



genomic
STANDARDS *consortium*
biodiversity working group

Biodiversity Working Group Report

Robert J. Robbins

rjr8222@gmail.com

GSC-13, Shenzhen China

5-7 March 2012

GSC Biodiversity Working Group = GBWG

What is GBWG doing (and why)?

GBWG (with assistance from RCN4GSC) is reaching out to other communities to engage scientists at the interface of genomics and biodiversity.

The
Research Coordinating Network for the Genomic Standards Consortium
Project

FUNDED BY NSF, HOSTED AT UCSD, AND SUPPORTING THE WORLDWIDE SCIENTIFIC COMMUNITY

The NSF RCN4GSC project at UCSD has the mission to:

create a research coordination network to promote and integrate standards for genomic and metagenomic data and metadata within an international community. The network is based on the existing Genomic Standard Consortium and will be extended under this award to include ecological data standards such as Ecological Metadata Language, biodiversity standards such as Darwin Core, and environmental research programs such as the Global Lake Ecological Observatory Network and Long Term Ecological Research.

The
Research Coordinating Network for the Genomic Standards Consortium
Project

FUNDED BY NSF, HOSTED AT UCSD, AND SUPPORTING THE WORLDWIDE SCIENTIFIC COMMUNITY

The NSF RCN4GSC project at UCSD has the mission to:

create a research coordination network to promote and integrate standards for genomic and metagenomic data and metadata within an international community. The network is based on the existing Genomic Standard Consortium and will be extended under this award to include ecological data standards such as Ecological Metadata Language, **biodiversity standards such as Darwin Core**, and environmental research programs such as the Global Lake Ecological Observatory Network and Long Term Ecological Research.

The
Research Coordinating Network for the Genomic Standards Consortium
Project

FUNDED BY NSF, HOSTED AT UCSD, AND SUPPORTING THE WORLDWIDE SCIENTIFIC COMMUNITY

The NSF RCN4GSC project at UCSD has the mission to:

create a research coordination network to promote and integrate standards for genomic and metagenomic data and metadata within an international community. The network is based on the existing Genomic Standard Consortium and will be extended under this award to include ecological data standards such as Ecological Metadata Language, biodiversity standards such as Darwin Core, and environmental research programs such as the Global Lake Ecological Observatory Network and Long Term Ecological Research.

RCN4GSC held an initial biodiversity engagement workshop in San Diego in March, 2011

Why the big push into biodiversity?

- Biodiversity is less a field of biology than a perspective (that of variance) into biology.
- Diversity is a sine qua non of biology; no diversity, no evolution.
- Genetics / genomics are equally central to biology.
- Probably half of the world's biomass and by far most of its biodiversity exists in microbial communities, many of which have been effectively invisible until recently. Understanding this biological dark matter is critical to a full understanding of the biosphere.

- Metagenomics tools are showing that, compared to macro-scale organisms, the diversity of microbial communities is staggering.

Intra-species bacterial genetic diversity is greater than that among the great apes; intra-genus bacterial diversity is greater than that among all the mammalia.

- Metagenomics tools are showing that a full understanding of macro-scale organisms will depend on an understanding of their interactions with their associated microbiomes.

Understanding how different ants optimize nutrient acquisition, and thus how they function in their niches, depends on an understanding of their associated gut microbiomes.

Why the big push into biodiversity?

- Metagenomics tools are showing that, compared to

**The connection between genomics/
metagenomics and biodiversity
seems obvious and profound.**

associated microbiomes.

Understanding how different ants optimize nutrient acquisition, and thus how they function in their niches, depends on an understanding of their associated gut microbiomes.

Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Characterize Benefits

Identify Opportunities

Assess Challenges

Propose Next Steps

Purposes of the March, 2011, RCN4GSC Workshop

→ Develop Conceptual Framework

Characterize Benefits

Identify Opportunities

Assess Challenges

Propose Next Steps

Efforts to discover, elucidate, and understand the planet's biodiversity must span the macro, meio, and micro scales and must integrate traditional tools with emerging genomic and metagenomic methods. To pursue these in isolation would violate biological reality and be counter-productive to knowledge discovery.

Integration/interoperability between macro and genomic biodiversity informatics is scientifically desirable, technically feasible, and should be tackled.

Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Integrative comparative analyses of genomic/macro biodiversity data sets will enable more efficient and effective research, and will support more far-reaching interdisciplinary research and discoveries.

→ Characterize Benefits

Coherent efforts to deploy standard genomic and metagenomic methods across the multiple dimensions of biodiversity will achieve economies of scale, produce mutual benefits, generate multiplier effects, leverage support, and yield more comprehensive science.

Identify Opportunities

Assess Challenges

Propose Next Steps

Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Characterize Benefits

→ Identify Opportunities

Assess Challenges

Propose Next Steps

Opportunity to benefit multiple communities through harmonization of tools and concepts across communities.

Opportunity to advance biology through substantial extension and refinement of such fundamental concepts as individual or species.

