SGBC
SLU Global Bioinformatics Centre

SLU
Swedish University of Agricultural Sciences

eB3Kit + Galaxy

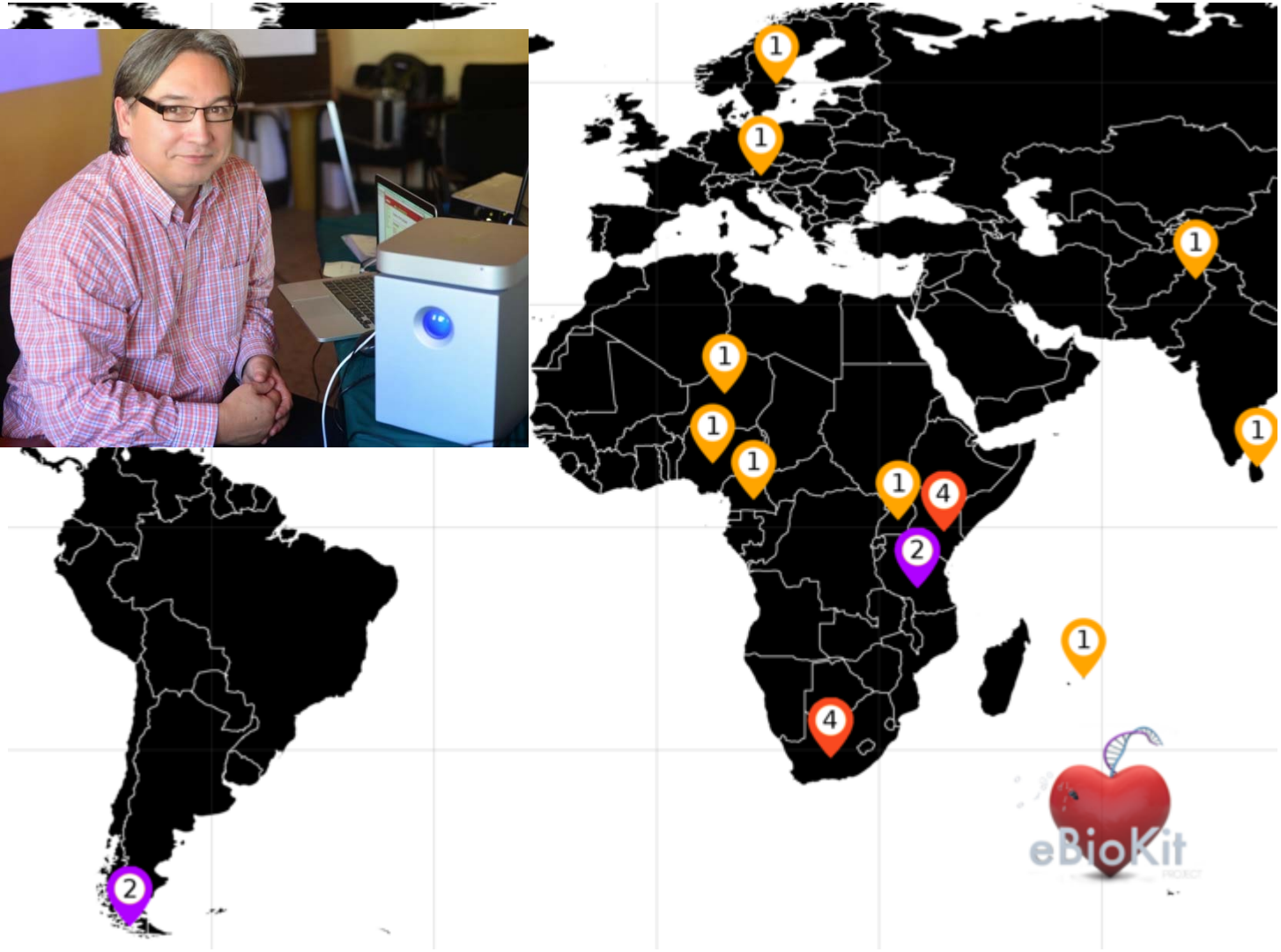
B3AFRICA - WORKPACKAGE 4

BRIDGING BIOBANKING AND BIOMEDICAL RESEARCH
ACROSS EUROPE AND AFRICA

Rafael Hernández de Diego & Tomas Klingström

2016-09-19

SGBC



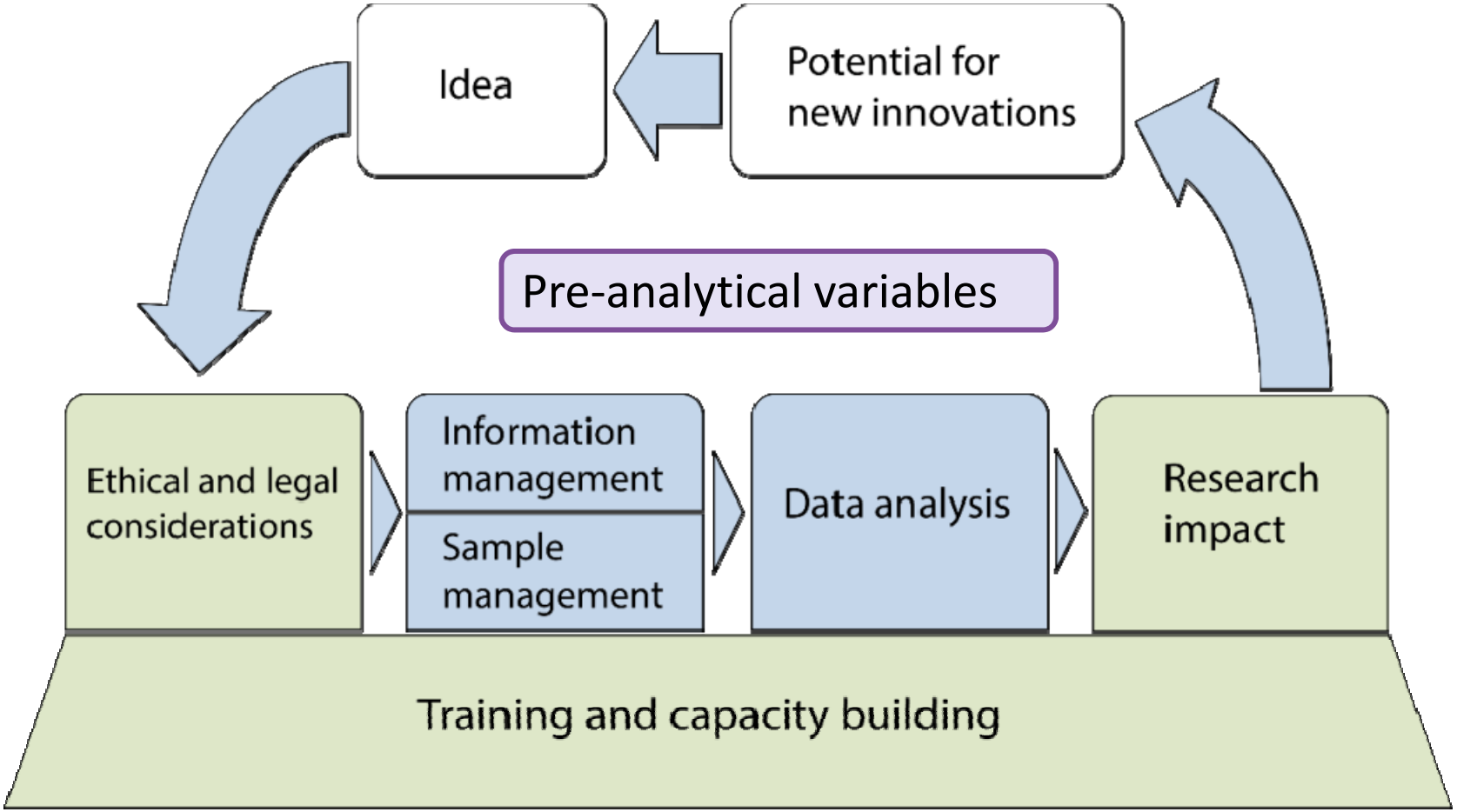
OVERALL AIM B3AFRICA

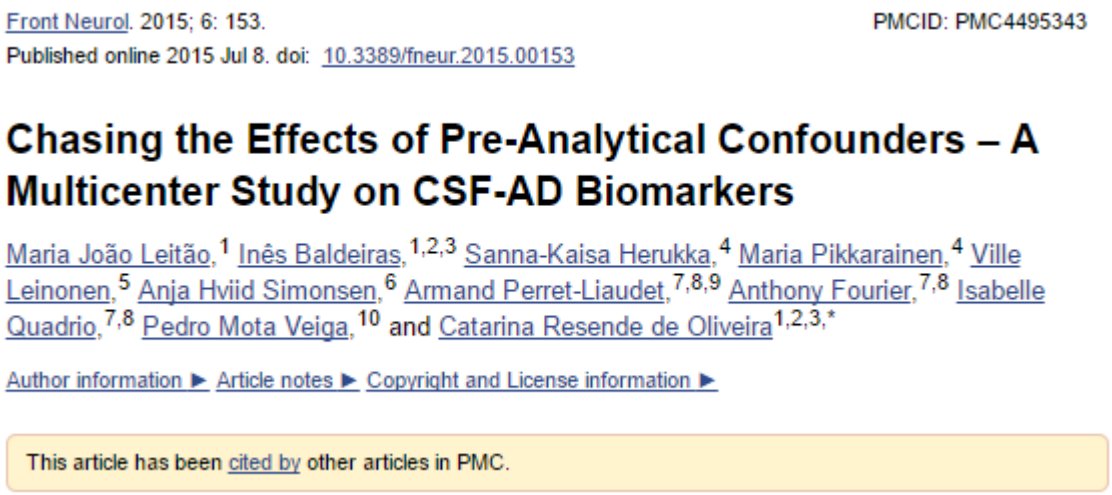
1. Create an effective informatics platform for research integrated biobanks
2. Support the ethical and legal framework for collaboration across borders

