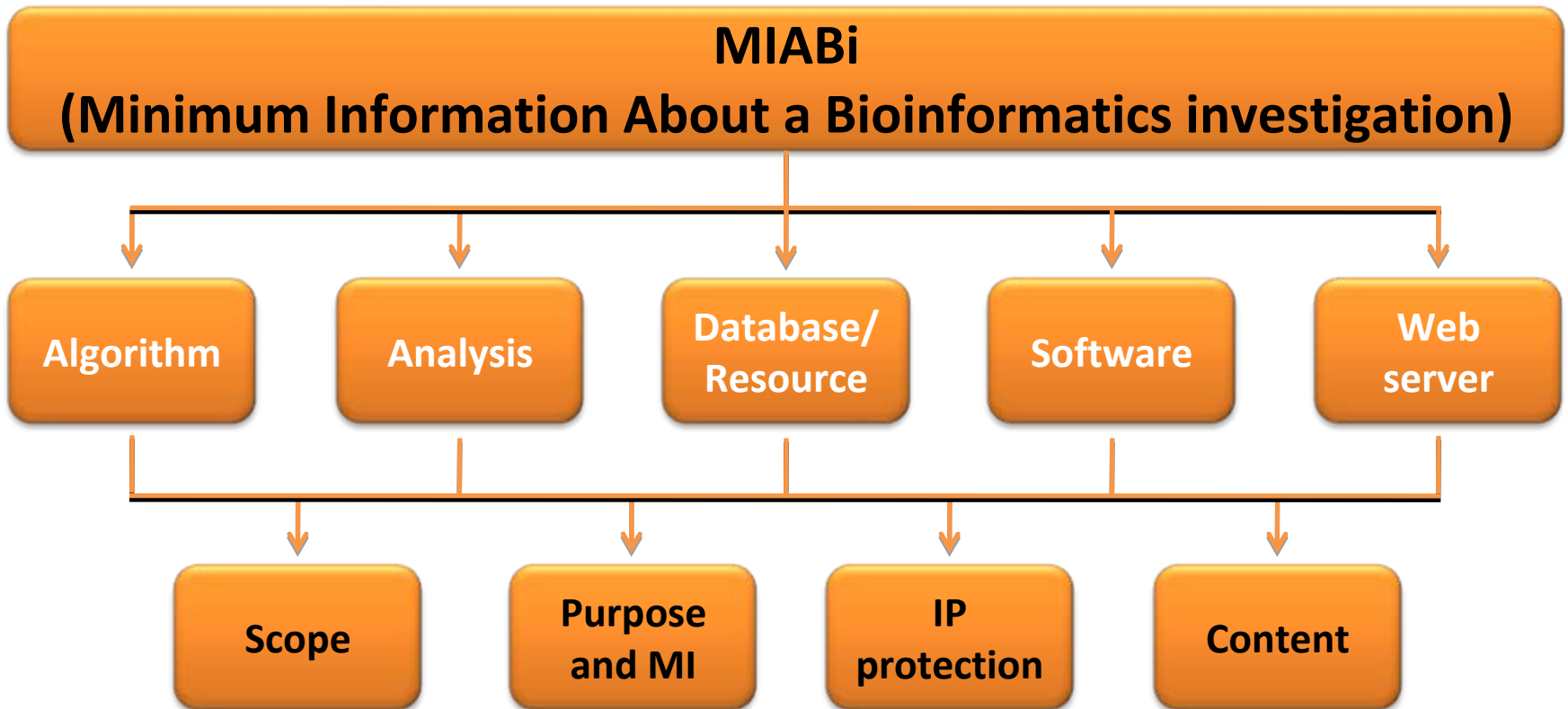
An open-access document repository initiative of APBioNet

[Datasets/Supplementary Materials](#)

