When History Matters

Assessing Reliability for the Reuse of Scientific Workflows

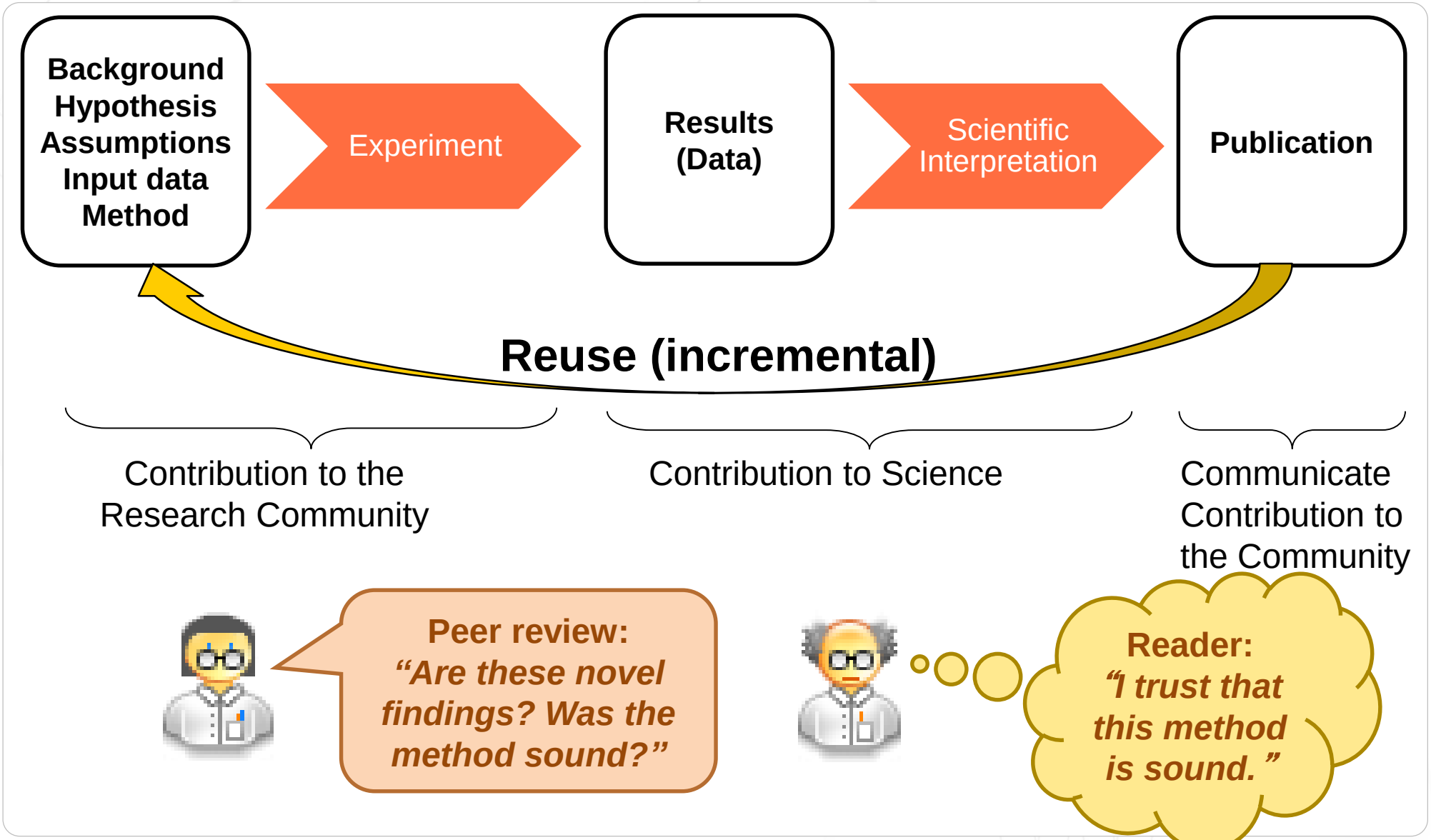
José Manuel Gómez-Pérez, Esteban García, Aleix Garrido, José Enrique Ruiz, Jun Zhao, Graham Klyne