INFORMATION TRANSFER

External validation

Collaboration





Front Neurol. 2015; 6: 153. PMCID: PMC4495343
Published online 2015 Jul 8. doi: 10.3389/fneur.2015.00153

Chasing the Effects of Pre-Analytical Confounders – A Multicenter Study on CSF-AD Biomarkers

Maria João Leitão,[1] Inês Baldeiras,[1,2,3] Sanna-Kaisa Herukka,[4] Maria Pikkarainen,[4] Ville Leinonen,[5] Anja Hviid Simonsen,[6] Armand Perret-Liaudet,[7,8,9] Anthony Fourier,[7,8] Isabelle Quadrio,[7,8] Pedro Mota Veiga,[10] and Catarina Resende de Oliveira[1,2,3,*]

Author information ► Article notes ► Copyright and License information ►

This article has been cited by other articles in PMC.

Three biomarkers

- Aβ42 increase IF high CSF total protein content AND (IF not centrifuged OR centrifuged at room tempearature).
- Aβ42 decrease if CSF is freeze-thawed more than 3 tims.
- Tau – not significantly altered under testing conditions.
- Phosphorylated Tau (pTau) – not significantly altered under testing conditions.

The Russian connection: See all our coverage | Read about our investigation

You're Drinking the Wrong Kind of Milk

And we're not talking about soy or almond, either.

