GB-500: Introdução a Workflows Científicos e suas Aplicações

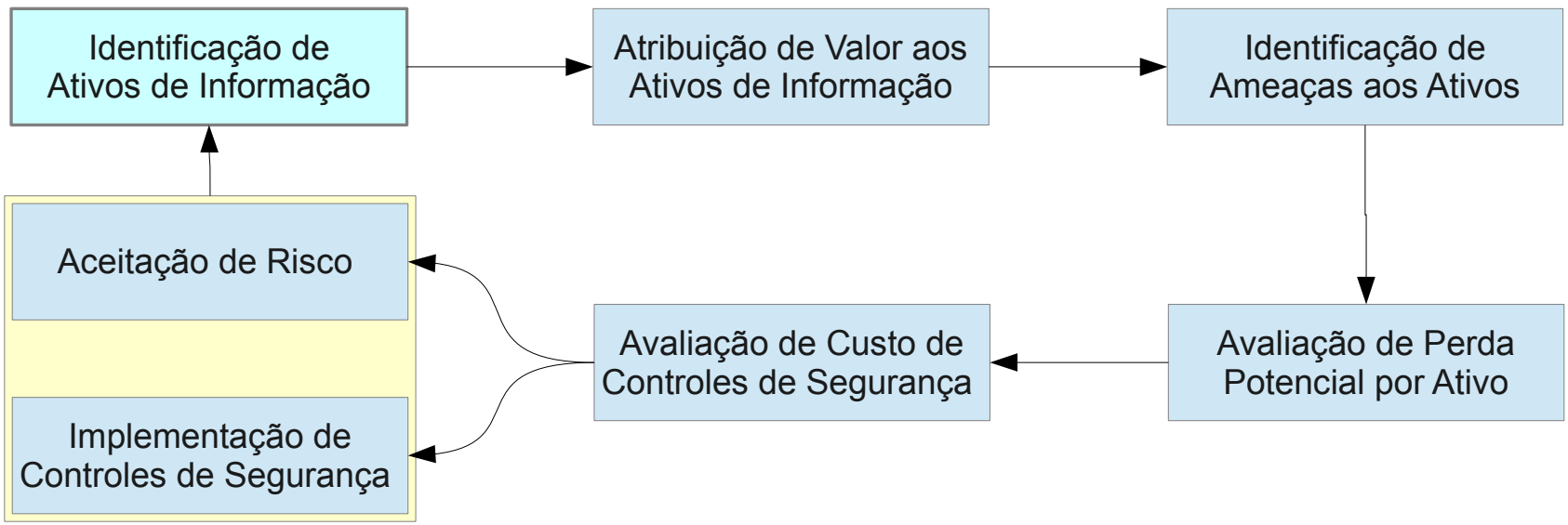
Professores: Luiz Gadelha e Kary Ocaña

Programa de Pós-Graduação em Modelagem Computacional, P3/2015
Laboratório Nacional de Computação Científica

18 de junho de 2015

Laboratório
Nacional de
Computação
Científica

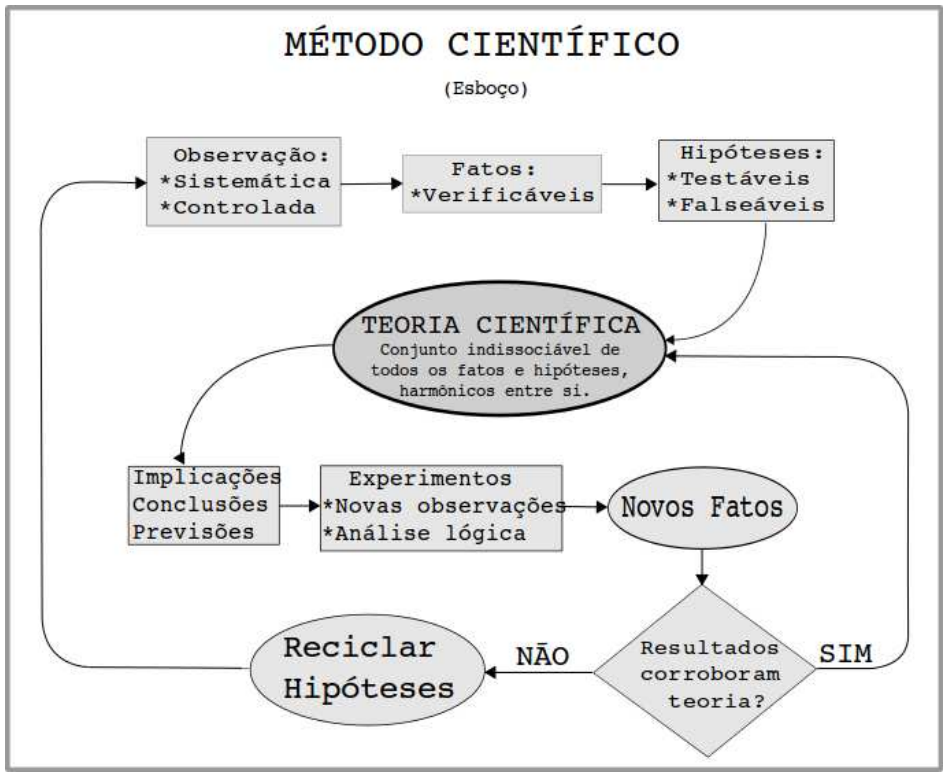
Análise de Risco em Segurança da Informação