Semantic Web in Use - ISWC
23rd October 2013

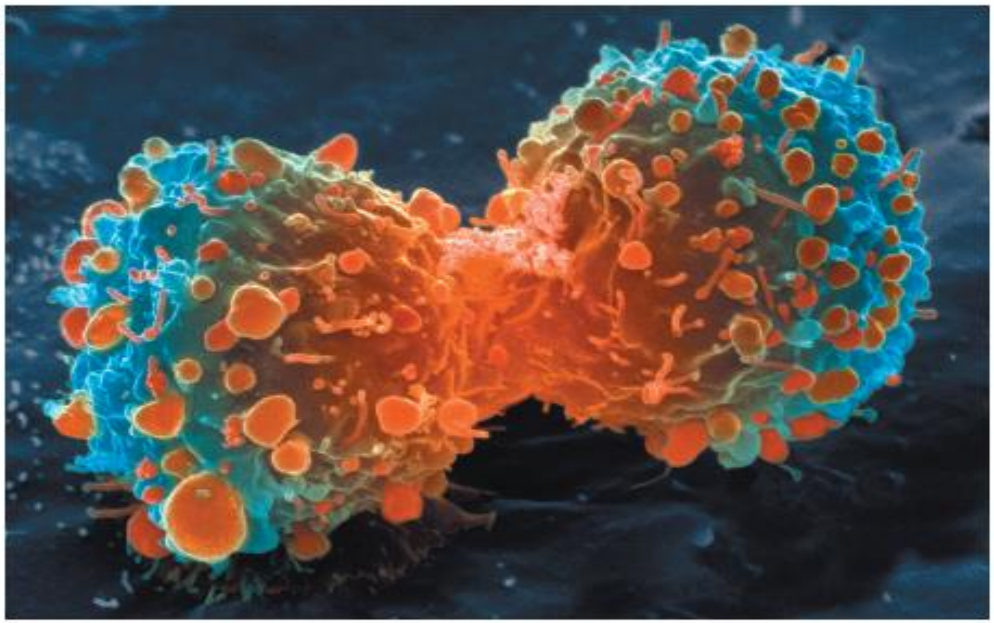
iSOCO
enabling the networked economy

Workflow 4Ever

The research life cycle



COMMENT



S. GSCHMEISSNER/SPL

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical suc

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this lowe

investigators must reassess their approac translating discovery research into grea clinical success and impact.

Many factors are responsible for the h failure rate, notwithstanding the inh ently difficult nature of this disease. C tainly, the limitations of preclinical to such as inadequate cancer cell line

47 of 53 “landmark” publications could not be replicated

Inadequate cell lines and animal models

Nature, 483, 2012

Carole Goble, JCDL 2012 Keynote

nature biotechnology

NATURE BIOTECHNOLOGY | COMPUTATIONAL BIOLOGY | NEWS FEATURE

My data are your data

Vivien Marx

Nature Biotechnology **30**, 509–511 (2012) | doi:10.1038/nbt.2243
Published online 07 June 2012

Encouraging more broad and inclusive data sharing in today's world will involve concerted community efforts to overcome technical barriers and human foibles. Vivien Marx investigates.

Introduction

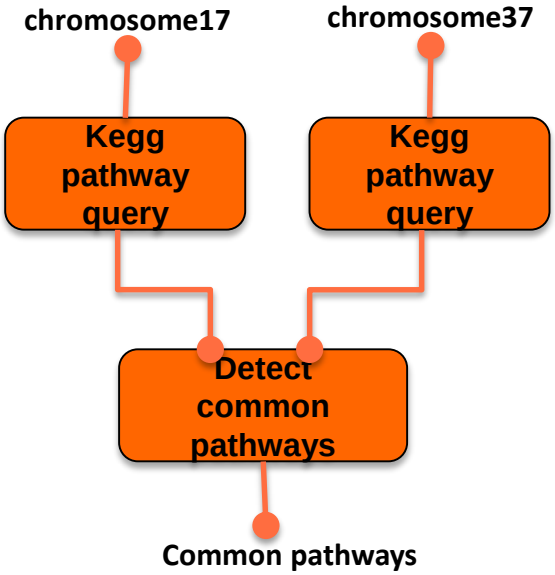
Introduction • References • Supplementary Information

In January, over 50 researchers from 30 academic and commercial organizations agreed on a standard for describing data sets. The BioSharing