JOSH HARKINSON MAR. 12, 2014 6:00 AM

TOP NEWS

A Vet Turns to Farming His Wounds

TEXT AND PHOTO

Critique of the hypothesis

- Top

A1 milk and diabetes mellitus type 1 (DM-1)

Reasons why the between-country correlations are far from conclusive are as follows:

- Were the individuals who developed diabetes the ones who took the A1 β-casein? What human evidence we have suggests that any influence of milk (not just A1) on DM-I operates in infancy (Borch-Johnsen *et al*, 1984; Mayer *et al*, 1988) and this is usually consumed in the form of infant formulas. Only a small percentage of total milk consumption in any particular country is in infant formulas. National average consumption estimates for milk and products cannot serve as quantitative assessment of the A1 β-casein consumption from infant formulas. These formulas usually contain increased whey and reduced casein and the milk protein used for their manufacture does not always come from within the country where the formulas are used.
- Confounding cannot be excluded. To take an example, people in Finland, with a high rate of DM-I have a high frequency of HLA haplotypes that indicate susceptibility to diabetes (Reijonen *et al*, 1991). The island of Sardinia has the highest DM-I incidence in the Mediterranean region; emigration studies and HLA types show that this is, in the main, genetically determined (Muntoni & Muntoni, 1999).
- Breast feeding can in some communities be negatively associated with DM-I in case–control studies (Jones *et al*, 1998). There are socioeconomic, and hence environmental differences between breast-fed and formula-fed infants and breast fed are not drinking A1 milk.
- Nutritional scientists have experienced the unreliability of correlation studies of food intake and chronic disease. It was earlier claimed that sugar consumption correlated with CHD (Yudkin, 1964); that countries' fat consumption correlated with breast cancer (World Cancer Research Fund, 1997) and that their meat consumption correlated with colon cancer (Armstrong & Doll, 1975). Closer human research has shown these associations to be spurious or uncertain (FAO/WHO Expert Consultation, 1998; Committee of Medical Aspects of Food & Nutrition Policy, 1998; Truswell, 2002).
- There are particular difficulties with the A1 β-casein-DM-I correlations. There is more uncertainty with national figures for A1 β-casein than for total milk casein and more uncertainties for these than for average milk consumption. Some important developed dairy countries were not included, for example, Netherlands, Ireland, nor were any of the emerging and developing countries.

WHY I AM EXCITED ABOUT THE PROJECT

(Freedman, L. et al. (2015) Plos Biology 13, e1002165.)