Segurança em Ciência, Tecnologia e Inovação

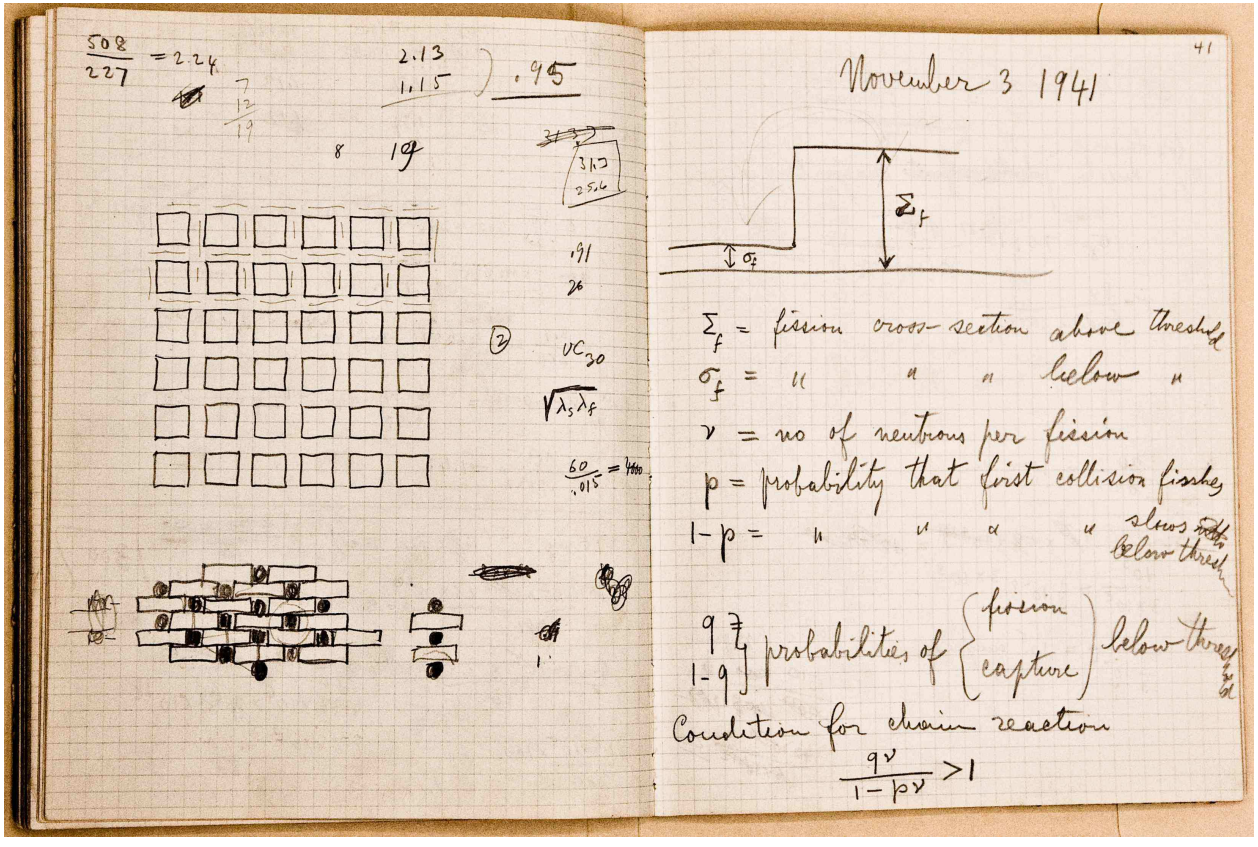
- Quais seriam os principais ativos de informação em CT&I?
- Quais são as principais ameaças a estes ativos?
- Qual é a perda potencial por ameaça para cada ativo?
- Quais são os custos dos respectivos controles de segurança?

Segurança em CT&I: Método Científico



Fonte: Wikimedia Commons.