Opportunity for consilience resulting from growth of aggregators (e.g., GBIF, EOL, Genbank, GSC, KNB, LTER, NEON, DATAONE, EMP) that are serving most of the data.

Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Will need to harmonize fundamentally different sampling units in macro and genomic surveys and inventories.

Characterize Benefits

Will need to deal with inconsistent implementation by INSDC of different standards, e.g., GSC and Barcode standards such as occurrence, collector data, etc.

Identify Opportunities

Will need to extend data “adherence” measures from **presence** of data, to **quality** of data.

→ Assess Challenges

Will need to reach out to a multitude of fragmented biodiversity communities.

Propose Next Steps

Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Characterize Benefits

Identify Opportunities

Assess Challenges

→ Propose Next Steps

Write and publish a white paper on biodiversity, genomics, and metagenomics.

Scope and implement a focused pilot project/working group to (a) explore how to integrate genomic and macro biodiversity data and (b) explore the integration of the Darwin Core and GSC standards by mapping and reconciling the elements of the Darwin Core and GSC checklists.

Develop a set of use cases.

Examine the lessons learned from the test/pilot project and use-cases to ascertain appropriate next steps.



Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Write and publish a white paper on

All of this seems dependent upon
Conceptual harmonization
Nomenclatural standardization

Examine the lessons learned from the test/pilot project and use-cases to ascertain appropriate next steps.



Purposes of the March, 2011, RCN4GSC Workshop

Develop Conceptual Framework

Write and publish a white paper on

All of this seems dependent upon
Conceptual harmonization
Nomenclatural standardization

How important is this?

test/pilot project and use-cases to
ascertain appropriate next steps.

孔夫子



孔夫子



When asked,
“What is the first thing to be done, in order to
administer the government?”

子曰
必也正名乎



The master said:

What is necessary is to rectify names.

子曰
必也正名乎



When challenged as to whether this was truly the top priority, the master responded...

名不正
則言不順
則事不成



If names be not correct, language is not in accordance with the truth of things. If language be not in accordance with the truth of things, affairs cannot be carried on to success.

名不正
則言不順
則事不成



If names be not correct, language is not in accordance with the truth of things. If language be not in accordance with the truth of things, affairs cannot be carried on to success.

正名



Conclusion:

The rectification of names – i.e., achieving conceptual harmonization and developing nomenclatural standards – is a critical first step (perhaps THE critical first step) in any administrative undertaking.

Really?! What's in a name?

Does it really matter if we have subtle differences in meanings and referents for words?

Examples:

- Personhood

Examples:

- Personhood
- Gene – locus – position distinctions

Thought experiment:

Imagine you have a strain of *E. coli* in which a translocation has moved the lac operon 180 degrees around the chromosome. What is the best way to describe the situation?

The lac gene has moved to a new locus.

The lac locus has moved to a new position.

Examples:

I once asked this lac operon question of a room full of *E coli* researchers at an ASM meeting.

The result:

They were unanimous in their feeling that this was a stupid, trivial question.

Examples:

I once asked this lac operon question of a room full of *E coli* researchers at an ASM meeting.

The result:

They were unanimous in their feeling that this was a stupid, trivial question.

They split 50:50 on which answer was in fact trivially correct.

Food for thought:

Such differences of opinion are of little consequence for the practice of science.

But they make the development of formal information-management systems devilishly difficult.

Paradoxes arise from conceptual fuzziness:

- Ship of Theseus

Problems associated with persistence of individuality and replacement of parts.

- White Horse Dialogue

Problems associated with the interacting relationships of individuals, populations, sets, categories, and their attributes.

Paradoxes arise from conceptual fuzziness:

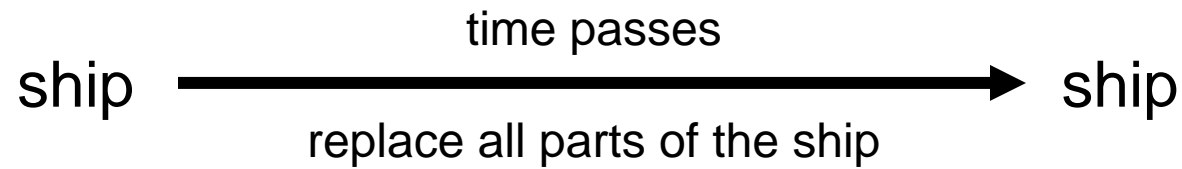
It is not a good idea to have paradoxes lurking near the core of formal information-management systems.