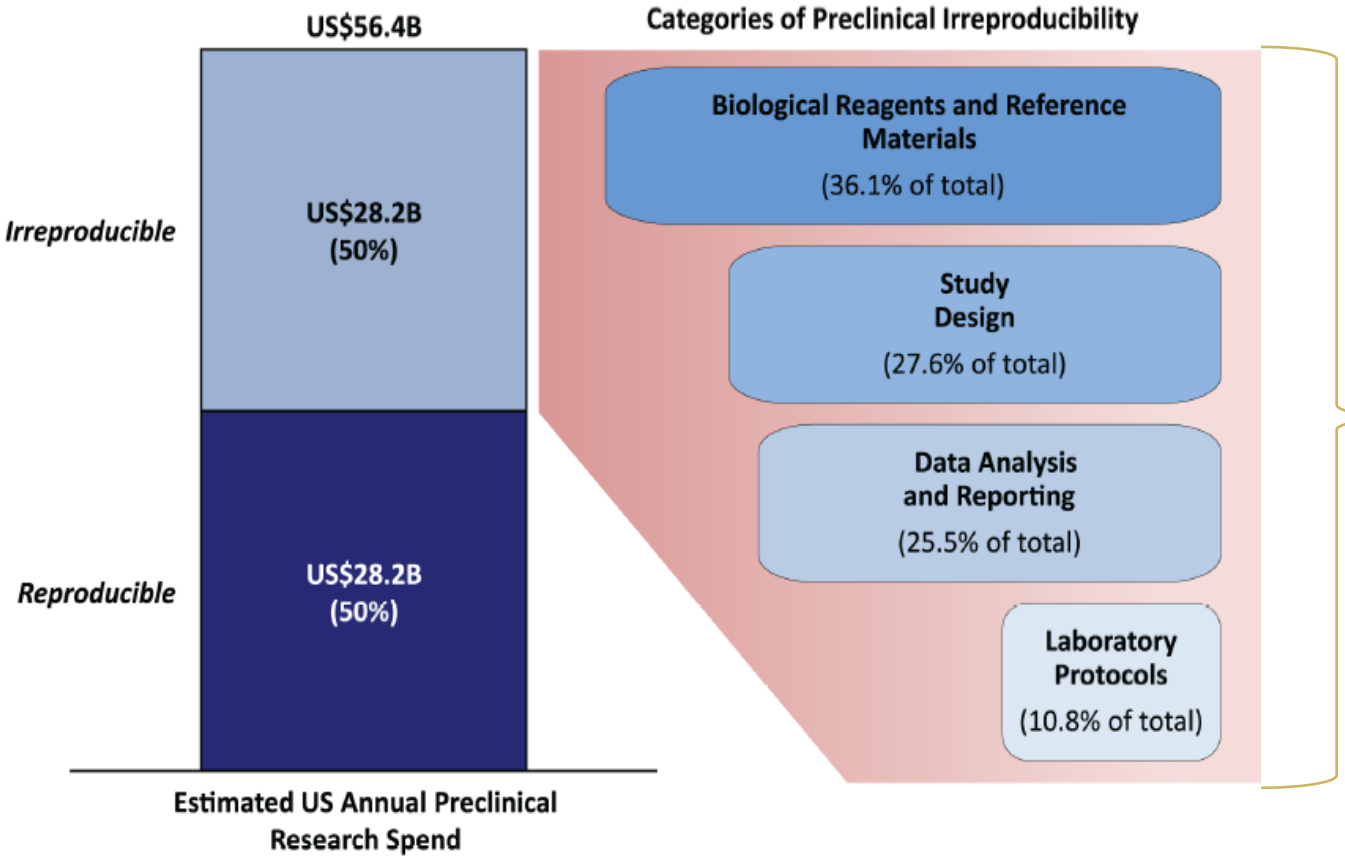
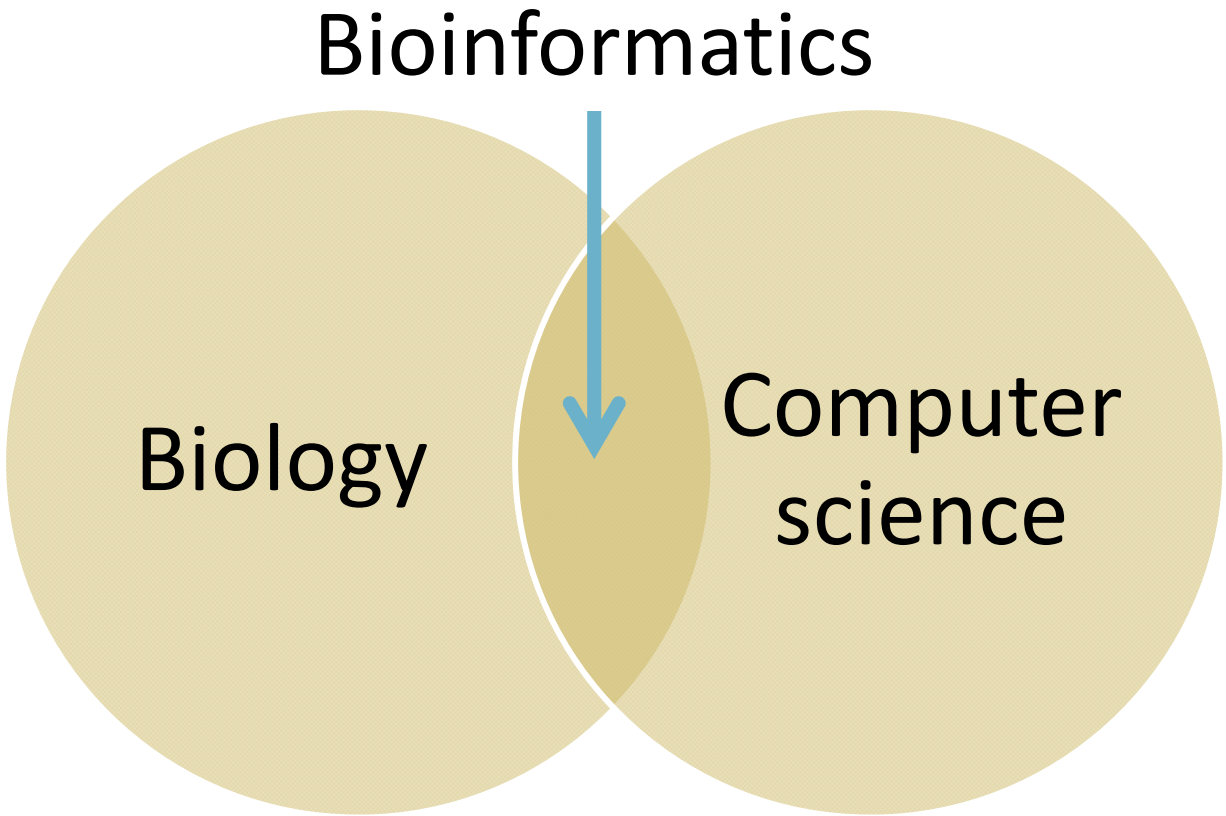