Segurança em CT&I: Ativos



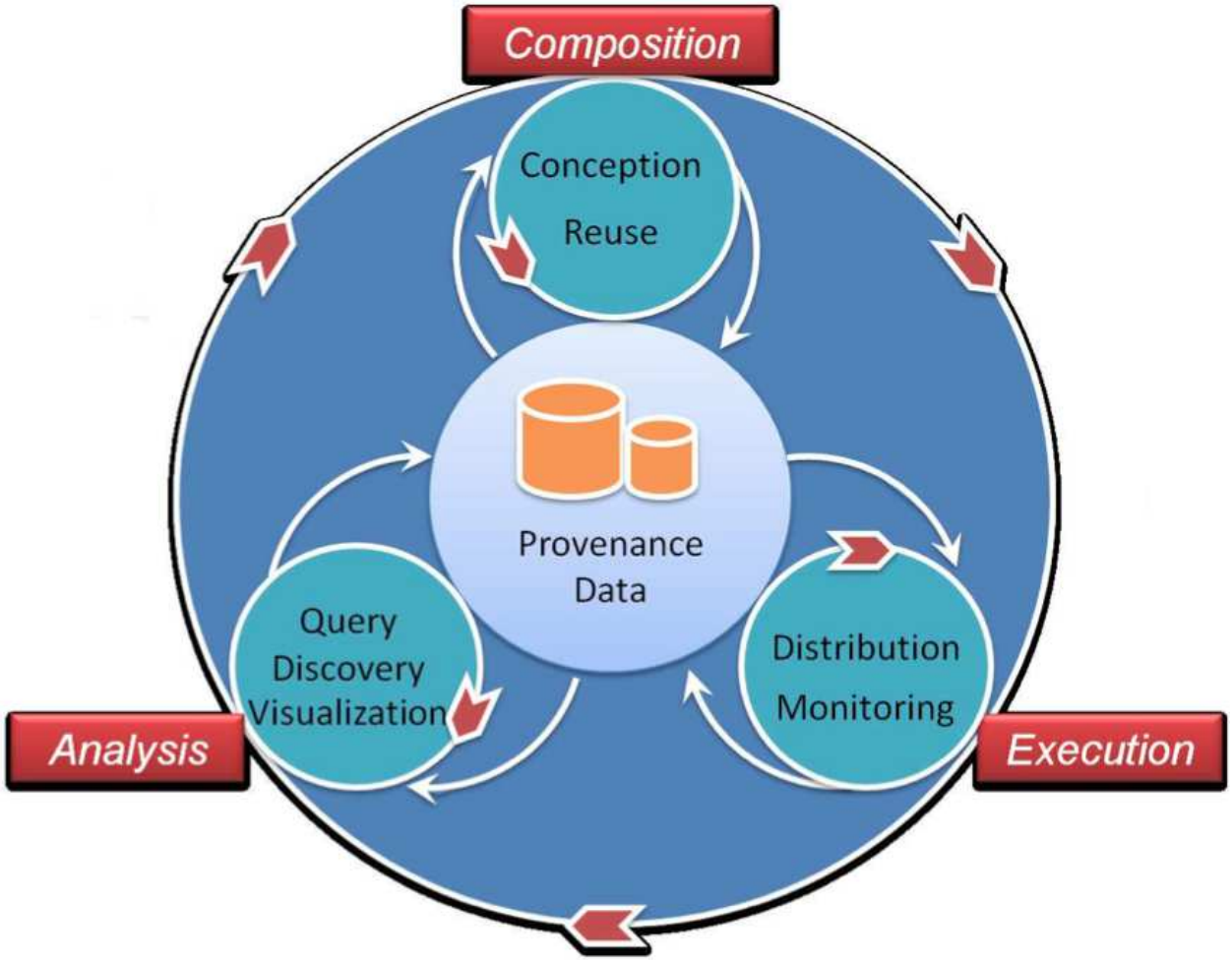
Caderno de laboratório de Enrico Fermi com anotações sobre a primeira reação nuclear em cadeia.
Fonte: The University of Chicago Library.

Proveniência

- Informações de **proveniência** descrevem o histórico de concepção e execução de um experimento.
- Os registros de proveniência são análogos aos *cadernos de laboratório* dos experimentos de bancada.
  - plano do experimento;
  - parâmetros iniciais;
  - descrição dos resultados.

S. B. Davidson and J. Freire, Provenance and scientific workflows: challenges and opportunities. *Proc. of the International Conference on Management of Data* (SIGMOD 2008), pp. 1345–1350. ACM, 2008.

Computational Scientific Experiment Life Cycle



M. Mattoso, C. Werner, G. Travassos, V. Braganholo, E. Ogasawara, D. Oliveira, S. Cruz, W. Martinho, and L. Murta, Towards supporting the life cycle of large scale scientific experiments. *International Journal of Business Process Integration and Management* 5(1):79–92, 2010.

Threat evaluation in e-Science

- We propose that the main information asset of these systems is given by provenance traces describing the intellectual process of a computational scientific experiment.
- This information is particularly vulnerable in current e-Science infrastructures since they often are transferred to third-party computational resources which scientists have little control of.