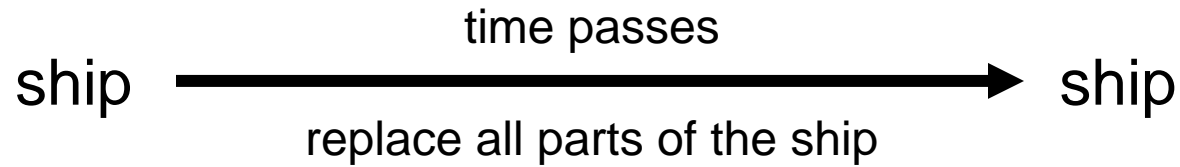
genomic
STANDARDS *consortium*
biodiversity working group

First Paradox

Ship of Theseus:

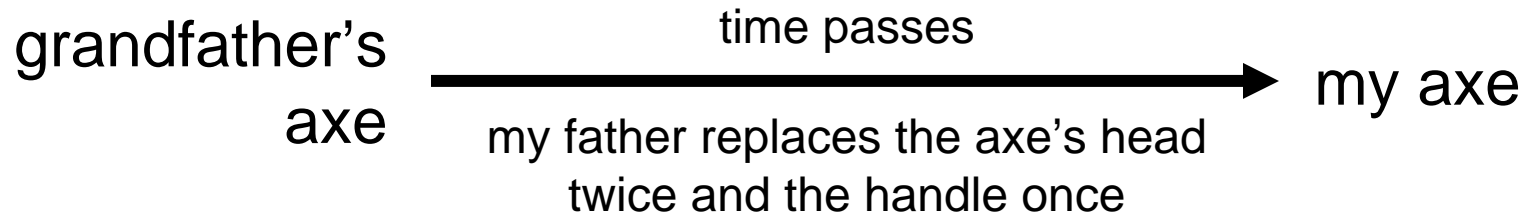


Ship of Theseus:



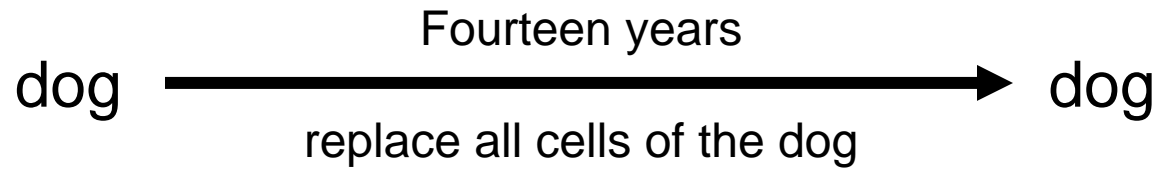
Same ship?

Grandfather's axe:

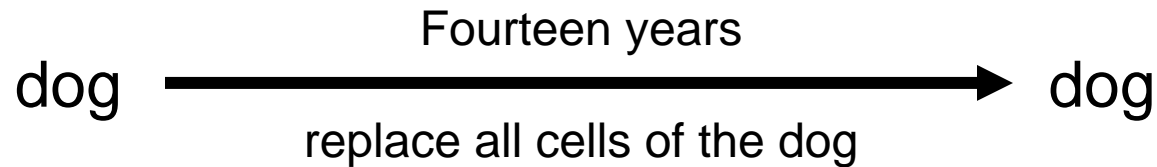


Same axe?

Biological Example I:

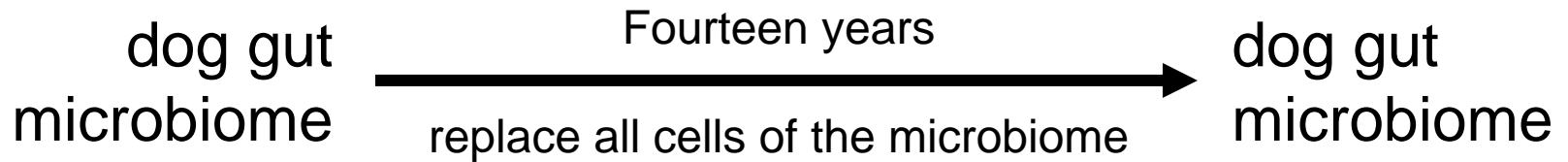


Biological Example I:



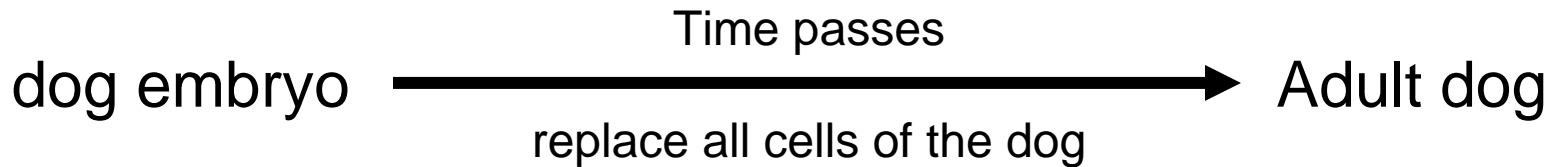
Same dog?

Biological Example II:



Same microbiome?