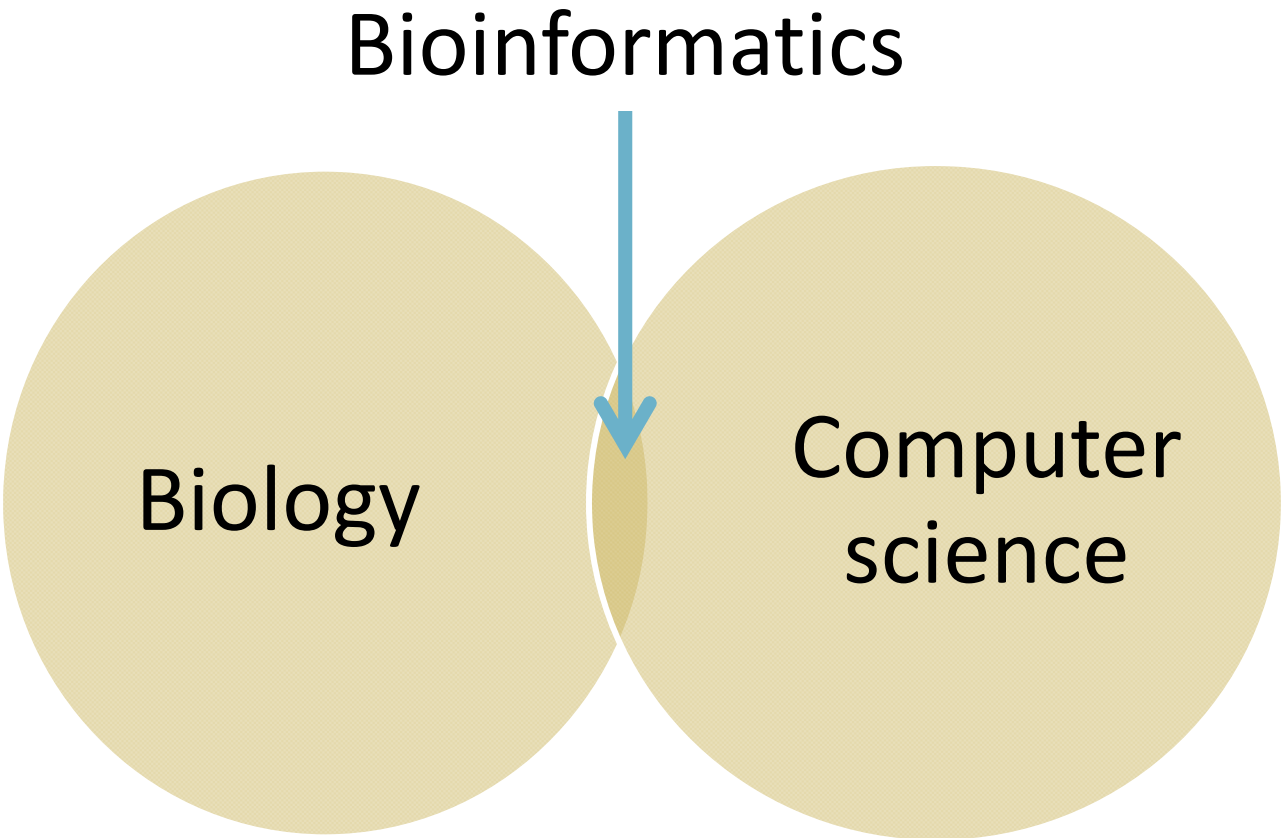
Fig 2. Estimated US preclinical research spend and categories of errors that contribute to irreproducibility. Note that the percentage value of error for each category is the midpoint of the high and low prevalence estimates for that category divided (weighted) by the sum of all midpoint error rates (see S1 Dataset). Source: Chakma et al. [18] and the American Association for the Advancement of Science (AAAS) [19].