Threat evaluation in e-Science

The Eavesdropper. Eugene de Blaas, 1906. (Source: Wikimedia Commons)

Threat evaluation in e-Science

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

News › World news › NSA

NSA accused of spying on Brazilian oil company Petrobras

Accusations that NSA is conducting intelligence-gathering operations that go beyond its core mission of national security

Jonathan Watts in Rio de Janeiro
theguardian.com, Monday 9 September 2013 16.55 BST
Jump to comments (197)

Petrobras is the largest company in Brazil – majority owned by the state – and one of the 30 biggest businesses in the world. Photograph: Bloomberg/ Getty

The New York Times

Americas

N.S.A. Spied on Brazilian Oil Company, Report Says

By SIMON ROMERO
Published: September 9, 2013

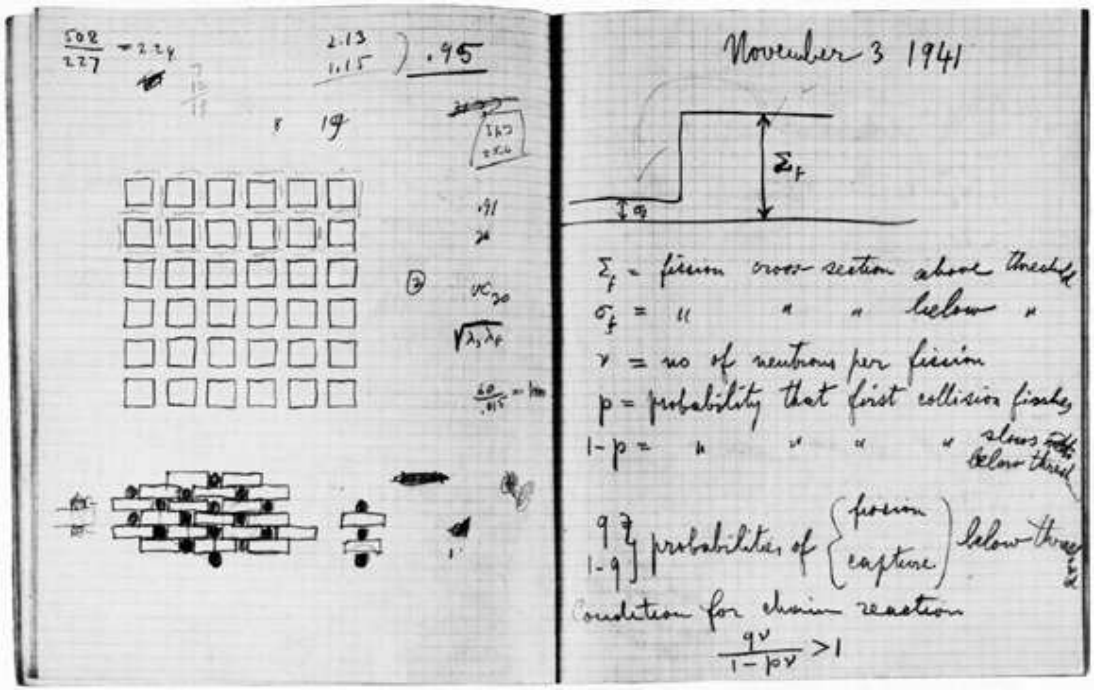
RIO DE JANEIRO — The National Security Agency spied on Petrobras, Brazil's giant national oil company, according to a report here on Sunday night by the Globo television network, in the latest revelation of the agency's surveillance methods that have raised tension between Brazil and the United States.

FACEBOOK
TWITTER
GOOGLE+
SAVE

Threat evaluation in e-Science

{S1} Illegitimate claim of attribution.
{S2} Unauthorized access to private data.
{S3} Intentional modification of provenance records.
{S4} Dissemination of illegitimate provenance data.
{S5} Obstruction of provenance information collection and access.

Provenance records vs. lab notebooks



Source: Enrico Fermi's Laboratory Notebook, The University of Chicago Library.

- Provenance records are analogous to lab notebooks:
  - plan of the experiment;
  - initial parameters;
  - description of results.

Provenance records vs. lab notebooks

- Guidelines for maintaining lab notebooks:
  - "Shows completion before the dates of other prior art references;"
  - "Provides evidence for proving inventorship or first-to-invent."
  - "Use indelible ink for entries."
  - "All details of an experiment should be listed, signed, dated, and witnessed. This includes data and final results of experiments, protocols and design of experiments, calculations on which the results are based, ..."

Maintaining Laboratory Notebooks, University of Minnesota, 2014.

Problemas de Segurança em Computação Distribuída

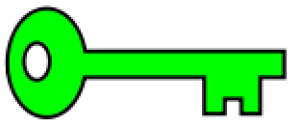
- Questões:
  - Como gerenciar identidades através de domínios administrativos?
  - A *Grid Security Infrastructure* (GSI) utiliza técnicas de criptografia e certificação digital para gerenciar identidades através de domínios administrativos com segurança.

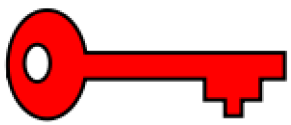
Criptografia e Certificados Digitais

- Propriedades desejáveis para a realização de transações eletrônicas de forma segura:
  - Confidencialidade: sigilo para o conteúdo;
  - Integridade: conteúdo não pode ser alterado;
  - Autenticação: garantir a identidade dos participantes;
  - Não-repúdio: os participantes não podem negar o conteúdo.