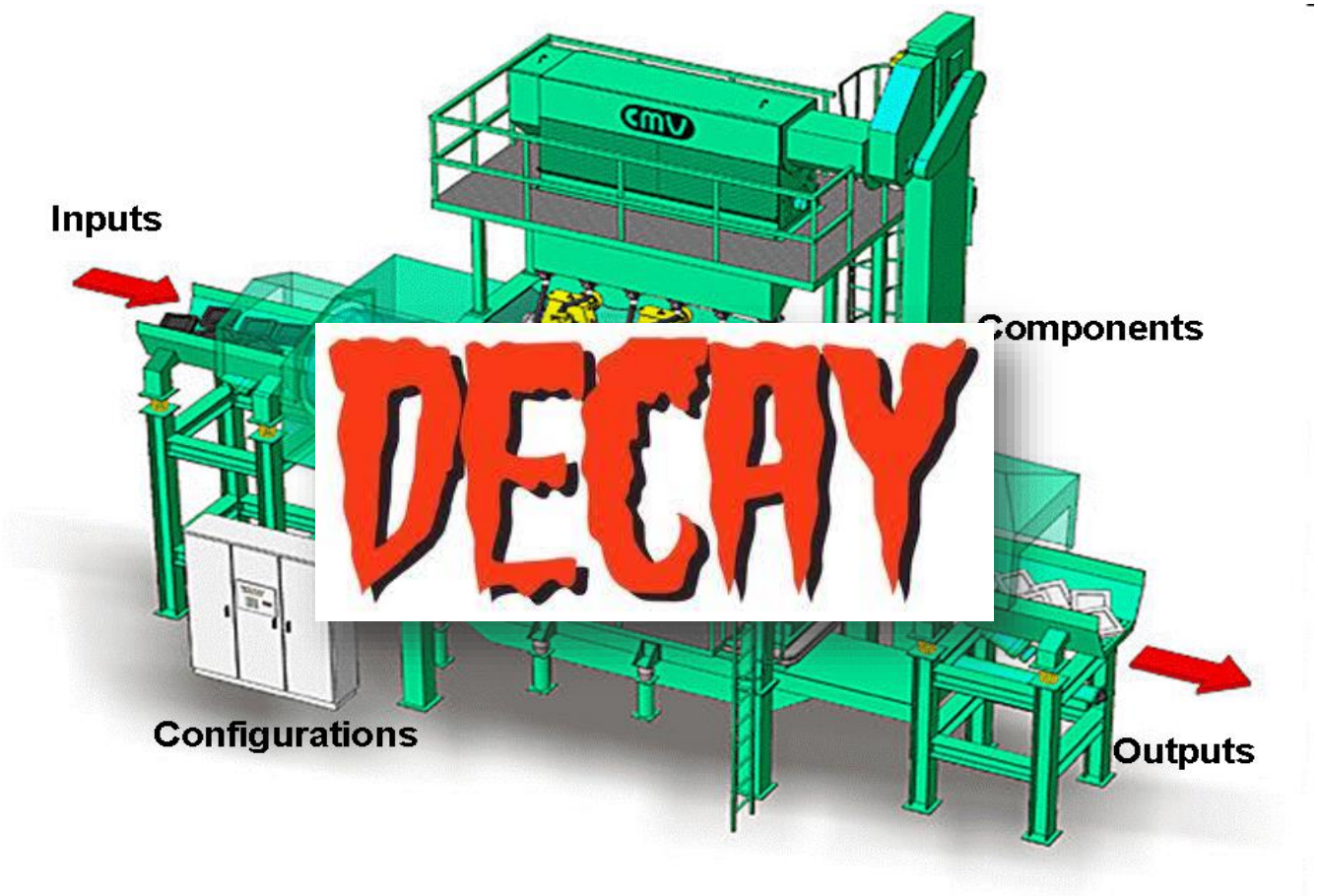
日本語要約

Workflows as means to describe, re-run and reuse scientific methods

- Explicit description of computational methods
- Coordinated execution of computational resources
- Provenance trails
- Repeatable, comparable
- Increasingly adopted in experimental science

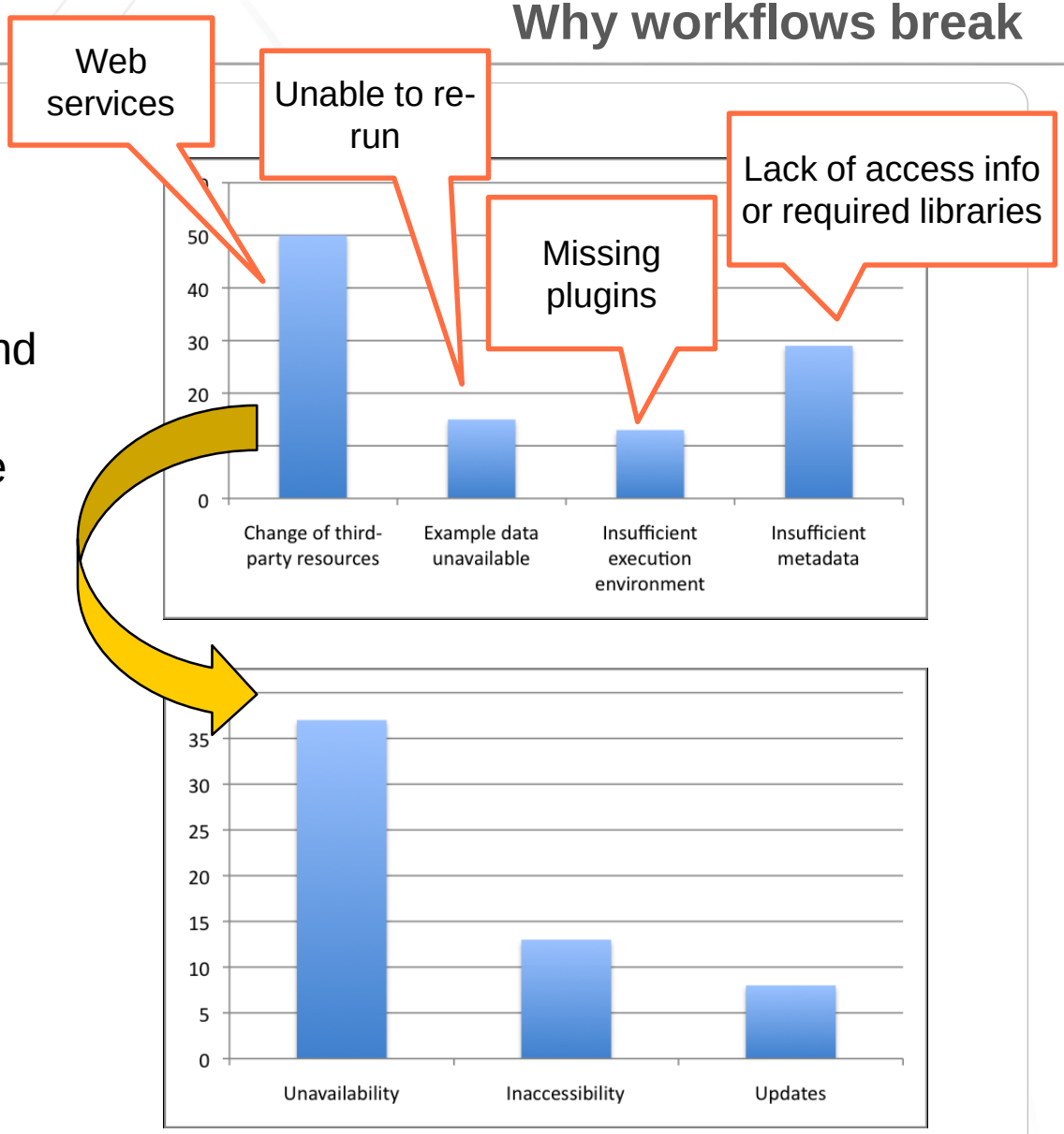


Not only the lack of transparency in the methods makes results irreproducible and not reusable. Eventually workflows just break