Biological Example III:



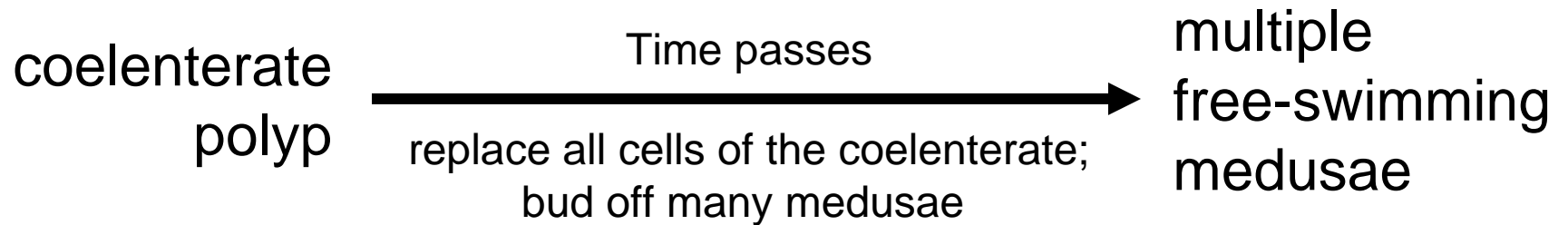
Same dog?

Biological Example IV:



Same person?

Biological Example V:



Same coelenterate(s)?

Food for thought:

Many of our most fundamental concepts about biological entities and identities (and therefore the anchors on which to attach attributes) seem closely tied to the idea of hierarchical aggregations of physically connected, clonally derived, genetically identical cells – i.e., multicellular bodies – that inherit their genetic information linearly from their immediate ancestors.

Food for thought:

Many of our most fundamental concepts about

**Virtually none of these conditions
are generally applicable to
microbial communities.**

information in many from their immediate
ancestors.



genomic
STANDARDS consortium
biodiversity working group

Second Paradox

White Horse Dialogue - 白馬論

Asked the question, can it be that a white horse is not a horse?

<http://ctext.org/gongsunlongzi/bai-ma-lun>

白馬非馬

White Horse Dialogue - 白馬論

Asked the question, can it be that a white horse is not a horse?

<http://ctext.org/gongsunlongzi/bai-ma-lun>

But, what if instead of “white horse not horse?” we asked “feathered dinosaur not dinosaur?”

Hmmm. That seems less far-fetched and in fact becomes a legitimate question regarding the classification of biodiversity over evolutionary time.

白
馬
非
馬

Food for thought:

The number of possible bacterial generations since John Snow did his epidemiological work on cholera is roughly the same as the number of hominid generations since...

Food for thought:

The number of possible bacterial generations since John Snow did his epidemiological work on cholera is roughly the same as the number of hominid generations **since the last common ancestor of chimps and humans.**

Food for thought:

The number of possible bacterial generations since John Snow did his epidemiological work on cholera is roughly the same as the number of hominid generations since the last common ancestor of chimps and humans.

Maybe our macro-scale based intuition about evolutionary events (and the applicability of “feathered dinosaur not dinosaur” thoughts) will not apply well at the microbial level.

More food for thought:

Other scaling issues arise across the mega- and micro-biological scales. E.g., should microbial samples taken one meter apart be considered equivalent?

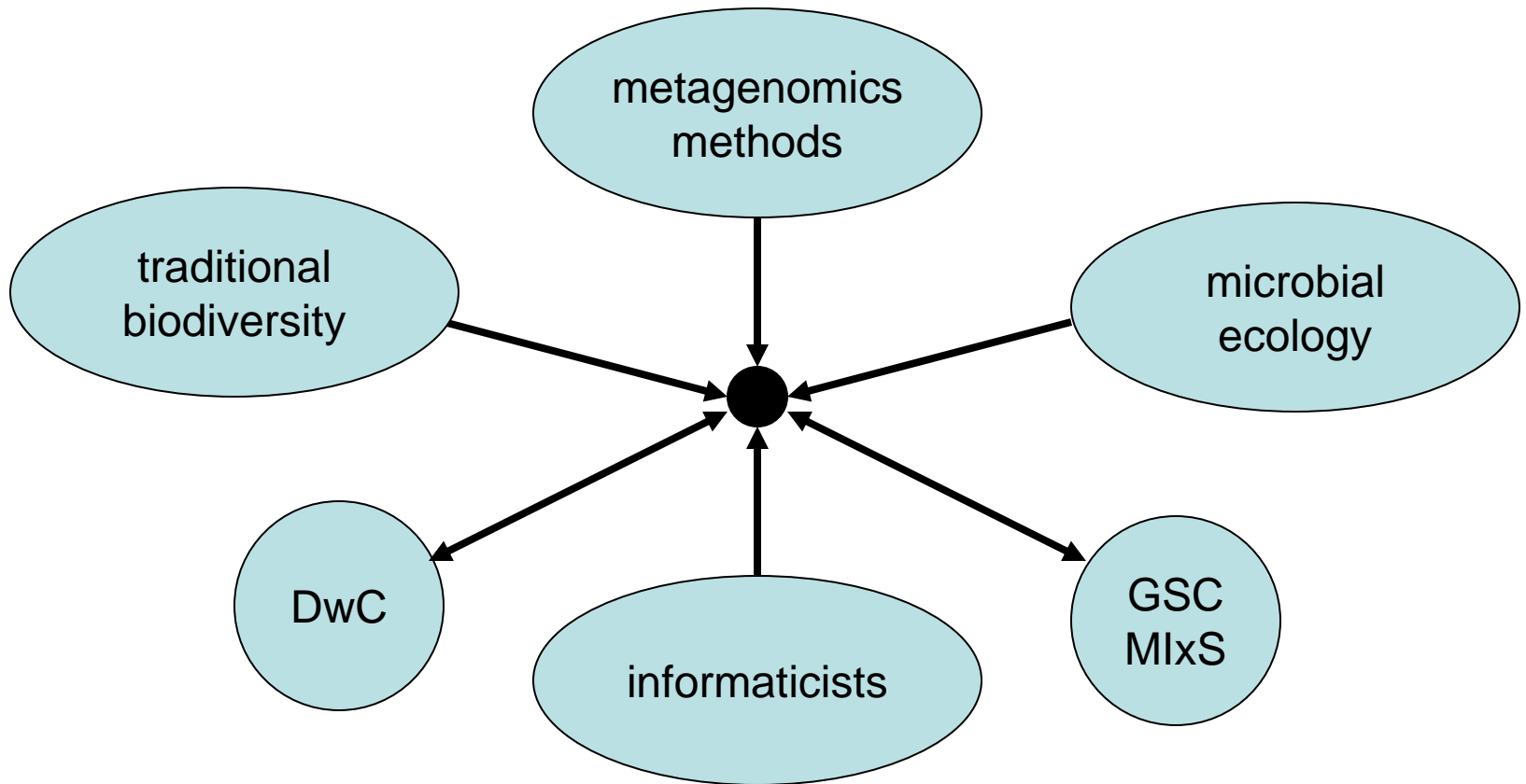
More food for thought:

Other scaling issues arise across the mega- and micro-biological scales. E.g., should microbial samples taken one meter apart be considered equivalent?

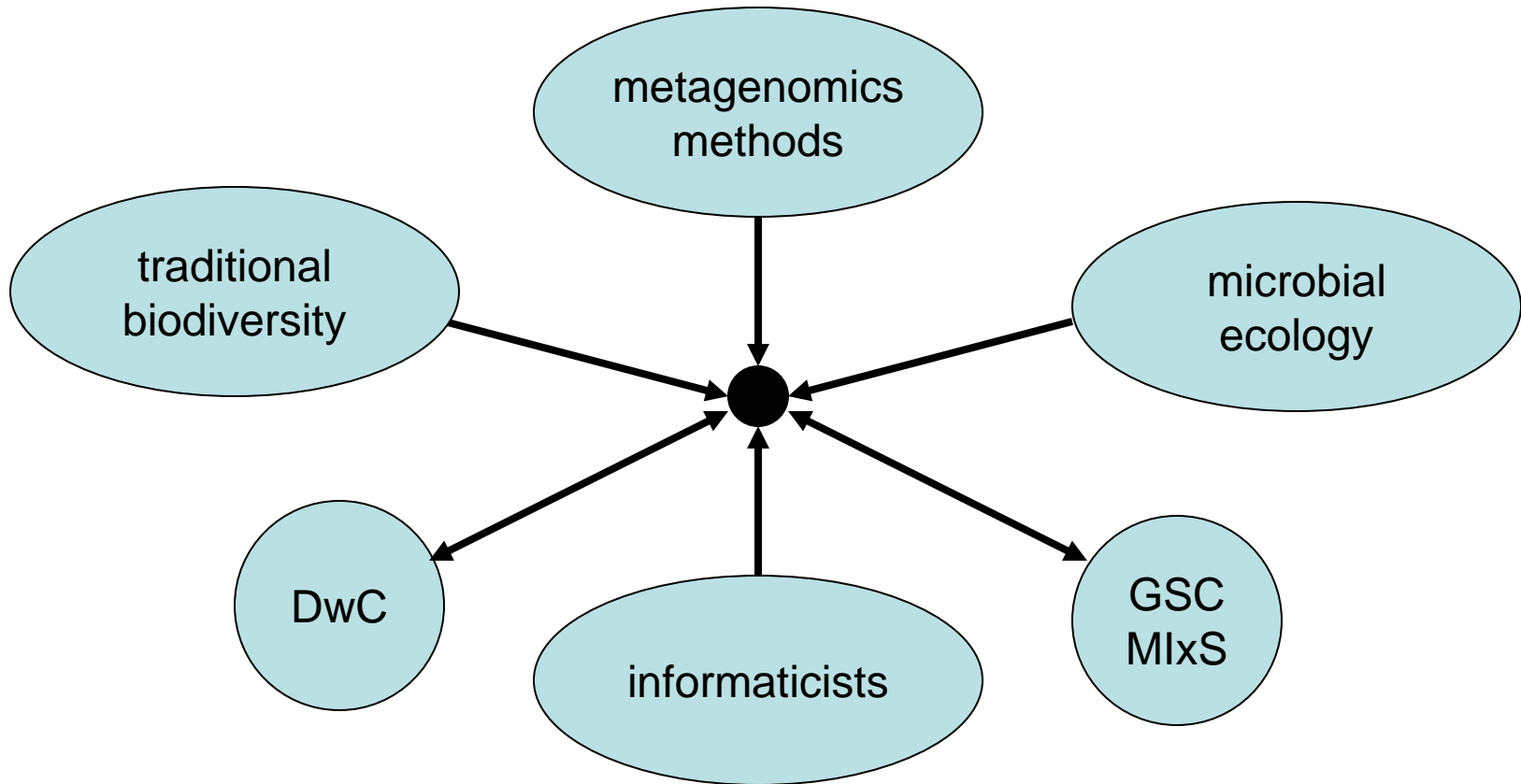
Scaled for body size, that same distance would be