Criptografia e Certificados Digitais

- Na criptografia de chave assimétrica uma entidade possui duas chaves criptográficas:

Chave pública: 

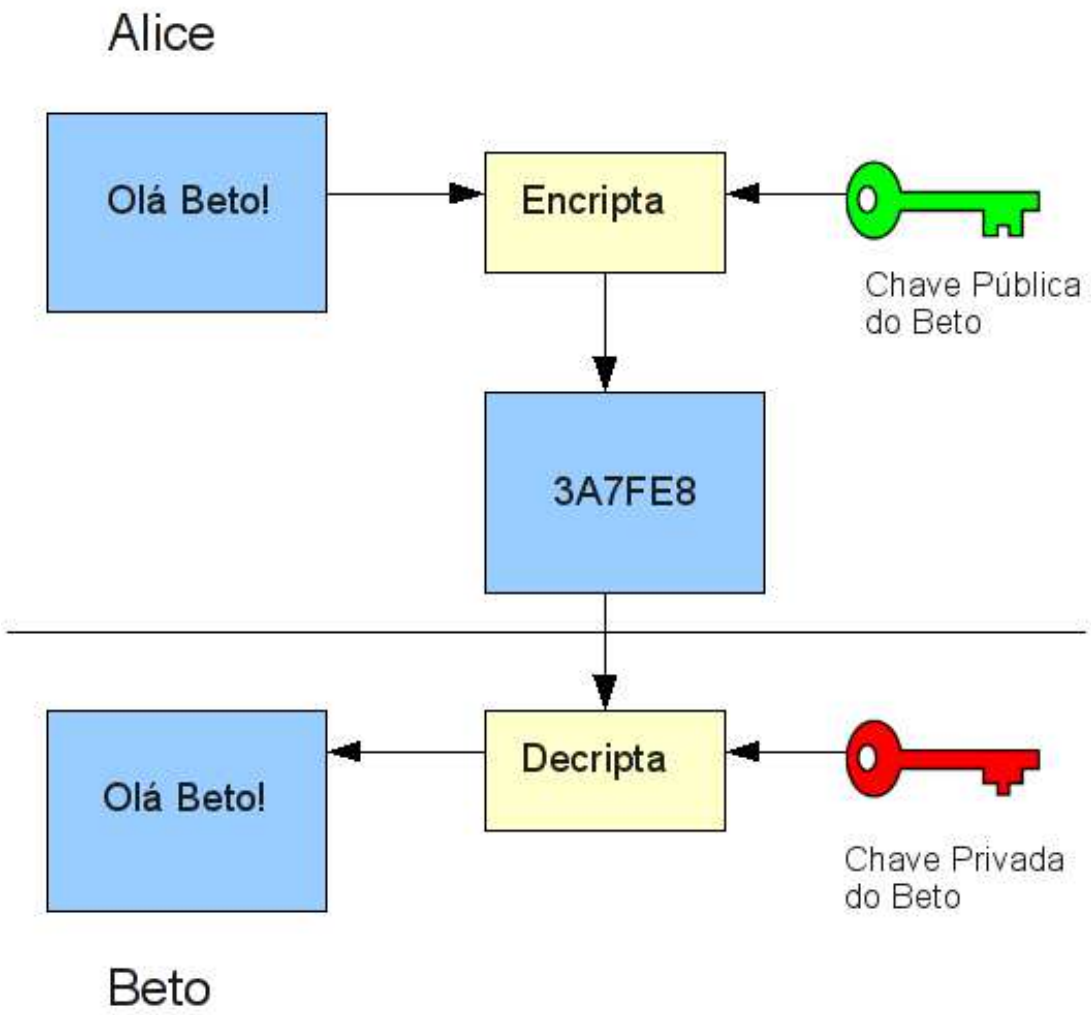
Chave privada: 

- As chaves são mutuamente inversas: o que uma encripta a outra decripta.
- A chave privada deve ser mantida em sigilo.
- A chave pública pode ser divulgada.

