

Return on Investment for HPC at Loughborough

Experiences

Aaron Turner, IT Services Senior Specialist in HPC, Loughborough University
Lizzie Gadd, Research Office, Loughborough University

Why Calculate Return on Investment?

Assess value of business units and show value

Ensure business units are delivering and discover when they are not

Inform senior management, stakeholders to inform strategy

Part of making a case for increased investment

Potentially required by funders in some cases

Want to know if:

Return > 0 , Return $>$ investment, Return \gg investment

Return < 0

What is Return?

Depends on stakeholder

For the university (publications, grant income, reputation, etc.)

For researchers (career prospects)

For funders/society (impact of research)

What is Return?

Depends on stakeholder

For the university (publications, grant income, reputation, etc.)

More . more income

For researchers (career prospects)

For funders/society (impact of research)

What is *not a* Return?

KPI/Metric is *not necessarily* the same as a return

But we measure KPIs as proxies in some instances

What is *not* a Return?

KPI/Metric is *not necessarily* the same as a return

But we measure KPIs as proxies in some instances

Not the same as service sustainability

Policy decision on returning income to service

What is Investment?

Many ways of calculating what the investment is

What is actually included?

Often infrastructural elements are shared

What is Investment?

Many ways of calculating what the investment is

What is actually included?

Often infrastructural elements are shared

Different stakeholders, different views of what the investment is

What is Investment?

Many ways of calculating what the investment is

What is actually included?

Often infrastructural elements are shared

Different stakeholders, different views of what the investment is

Regional centres: EPSRC view != regional consortium view

Important Factors for useful ROI calculation

Specific

Measurable

Accurate

Relevant

Timely

Important Factors for useful ROI calculation

Specific

Measurable (Meaningful)

Accurate (Automatable)

Relevant (Reproducible)

Timely

Specific

Needs to measure things that are specifically attributable to HPC

Specific

Needs to measure things that are specifically attributable to HPC

Which bits of a grant are just HPC?

Do we count the whole grant? Is that valid?

Measurable

Not all impact is actually measurable

We can only really work with what we can actually measure objectively

Not everything is meaningfully convertible to £

Accurate

Can't make sense of metrics if they are not accurate.

Sometimes data in management information systems isn't accurate.

Sometimes data in institutional repositories isn't accurate.

Can we get a confidence measure?

Sensitivity analysis? Imputation?

Relevant

Data that actually informs decision making

Can be directly compared against similar types of facilities

Timely

Needs to be

sufficiently fresh

frequent

cover appropriate time periods

Impact may not emerge immediately for projects

Heisenberg's Uncertainty [principle] of Return Timeliness

Metrics : Relatively Easy to Gather

Outputs

#/size/type grants/income
Papers, publication quality
Patents, spin outs
Commercial income
SME engagements
Projects
Users: PIs, researchers, PhD
PhD students graduated
Satisfaction surveys

Inputs

Time used
Facility cost
System properties

Metrics : Hard/Harder to Determine

Proportion of outputs directly attributable to HPC alone/partially

± Halo effect → grants not using HPC that are still enabled by it?

Effect/value of SME engagement

Disambiguating some system costs

What is the 20 year impact?

Effect on a researcher's career long term?

Improved retention of a researcher whose career has been improved?

Effect does providing seed time have?

Hypotheticals

Would waiting have allowed a smaller investment with a better return?

Opportunity costs of spending at time X rather than delaying to Y?

Would a different system architecture have offered better return?

Would a larger system have been better?

Is the system size at any given time appropriate

too big at one point, too small at another?

Driven by CAPEX/OPEX

What would have happened if the HPC was not there: other research?

Which Metrics Are Important?

We tend to assume what is and what is not important

We may assume what is easiest to gather is likely to be important

Confirmation bias

Ideally we need statistical analysis

Assuming we can get reliable enough data to even do this.

Or some clever Deep Learning technique?

Comparison

Very hard to convert all metrics into £ value

R:I hard to compute

Comparison

Very hard to convert all metrics into £ value

Should they be?

R:I hard to compute

Loughborough's Experience

Finding the right data was difficult

Required working with multiple other business units to find data

Data was not clean

Relies on people to input the right data

Poor ability to integrate data

Identity management was an issue

It is hard to quantify errors without a statistical analysis

It took time

Finding the data

Tends to exist in multiple systems

Integration between systems is often poor

Often there is no single key to link data relating to a project on HPC

Makes it hard to automate

Cost of data gathering is high

Affects specificity, accuracy and ability to cost-effectively conduct ROI exercises

Data is not clean

Sometimes same paper multiply entered in databases, with small variations

Need a model of those variations to deduplicate

It may not be clear who had what status on a project at a point in time

Data may be missing

Need a model to impute

How do you know without good integration?

Affects ROI accuracy if cleaning is not reproducible, specific, sensitive

Poor Integration

Many data sources and systems

Need good, stable APIs, time to integrate

Cannot automate overall process

Grant information not linked to HPC Project IDs

Not all work grant funded (e.g. Seed time)

Lack immediate links to financial DBs

Judgement currently required to integrate

Poor reproducibility

Uncertain accuracy

Identity Management

Many forms of ID over time/role

Was a person staff, student at a particular point in time?

Recycling of UNIX type usernames

Hard to find out who graduated

Quantifying Errors

Sources of errors:

- Lack of data

- Cleanliness of data

- The use of judgement to deal with some of the above

- Inaccuracy in data cleaning methods

Quantifying:

- Estimating/rule of thumb

- Proper statistical analysis via sensitivity analysis of some sort

Time

Between 2 and 3 person-weeks

Loughborough's ROI

Can't release the figures

Healthy (> 1) based on grant income £

Loughborough's ROI

Can't release the figures

Healthy (> 1) based on grant income £

But:

The results aren't 100% accurate

Judgement calls were made

What Loughborough is doing : Short Term

Record grants as part of HPC projects

Use grant IDs to track papers

Possibly mint pseudo-grants IDs

Track via new services

e.g. ResearchFish, ORCID

What Loughborough is doing : Change Academy

Integrate data sources

Work towards automation to allow ROI measurement to be:

cheaper

more timely

Work towards other facilities using same methodology

Improvements in identity management (*not just for ROI*)

What Loughborough is doing : Next time

Ideally proper statistical analysis

Stats lies outside ITS/RO/etc. core competencies at present

Quantify accuracy to some extent

Will need to be done again by hand pending Change Academy work

Work out how best/better to present the results

Recommendations

Ensure data has keys to enable links between data

Ensure the data is input automatically where possible or is sanity checked

Fix identity management

Ensure the ability to automate connection to databases, APIs

Use statistical techniques to quantify confidence in figures and quote them

Pick the metrics that are SMART, but what is Relevant?

Find the best way to present the results

Ensure that all other major facilities are on the same basis

Work with RCUK to improve tools, approaches, etc.

Comments and Discussion

What are your experiences?

How does your institution view the value of ROI calculations?

What is best practice, how can we share it?

Who owns/should own the ROI calculation process?

What tools or central services help or could help?

e.g. ResearchFish

Are there other useful metrics?

How/can all metrics be boiled down to just £/\$/” ?

Modelling counter-factual scenarios and/or prediction?