Nature Precedings : doi:10.1038/npre.2010.5078.1 : Posted 21 Oct 2010

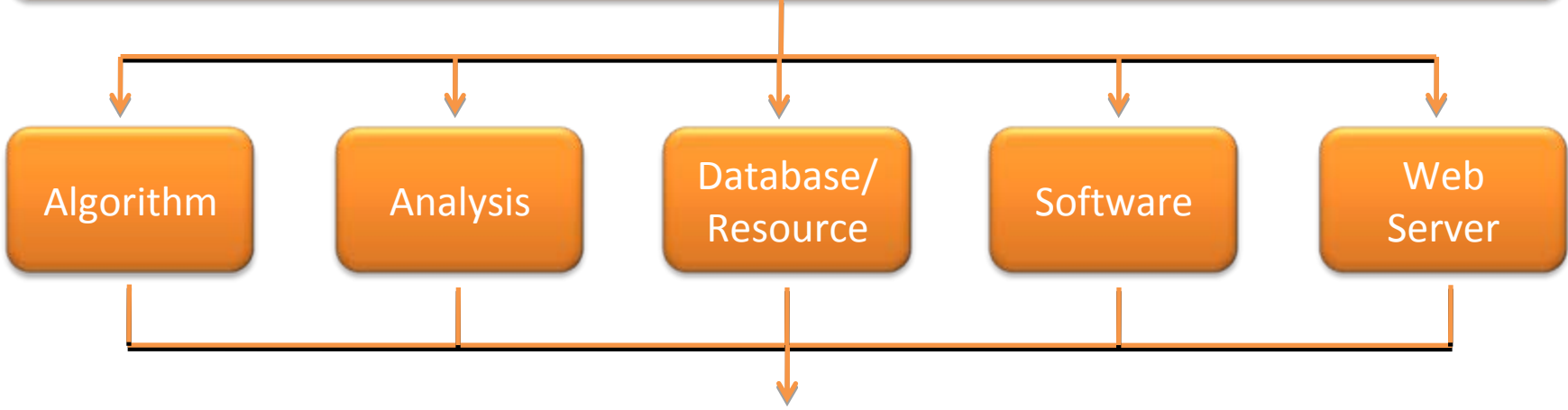
- Open Source
- Open Access
- Self-Deposition
- Escrow
- DocID
- Granularity

Sufficiency of Disclosure: MIABi



Supporting Reproducibility

MIABi



Scope of the MI for the subject bioinformatics investigation

Purpose and the MI for the subject bioinformatics investigation, including reproducibility, reusability, and validation of claims as published in scientific publications

Protection of intellectual property for the subject bioinformatics investigation conforming to MIABi standards

Content of the subject MI, including general descriptions, input data and parameters, output data, workflow of process, regression testing, performance and scalability, and provenance, where applicable

Reinstantiability

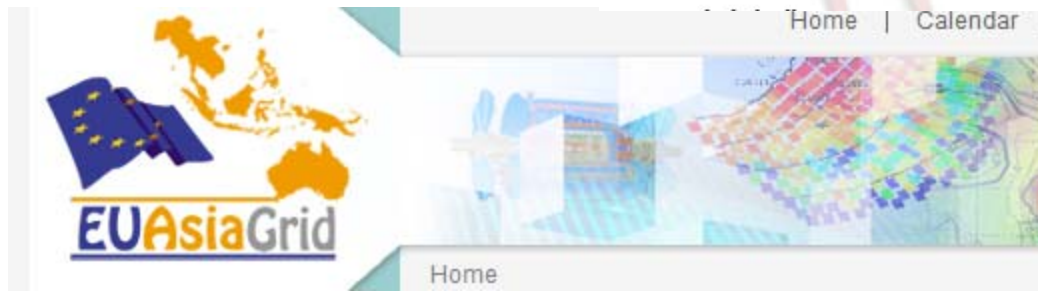
- BioSlax "LiveOS" images
- Lzm compressed files of deltas
- Citrix Xen Hypervisors and full hardware virtualisation
- Grid Computing on ***EUAsiaGrid.org***
- Cloud Computing on Amazon AWS



About

BioSLAX is a new live CD/DVD suite of bioinformatics tools that has been released by the resource team of the Bioinformatics Center (BIC), National University of Singapore (NUS). Bootable from any PC, this CD/DVD runs the compressed SLACKWARE flavour of the LINUX operating system also known as SLAX.