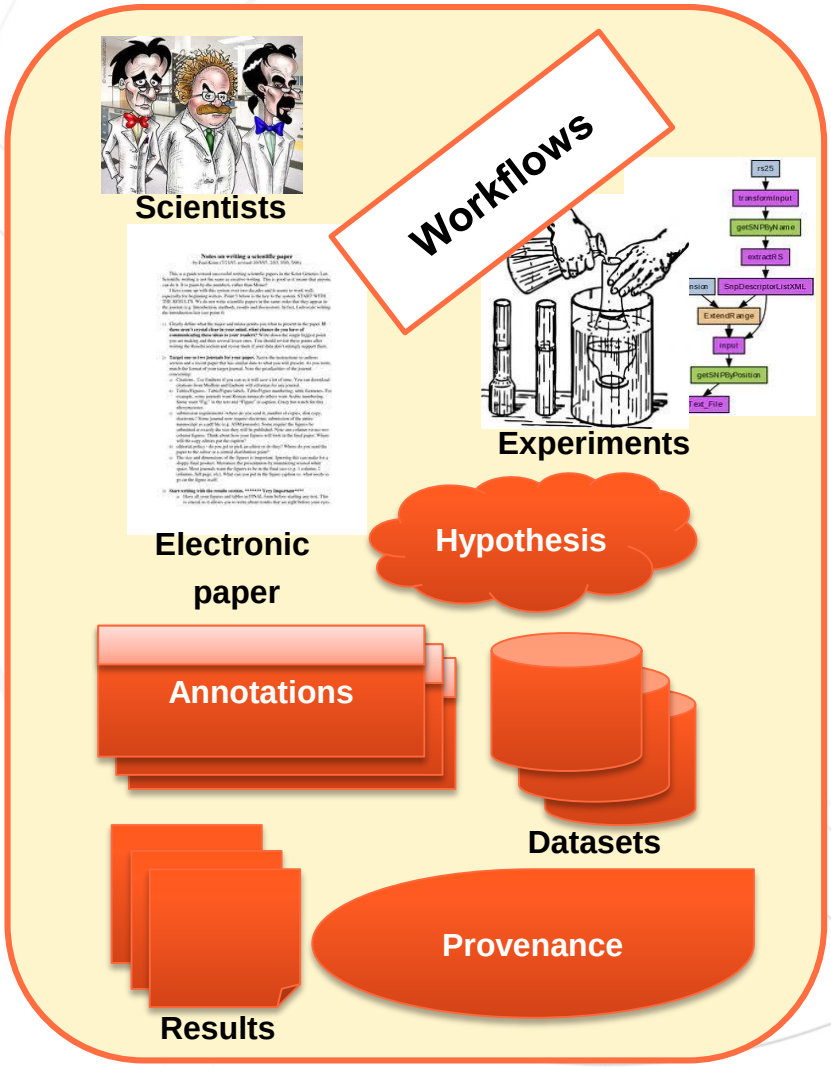
Why workflows break

- Tested ~100 Taverna workflows in myExperiment between 2007 and 2012
- Main domains included Genomics and Astronomy
- 75% failed to execute or produce the same results
- 91% in the case of older workflows
- Four main causes of decay
  - Volatile third party resources
  - Insufficient metadata
  - Sample input or output data not provided
  - Insufficient execution environment



J. Zhao, J.M. Gómez-Pérez, et al. Why Workflows Break - Understanding and Combating Decay in Taverna Workflows. In the proceedings of the IEEE eScience Conference (eScience 2012), IEEE CS, Chicago, USA, 2012.

Research Object



- Carriers of Research Context
- Uniquely identifiable and referentiable
- Metadata
  - Annotations, Manifests, Recipes, Permissions, Discourse
  - Provenance
- Lifecycle
  - Evolution, versioning
- Mixed Stewardship
  - Distributed, dispersed, local and external, 3rd party tenancy, heterogeneous
  - Decay
  - Graceful Degradation

OAI-ORE PROV-O

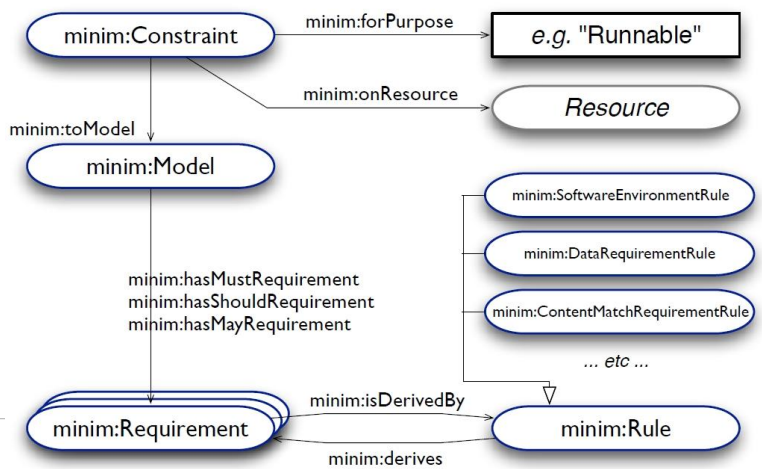
RO Primer: http://wf4ever.github.com/ro-primer
RO specification: http://wf4ever.github.com/ro

Aggregating info as ROs is not enough. Need to ensure it keeps fit for a purpose

Runnable workflow

- ☑ Execution environment present
- ☑ Workflow description present
- ☑ input files present
- ☑ parameter values defined
- ☑ required services accessible

MINIM model



- » Well-established tool for guiding practices to ensure safety, quality, and consistency in scientific communities
- » Checklists allow assessing a research object is **complete** wrt. the sufficient information for running the workflow and replicate its results
- » MINIM
- » Represents the requirements an RO must, should or may satisfy
- » Reusable checklists