mice: 25 miles elephants: 1250 miles

What is GBWG doing (conceptually)?



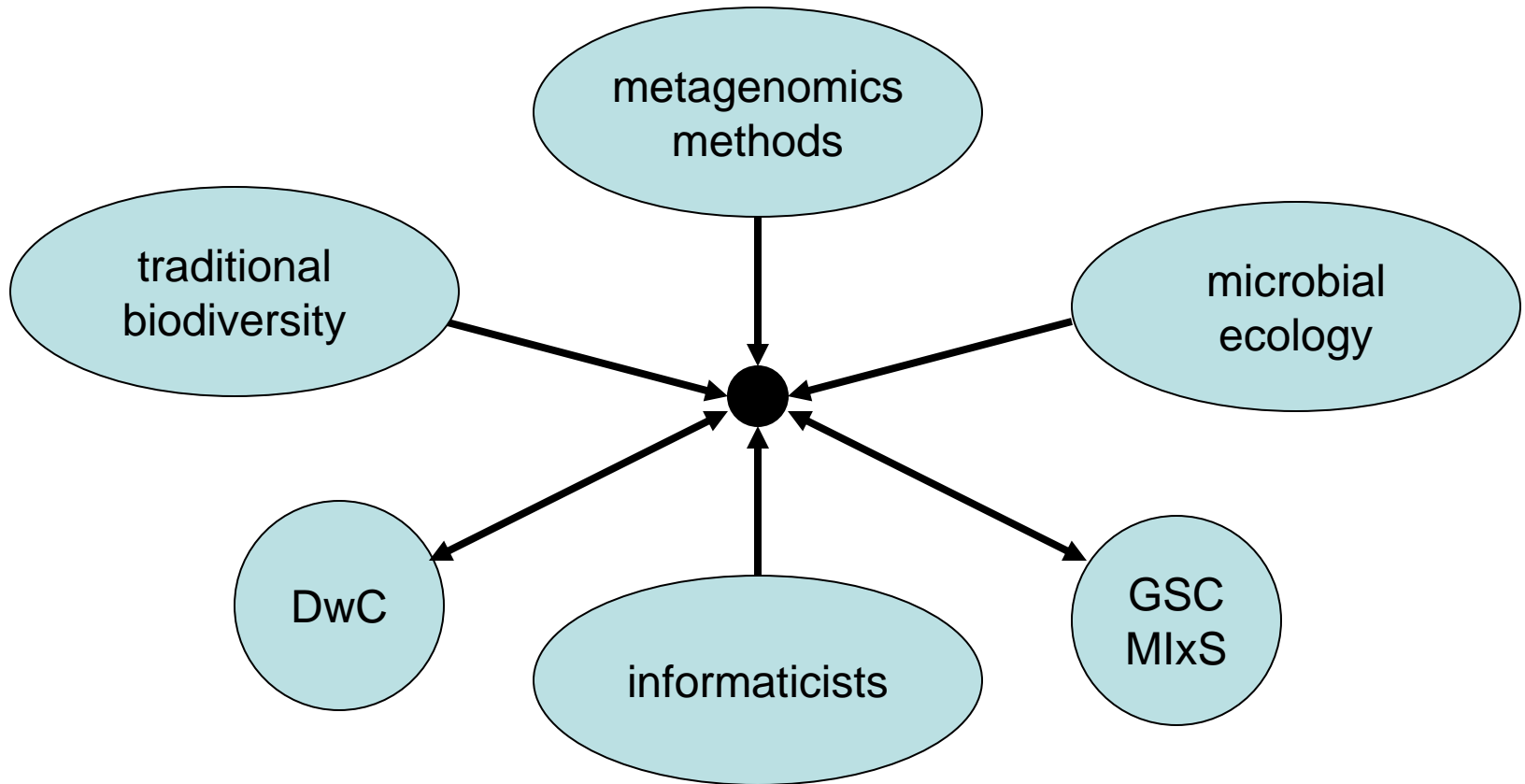
What is GBWG doing (conceptually)?