eB3Kit allow us to monitor this

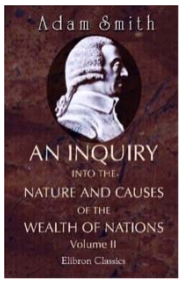
DESIGN PARADIGM



PURPOSE



Division of labour



- Make bioinformatics work more efficient by dividing labour tasks

Biologist/Medical specialist	→	Run standardized workflows
Bioinformatician (severe shortage)	→	Develop & standardize workflows
Systems admin	→	Maintenance (no biology skills)

WORKFLOWS

Galaxy Slim v0.1
ebiokit@gmail.com
Home
Workflows
Histories
Running 20140811-first-use
1
Workflow overview
2
History selection
3
Input selection
4
Custom parameters
5
Run workflow
6
Results
Details
Author: ebiokit
Description: No description provided
Workflow overview
1. Input dataset
2. Input dataset
3. Join
4. Group
5. Sort
6. Select first
7. Compare two Datasets
Cancel and go back
Next step

HARDWARE MANAGEMENT

Base cost: 2 999 USD
3.7 GHz quad-core 10 MB L3 cache
12 GB RAM
256 GB Flash memory
Firepro D300 GPUs with 2 GB GDDR5 VRAM each

Upgrades
Processor: 6-core 500 $, 8-core 2000 $ 12 core 3 500 $.
Ram: 16 GB 100 $, 32 GB 500 $, 64 GB 1 300 $.
Storage: 512 GB 300 $, 1 TB 800 $.
Graphics: D500 GPU 3 GB 400 $, D700 with 6 GB 1000 $.

CONNECTING TO HPCS

For ”simple” projects the major bottleneck for the eB3Kit is RAM.