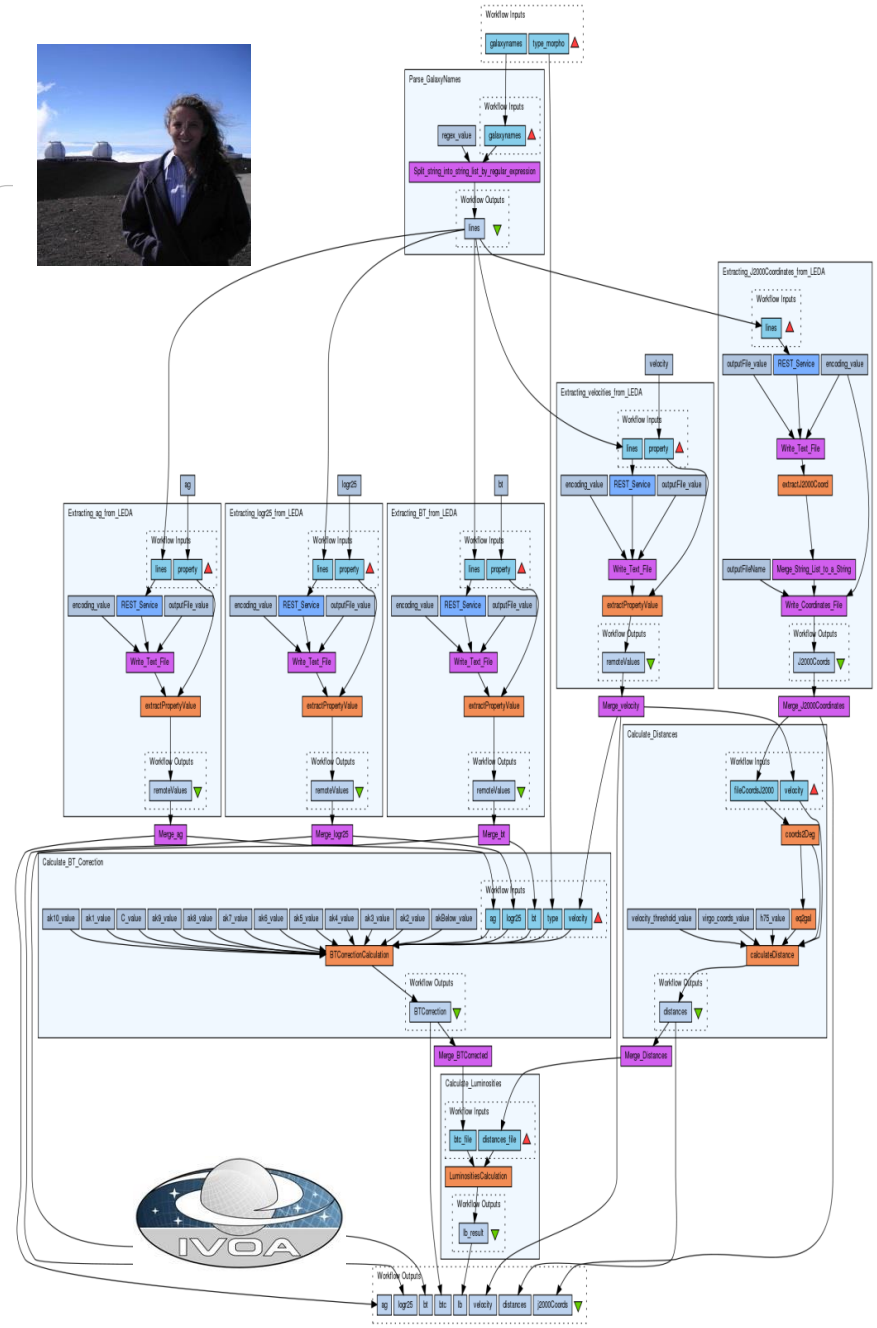
MINIM based on Gamble et al.'s MiM model

But history is important for reuse…

…to understand how you got here…

…and try to see what may be next…

Bob & Alice: A science love story?



Calculation of distances, magnitutes and luminosities using **HyperLEDA**

(http://www.myexperiment.org/workflows/2560.html)

Not reliable at all… I can't reuse this ☹

The workflow broke due to changes in the input data format when adopting VO standards for data querying. As a consequence a script using this dataset for calculating derived properties also broke

Bob's institution forgot renewing the domain and a web service was down during two days. Same with the input data, hosted in the same institution. Bob now using his own input dataset

Recent replacement of networking infrastructure (optic fiber and routers) caused connectivity glitches in the same institution

Data provider modified the output format of the responses from HTML to VOTable format to be VO compliant and achieve interoperability. This caused one of the scripts to fail

- **Completeness**

  How many of the requirements in the checklist are satisfied by the workflow at a given point in time?

  To what extent?

- **Stability**

  Are changes frequent over the life of the workflow?