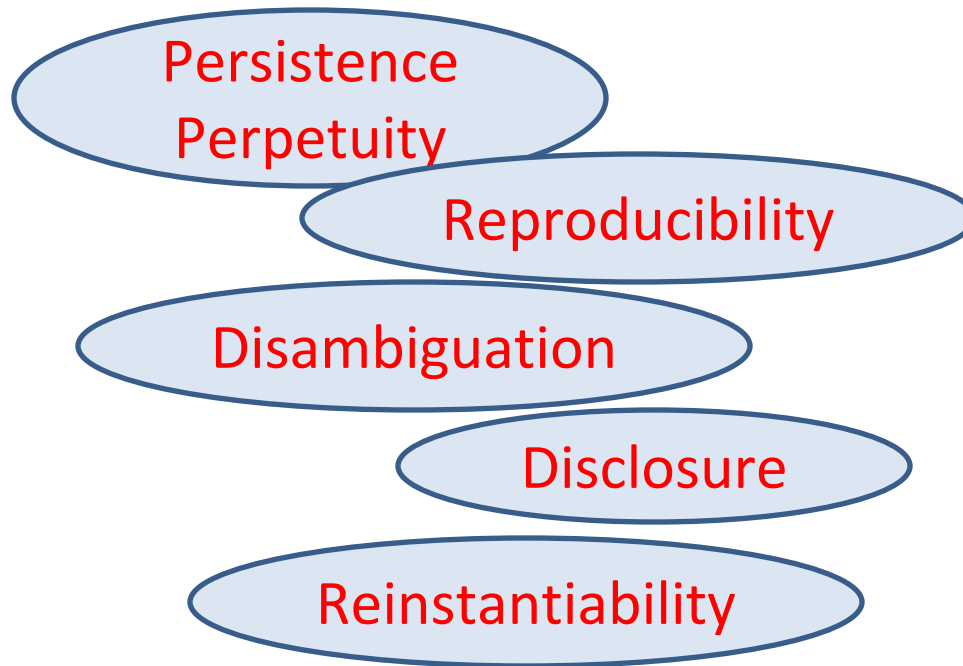
SLAX is becoming the live CD/DVD of choice because of its ability to modularize almost any application and plug it into the system on the fly. The system can also be installed to USB thumbdrives or directly to the PC as a regular Linux using the BioSLAX installer provided.

A screenshot of the BioSLAX desktop environment, showing a window with the OpenOffice 2.1 logo and various application icons.

People-Papers Platform: InCoB

- International Conference on Bioinformatics since 2002
- 2006: Publication of top papers in BMC Bio
- 2010: BMC Bioinformatics; BMC Genomics; Immunome Research; Bioinformation; IPSJ Transactions

Nature Precedings : doi:10.1038/npre.2010.5078.1 : Posted 21 Oct 2010



WIBMFI

Wouldn't it be MORE fun if:

- BioDB100 exemplar of how this will all work together
- MetaCurate all kinds of databases:
Maintain Meta-Information About Databases which curate biological knowledge
- Basis for Interoperability
- Building the BioInformation Infrastructure for future generations
- Cleaning up our act
TCP/IP and IETF: Internet Engineering Task Force

The BioDB 100 Opportunity

- InCoB Paper submission process
- Editors and Reviewers
- MIABi-Compliant Disclosures
- Authors-disambiguated; maybe Terminologies?
Ontologies?
- Deposited Databases, Datasets, Webservices
- Reinstantiability with virtual machines
- Reproducibility on grids and clouds
- Regression testing with bioinformatics software
- Microcitation and Nanocitation and Metacuration
- To the future and beyond!

Contact

Tan Tin Wee

Bioinformatics Centre Lab/

Dept of Biochemistry

National University of Singapore

Email: tinwee@bic.nus.edu.sg