Bringing together members of different communities (often introducing them to each other for the first time) to

explore scientific opportunities,

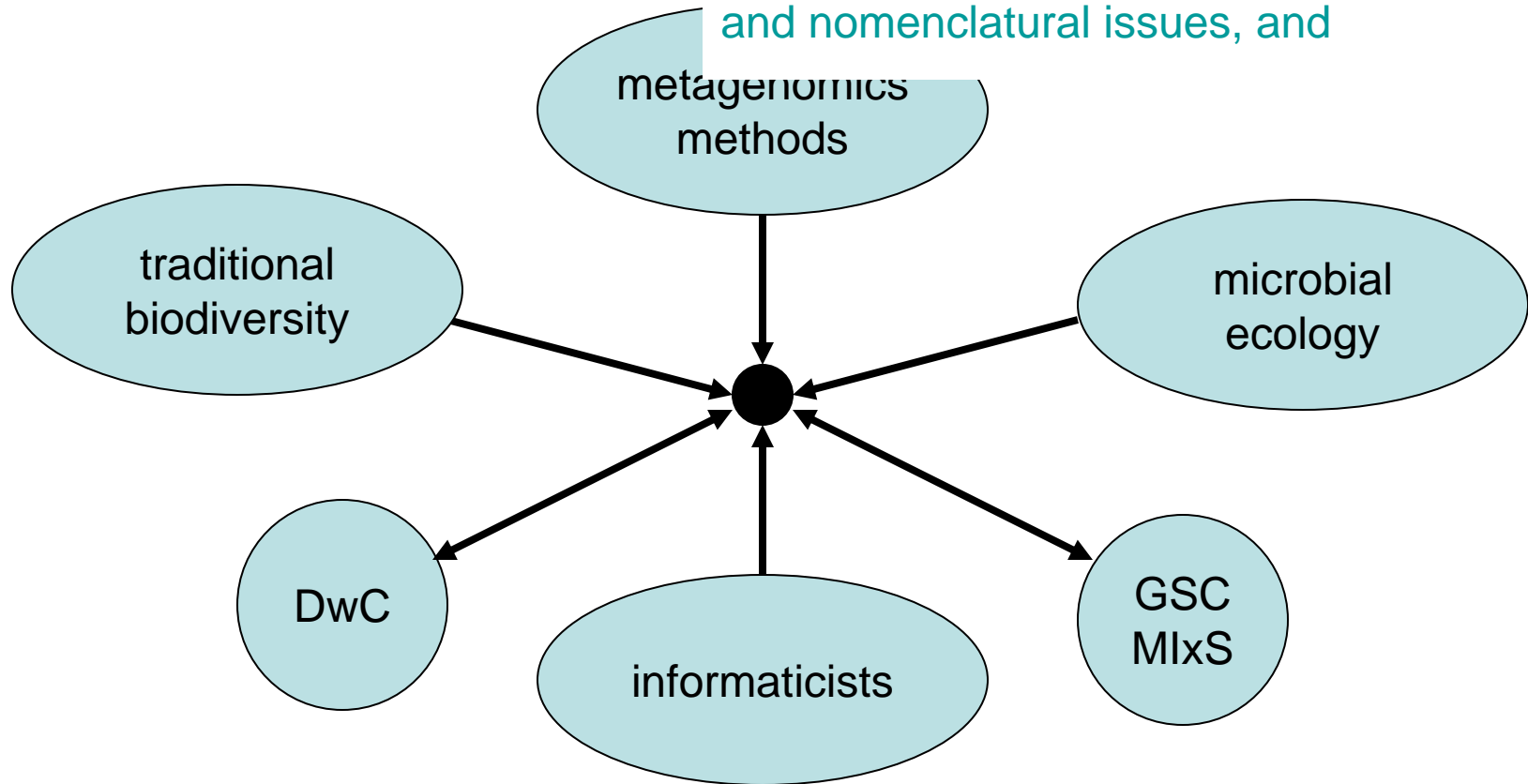
What is GBWG doing



Bringing together members of different communities (often introducing them to each other for the first time) to