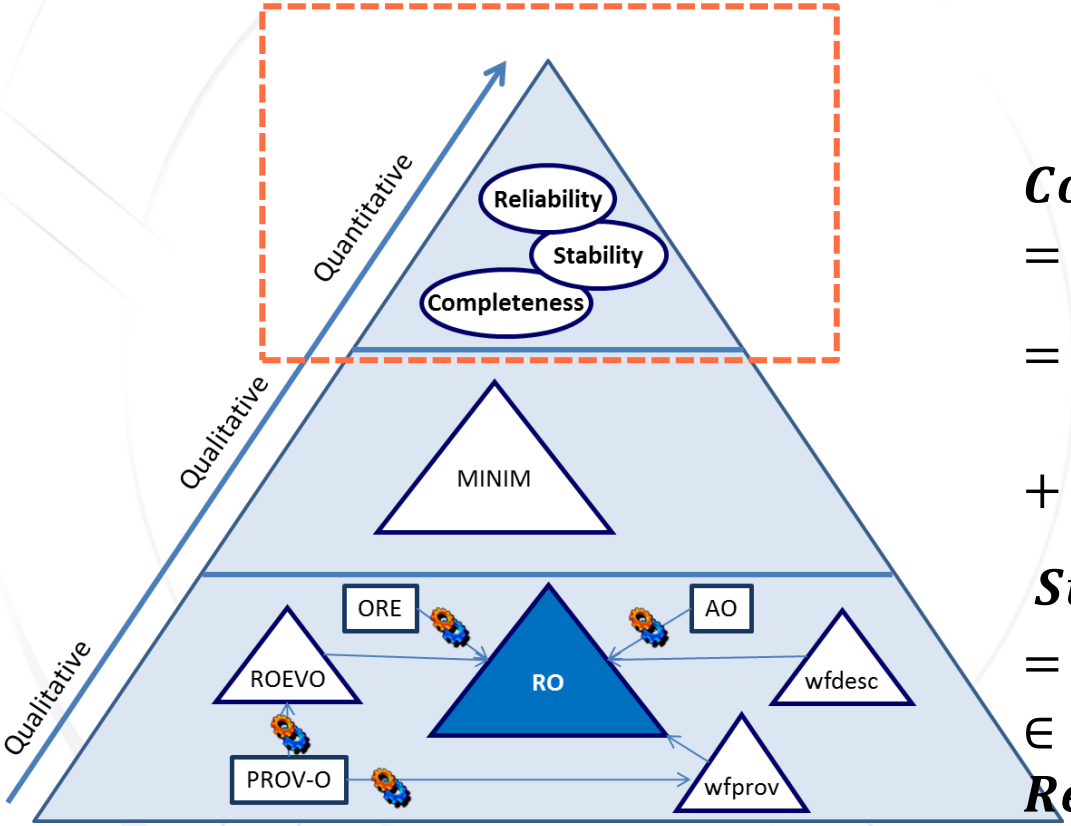
- **Reliability**

  Were those changes positive or negative?

  Does the workflow preserve its properties over time?

  Is it safe to reuse the workflow?

Quantifying reliability

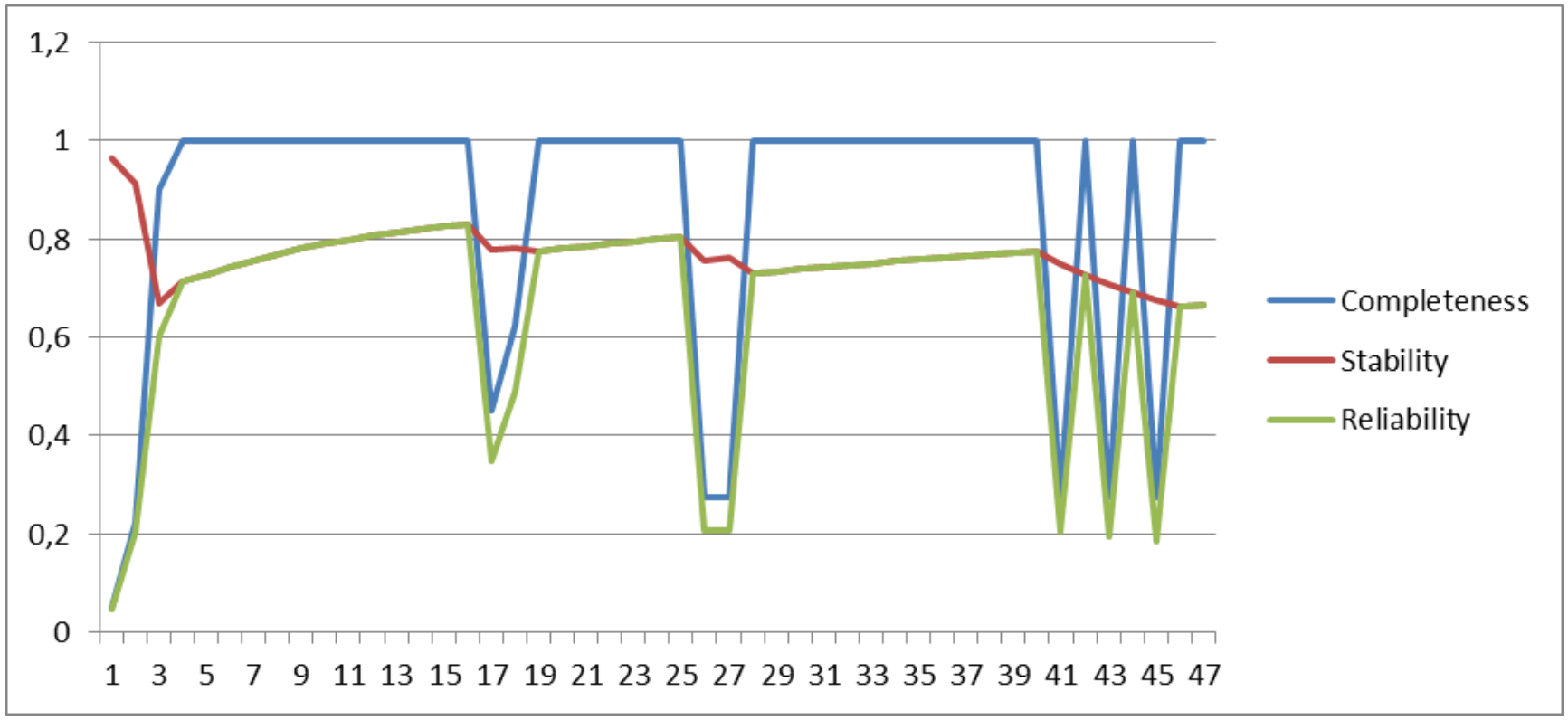


$$\boldsymbol{Completeness_{score(RO,t)}} = f(RO(t), requirements, type) = \alpha \frac{nSReq(RO(t), must)}{nReq(RO(t), must)} + (1-\alpha)\frac{nSReq(RO(t), should)}{nReq(RO(t), should)} \in [0,1]$$

