Dealing with the sins of bioinformatics

About the sins and virtues of SARA's HPC cloud in the context of e-BioGrid

Machiel Jansen

e-BioGrid

e-Biogrid is the **portal**

to the **BiG Grid** **eScience** environment

for **Life Sciences**

Mass spectrometry **Micro-array**

Nanoscopy and imaging

Structure determination and prediction

Next-generation sequencing

Biobanking

e-BioGrid

Get involved

We **offer support** to both **bioinformaticians** and **biologists** in the life-science research community.

We invite **you to share difficulties**, **obstacles** and **problems** you encounter in life-science research.

e-BioGrid

Who are we?

Joost Kok

Timo Breit

Irene Nooren

Jan Bot

Machiel Jansen

Coen Schrijvers

Floris Sluiter

Pieter van Beek

Tom Visser

BioInformatics

It is a **sinful** world.

It always **me-science** instead of **e-science**.

Professor Carol Goble, University of Manchester

nbic netherlands bioinformatics centre

BiG Grid the dutch e-science grid

MANCHESTER 1824

Sins

1. Parochialism and Insularity
2. Exceptionalism
3. Autonomy or death!
4. Vanity: Pride and Narcissism
5. Monolith Meglomania
6. Scientific method Sloth
7. Instant Gratification

Sins and reality

A **sin** is a **deviation** of an **ideal**.

Normative:

People **sin** when **they do not act as they are supposed to do**… relative to some divinely inspired criterion.

Pragmatic:

Instead of seeing sins **accept peoples behaviour** and understand their motives.

Sins and reality: push versus pull

Seeing **sins** means **pushing** people

In a **push strategy** the information about a product or service is **broadcast to an audience**. The information goes out from the provider to the user. **The user can only listen.**

Push strategy is a provider's monologue.

Push strategy is non-interactive

Pull

Accepting peoples choices means allowing them to pull

In a pull strategy the provider can broadcast information and the user can respond. The user can also search for a provider, find the provider and make a request for information about a product, service or infrastructure. **Pull strategy is a conversation between buyer and seller**. **Pull marketing is interactive.**

The virtues of bioinformatics

How are **bioinformaticians** supposed to work according to Goble?

Bioinformaticians should **reuse** each others **data** and **services**, use **workflows** (Taverna) and **collaborate**.

If they **don't** they **sin**

Sin 1: the wheel

Reinventing the wheel **is a sin**.

It **can be** very inefficient

but **is it** a **sin**?

Reinventing wheels may lead to innovation

The virtues of reinventing wheels

- "**reinventing** the **wheel**" is an important tool in the **instruction of complex ideas**

- It helps in **truely** understanding existing technology and may lead to **new solutions** to existing, or slightly different problems: it stimulates **innovation**.

Inventing the cloud (again)

SARA's HPC cloud is **yet another cloud**

and **yet it isn't**

Inventing the cloud

Users have **their own** managed **environment** and have everything **under their own control**.

Everybody can feel **special**

Which brings is to Goble's second sin: bioinformaticians think they are special

Biologist exceptionalism

I'm different. We are all individuals.

- I know there is already a gene name for that gene, but, I don't like it and it doesn't fit in with my schema.
- It would be better if I wrote the script I need so I know what it does, how it does it and how to modify it later because I haven't specified what it was supposed to do in the first place.

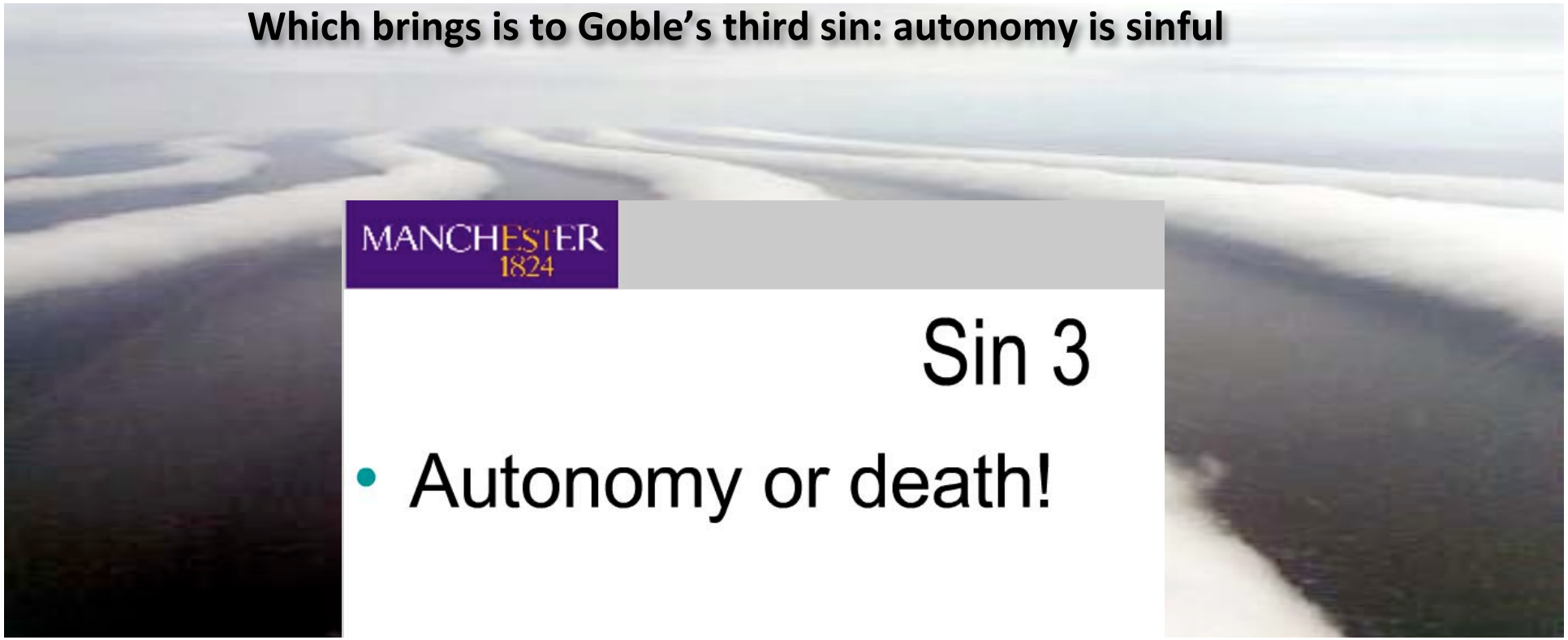
Inventing the cloud

Users are can keep **doing their own things**.

They are **not forced** to give up any **control**.

They are completely **autonomous**

Which brings is to Goble's third sin: autonomy is sinful



Sins on cloud

- All those **sinful** **bioinformaticians** are still **allowed to sin**.
- They have **their own environment**. But they can **share – if they want to**, not because they have to.
- The are **completely autonomous**. Root on their own system, install their own software, even allowed to reinvent their own wheels
- They **can learn from each other** – we bring them together. The cloud is a **central infrastructure** and user experiences **can be shared**.
- They are **not pushed**

The virtues of cloud

Yes you can **meet** and **share** and **reuse**

When you push a Life Scientist…