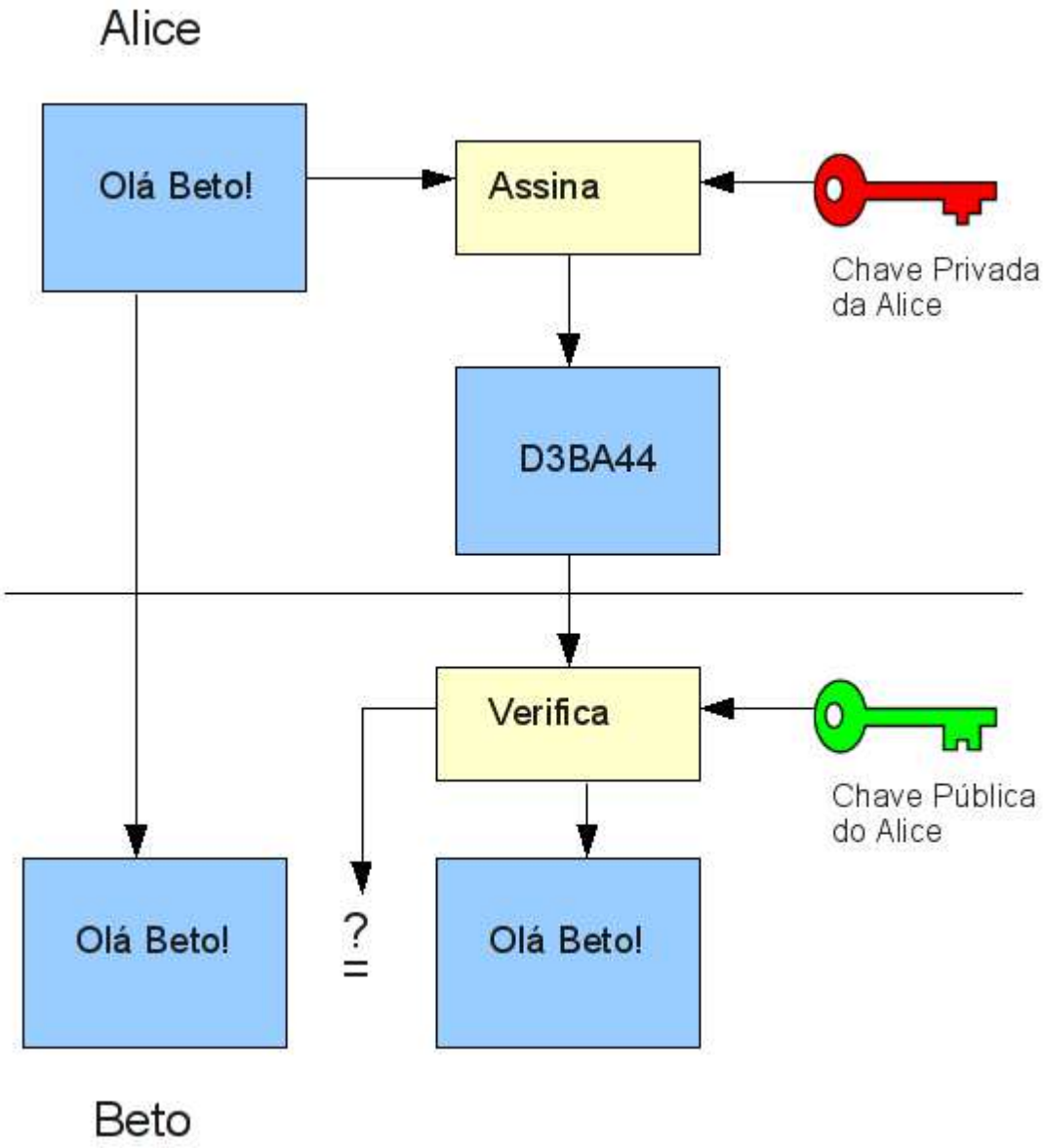
Criptografia e Certificados Digitais

- Aplicações de criptografia assimétrica:
  - Comunicação confidencial: remetente usa a chave pública do destinatário para criptografar a mensagem.
  - Assinatura digital: remetente usa sua chave privada para criptografar a mensagem.