$$\boldsymbol{Stability_{score(RO,t)}} = 1 - std\left(Completeness_{score}(RO, \Delta t)\right) \in [0.5,1]$$

$$\boldsymbol{Reliability_{score(RO,t)}} = Completeness_{score(RO,t)} \times Stability_{score(RO,t)}$$

Reliability over time (completeness x stabiity)

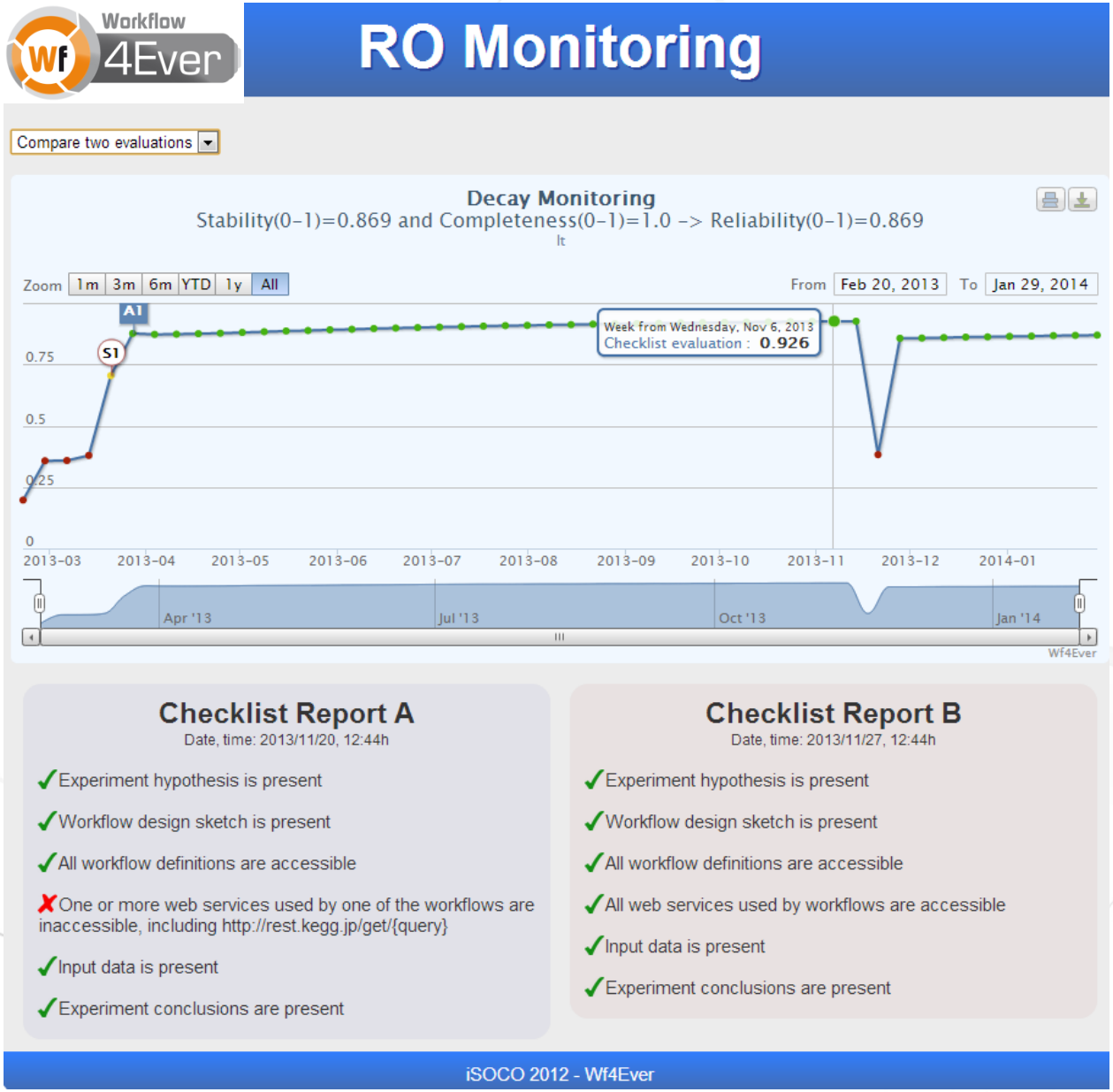


- Stability introduces an upper bound for reliability
- Reliability converges to 1 when an RO is stable and complete
- Empirically useful for assessing potential reuse

The RO monitoring tool

http://sandbox.wf4ever-project.org/portal

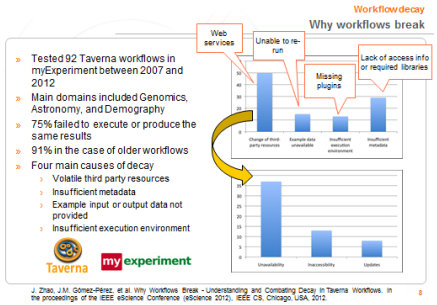
http://sandbox.wf4ever-project.org/rodl/ROs/Pack387