What is GBWG doing

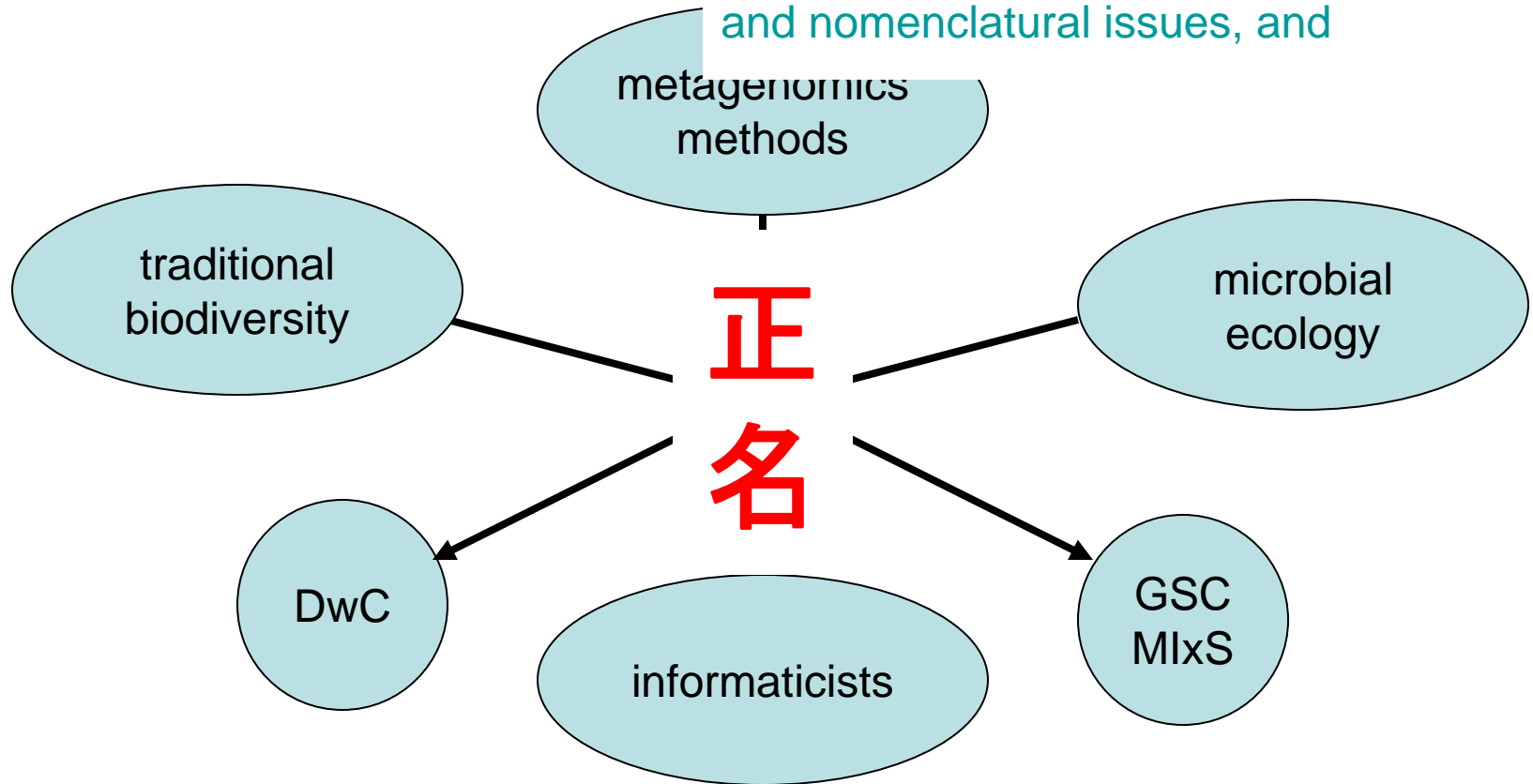
explore scientific opportunities, identify and consider potential conceptual and nomenclatural issues, and



Bringing together members of different communities (often introducing them to each other for the first time) to

What is GBWG doing

explore scientific opportunities, identify and consider potential conceptual and nomenclatural issues, and





HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)





HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)



Our guiding principles:

When it comes to standards,

$$n > (n + 1)$$

$$(n - 1) > n$$

What has GBWG done (practically)?

- Apr 2010 GBWG created at GSC9 @ JCVI
- Mar 2011 Initial meeting in San Diego
- Apr 2011 GSC11 @ Hinxton
- May 2011 Second meeting in San Diego
- Sep 2011 GSC12 @ Bremen
- Oct 2011 Engage TDWG at their annual meeting in New Orleans
- Jan 2012 BDWG planning workshop @ JGI
- Feb 2012 GBIF Hackathon @ Oxford
- Mar 2012 GSC13 @ Shenzhen

What will GBWG do (practically)?

May 2012 Biodiversity & Genomics Ontology Workshop @ U Kansas

Sum 2012 Asian Outreach / Connection ?

Jul 2012 GBIF engagement in Global Biodiversity Informatics Conference

Sep 2012 LTER engagement @ all scientist meeting ?

Sep 2012 GSC14 @ Oxford

Oct 2012 Participate in TDWG annual meeting in Beijing

Late 2012

DwC GSC lessons-learned, standards update meetings ?

Early 2013

What will GBWG do (practically)?

May 2012 Biodiversity & Genomics Ontology Workshop @ U Kansas

Sum 2012 Asian Outreach / Connection ?

Jul 2012 GBIF engagement in Global Biodiversity Informatics Conference

Sep 2012 LTER engagement @ all scientist meeting ?

Sep 2012 GSC14 @ Oxford

Oct 2012 Participate in TDWG annual meeting in Beijing

Late 2012

DwC GSC lessons-learned, standards update meetings ?

Early 2013

END