Encriptação



Assinatura Digital



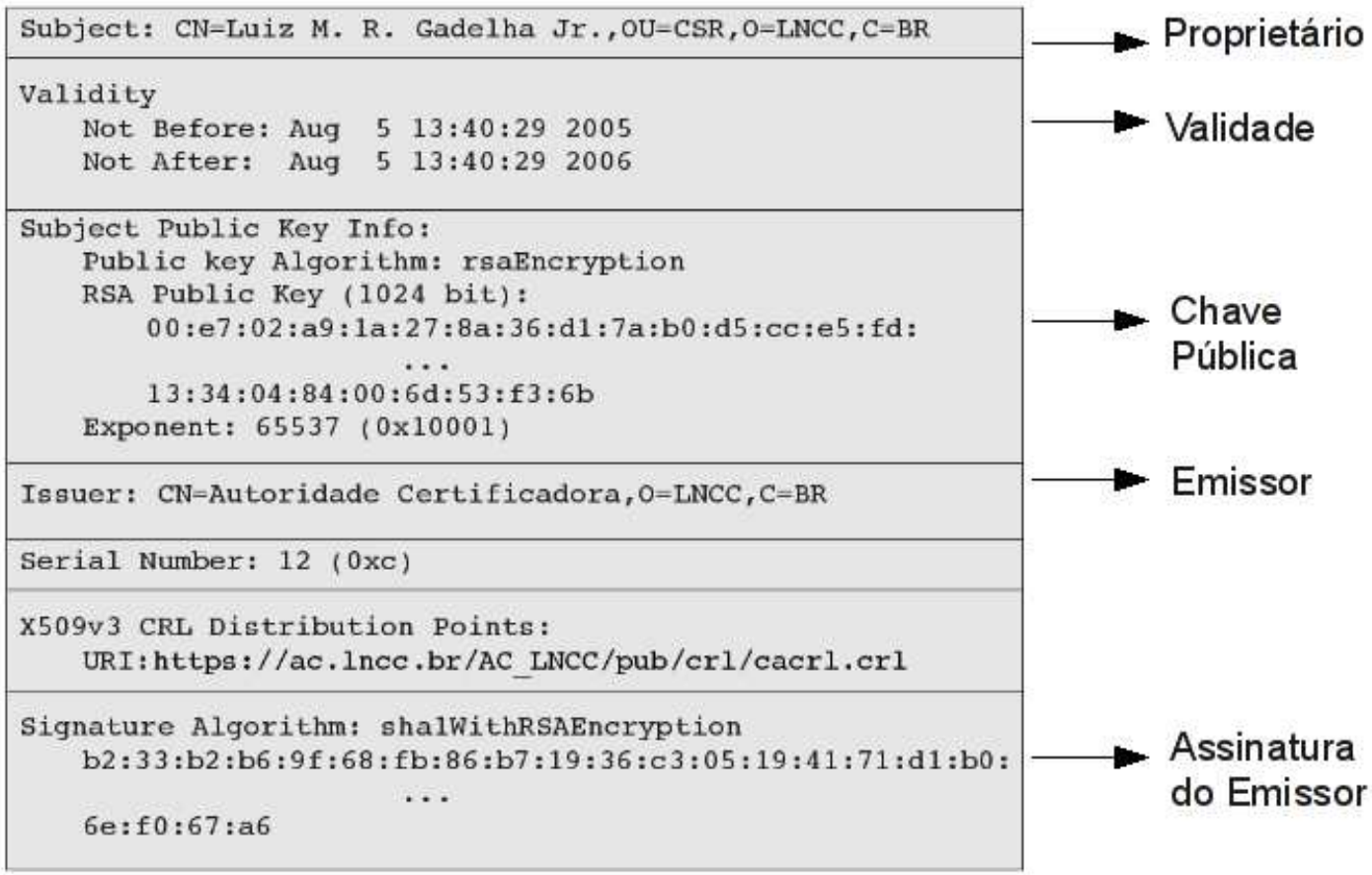
Criptografia e Certificados Digitais

- Problema: quem é o “dono” de um par de chaves?
- Uma solução é utilizar um terceiro de confiança, chamado Autoridade Certificadora (AC), tal que ele:
  - possua uma chave pública conhecida, na qual outras entidades possam confiar;
  - certifique a identidade de outras entidades e a relação de posse entre estas entidades e suas respectivas chaves criptográficas.

Criptografia e Certificados Digitais

- Para tal, a AC emite certificados digitais.
- Um certificado digital é um documento eletrônico que identifica uma entidade e a sua chave pública.
- Principais campos de um certificado digital:
  - nome da entidade (Subject);
  - chave pública da entidade (Subject Public Key Info).
  - nome da AC emissora (Issuer).
  - assinatura digital da AC emissora (Signature Algorithm).

Formato de um Certificado Digital



Criptografia e Certificados Digitais

- A AC age como terceiro de confiança.
- A própria AC possui um certificado digital.
- Esse certificado é auto-assinado ou é assinado por outra AC.

Certificado Digital de AC

```
Subject: CN=Autoridade Certificadora,O=LNCC,C=BR
```

```
Validity
    Not Before: Apr 27 12:53:36 2005 GMT
    Not After:  Apr 22 12:53:36 2025 GMT
```

```
Subject Public Key Info:
    Public key Algorithm: rsaEncryption
    RSA Public Key (2048 bit):
        00:b8:13:f4:a8:1e:33:61:60:63:1d:fa:aa:4c:d0:
                            ...
        de:af
    Exponent: 65537 (0x10001)
```