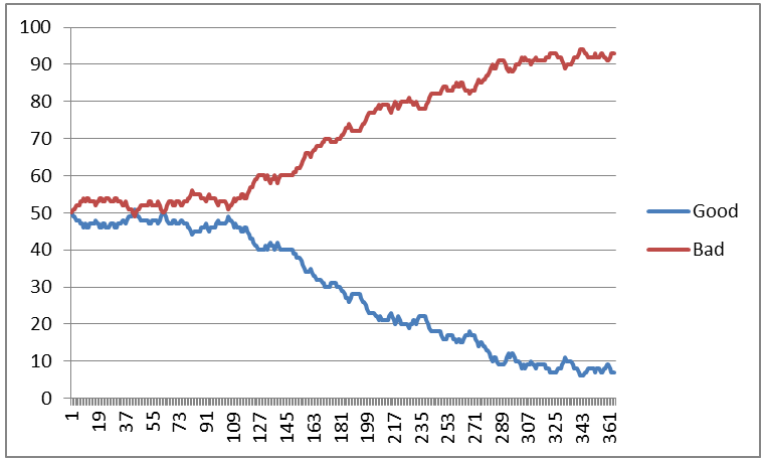


Evaluation settings

A notion of the history and evolution of workflow decay allows scientists for more effective and accurate decisions on workflow reuse

- Based on previous analysis of workflow decay
  - The probability of decay increases with time
  - 1 yr old: 50%, 5 yr. Old: 91%
- Model simulates the evolution of 100 worfklows for one year
- Two main groups of workflows
  - G1: Can be executed and well maintained
  - G3: Cannot be executed and unlikely to be fixed
  - Plus, G2: Currently run but not maintained
- In time, workflows move bw. the groups
- Initial vs. final distribution of workflow population
  - G1: (40, 7), G2: (20, 0), G3: (40, 93)





http://www.iaa.es

Evaluation team comprised of 9 astronomers from IAA-CSIC

User input

Q1: Would you reuse this workflow for your own experiments today?

Q2: Would you use it in three months from now?

- Two questions formulated on day 274
- Answers first based only on the completeness score and then using the reliability score and the RO monitoring tool
- Focus on the scores. No inspection of the workflows
- Reliability information improved 72% of Q1 answers and 76% in the case of Q2
- More realistic expectations on workflow reuse (32% vs. 38%)

Take home message

- In general, experimental resources are highly volatile
  - Mixed stewardship
  - Decay! Usually from 3rd party resources
- Research Objects
  - A little semantics goes a long way, once more
- Effective reuse requires looking at the present but also at the past
- Quality is important for reuse. So is the capability to find relevant stuff
- From the lab to the Web

Workflow 4Ever

http://www.wf4ever-project.org

Home | Publications | Deliverables | Press Room | Wiki | Contact us | Administrative Info

Wf4Ever addresses some of the challenges associated to the preservation of scientific experiments in data-intensive science, including:

- The definition of models to describe, in a standard way, scientific experiments by means of workflow-centric Research Objects, which comprise scientific workflows, the provenance of their executions, interconnections between workflows and related resources (e.g., datasets, publications, etc.), and social aspects related to such scientific experiments.
- The collection of best practices for the creation and management of Research Objects.
- The analysis and management of decay in scientific workflows.

To address these challenges, we are creating an architecture and tooling for the access, manipulation, sharing, reuse and evolution of Research Objects in a range of disciplines. This will result into the next generation RO-enabled myExperiment.

>>

Wf4Ever for Scientists

Wf4Ever for Developers

Wf4Ever for Citizens (Coming soon)

Should you wish to cite Wf4Ever, we suggest you to use the following reference: Belhajjame K, Corcho O, Garijo D, Zhao J, Missier P, Newman DR, Palma R, Bechhofer S, Garcia-Cuesta E, Gómez-Pérez JM, Klyne G, Page K, Roos M, Ruiz JE, Soiland-Reyes S, Verdes-Montenegro L, De Roure D, Goble CA: Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. In proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica2012), Heraklion, Greece, May 2012

Wf4Ever Blog

Wf4Ever in Beyond PDF2
3/4/13

ROs appear as a use case in the LDP WG
3/4/13

Prof. Goble gave invited talk at the Scholarly Communication Workshop, Pittsburgh, 14-15 January 2013
1/17/13

10 Best Practices for Workflow Design at SWAT4LS
11/21/12

Prof Goble's Keynote at JCL2012
6/18/12

Wf4Ever@ISWC2012
6/2/12

Taverna looks at the sky
5/18/12

Wf4Ever@IIPC General Assembly
5/16/12

Querying Provenance of Workflow Results

Dr. José Manuel Gómez-Pérez
Director R&D
jmgomez@isoco.com
T +34 913 349 797
http://www.isoco.com

Thank you for your attention!

Barcelona
Tel +34 935 677 200
Edificio Testa A
C/ Alcalde Barnils, 64-68
St. Cugat del Vallès
08174 Barcelona

Madrid
Tel +34 913 349 797
Av. del Partenón, 16-18, 1º7ª
Campo de las Naciones
28042 Madrid

Pamplona
Tel +34 948 102 408
Parque Tomás
Caballero, 2, 6º4ª
31006 Pamplona

Valencia
Tel +34 963 467 143
Oficina 107
C/ Prof. Beltrán Báguena, 4
46009 Valencia

http://www.wf4ever-project.org

iSOCO
enabling the networked economy

Workflow 4Ever