BlastX: 13-14 GB

Blast2GO: 30-40 GB

Post Assembly correction: 40 GB

Trinity (prokaryotes): ~ 60 GB

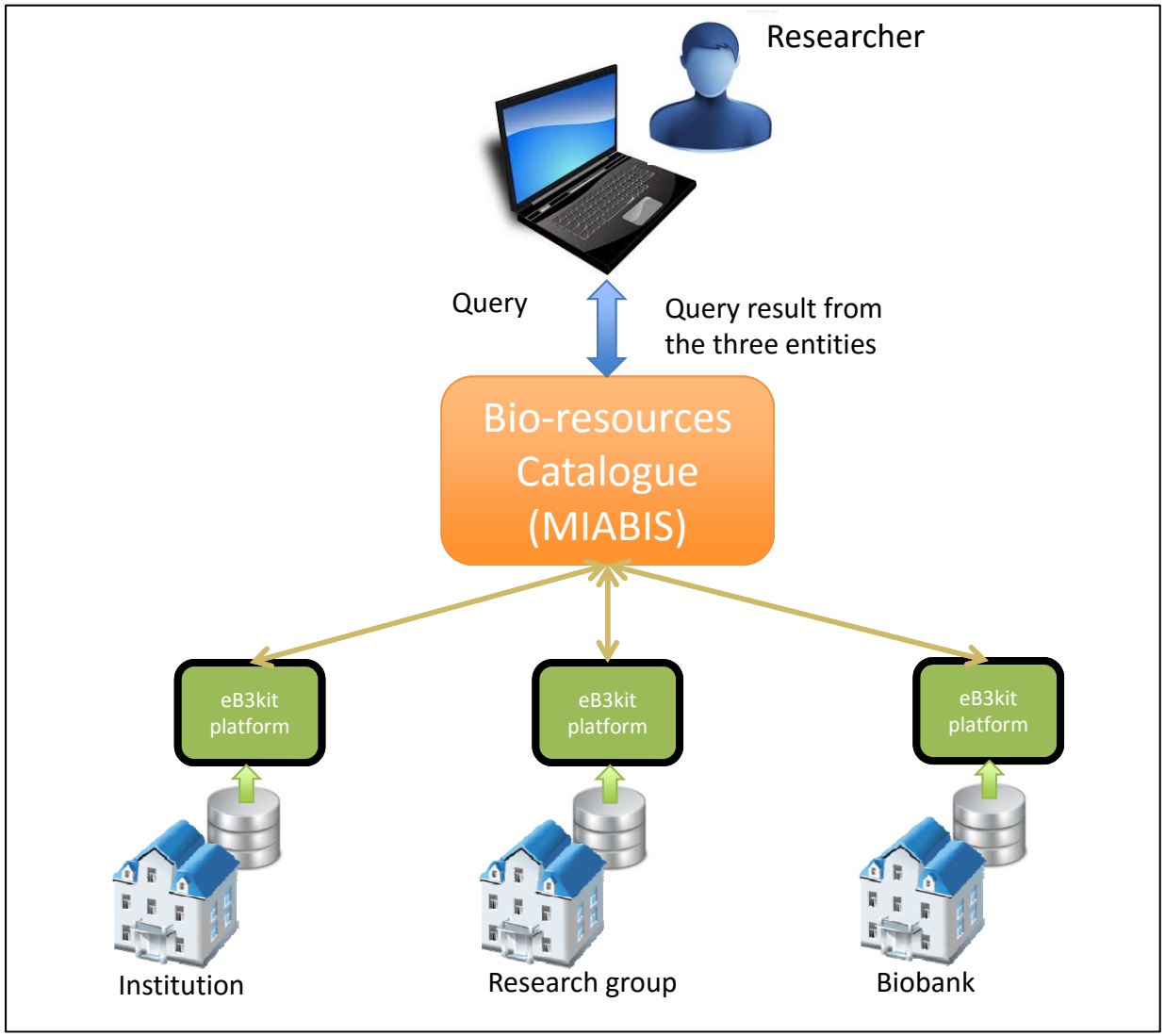
Trinity (eukaryotes): +200 GB

Some examples available to partners which we could connect to

Option	Model	Cores	RAM	Disk capacity	Sufficient for this analysis	H3ABioNet nodes with a server
Option A	Dell C6145	64	256GB	10TB	Yes	Rhodes, CPGR, ICIPE, IPT, CBIO
Option C	Dell C6145	24	128GB	10TB	Yes	MUHAS, NMIMR, KCCR, UDSM
Uganda HP 1	HP DL360e	16	64GB	3.6TB	Yes	UVRI
Uganda HP 2	HP DL385p	16	128GB	6TB	Yes	UVRI
Tanzania Dell 1	Dell	8	144GB	12TB	Yes	MDH

INSTALLATION

3 dockers (already installed in demo server):

- Galaxy docker
  - Port 8101 --> 8080 (internal)
  - Ubuntu server 14.04 LTS
  - MySQL 5
  - Galaxy code
- Galaksio docker
  - Port 8102 --> 80 (internal)
  - Ubuntu server 14.04 LTS
  - Apache 2 server
- *Data storage* docker
  - Keeps the databases and user's files out of the Galaxy docker.

LONG TERM AIM