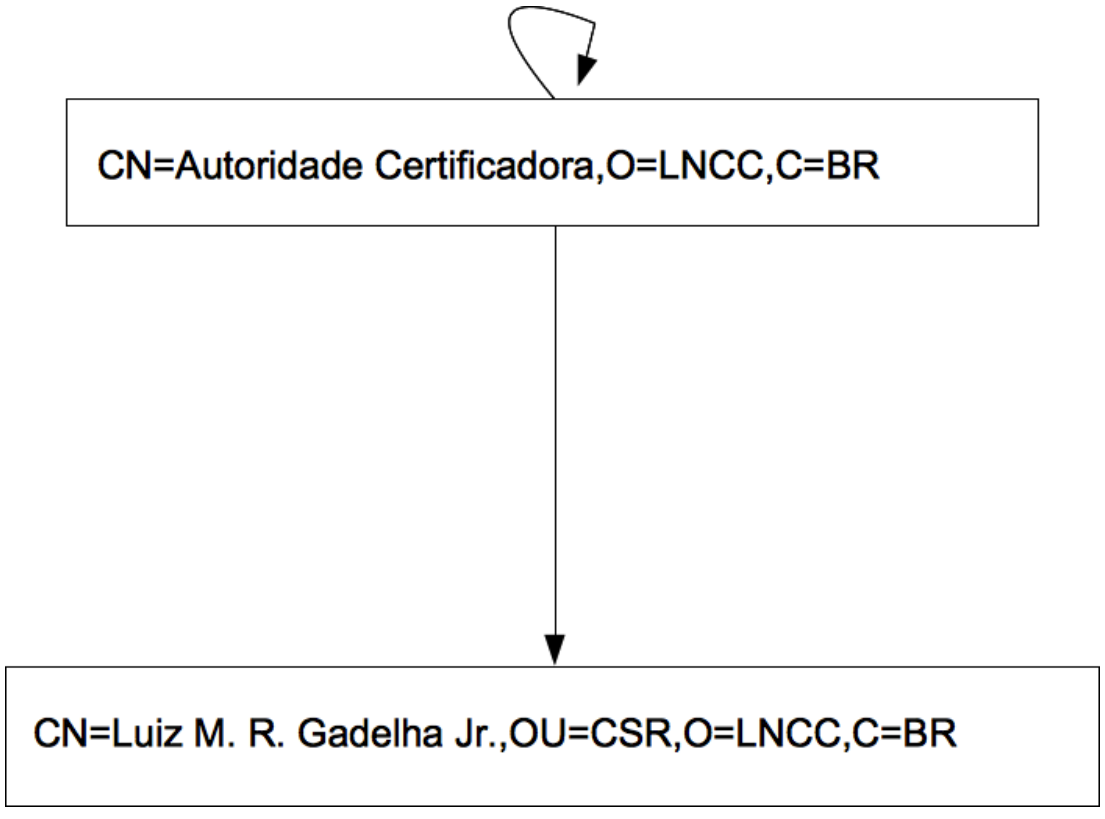
```
Issuer: CN=Autoridade Certificadora,O=LNCC,C=BR
```

```
Serial Number: bf:b8:fa:de:b3:59:78:f4
```

```
X509v3 CRL Distribution Points:
    URI:https://ac.lncc.br/AC_LNCC/pub/crl/cacrl.crl
```

```
Signature Algorithm: sha1WithRSAEncryption
    98:e9:ce:fc:6c:04:99:7c:b4:79:3a:4b:79:c1:c9:ab:bc:ba:
                            ...
    20:3d:b9:55
```

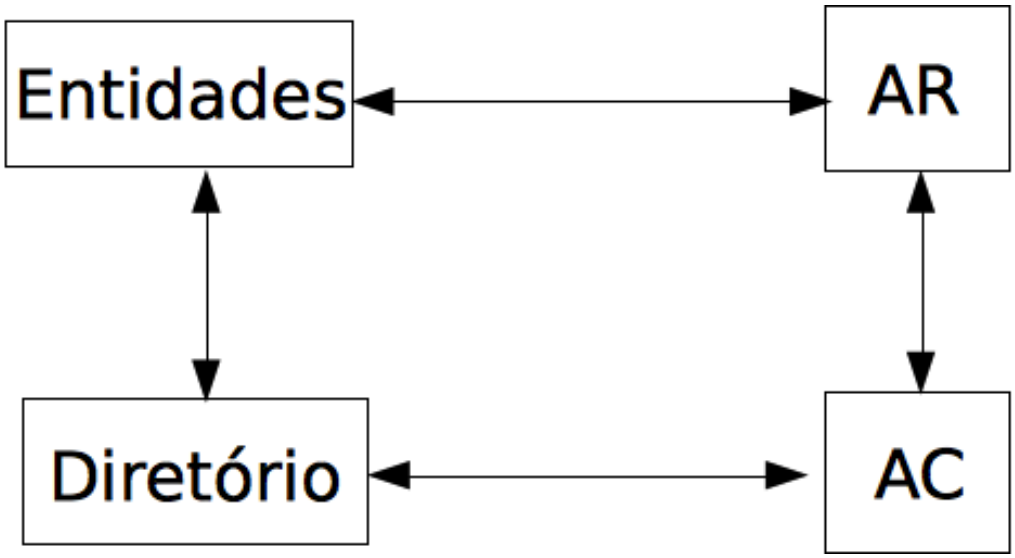
Formato de um Certificado Digital



- Uma autoridade de registro (AR) processa e faz a validação das solicitações de certificados.
- Gerencia a interação entre a entidade final e a AC.
- Diversos níveis de validação:
  - existência de caixa postal de e-mail;
  - validação presencial mediante apresentação de documentos.

- Um sistema de certificação digital deve manter um diretório (em LDAP ou HTTP p.ex.) para publicação de:
  - certificados emitidos;
  - listas de certificados revogados;
  - políticas e práticas de certificação.

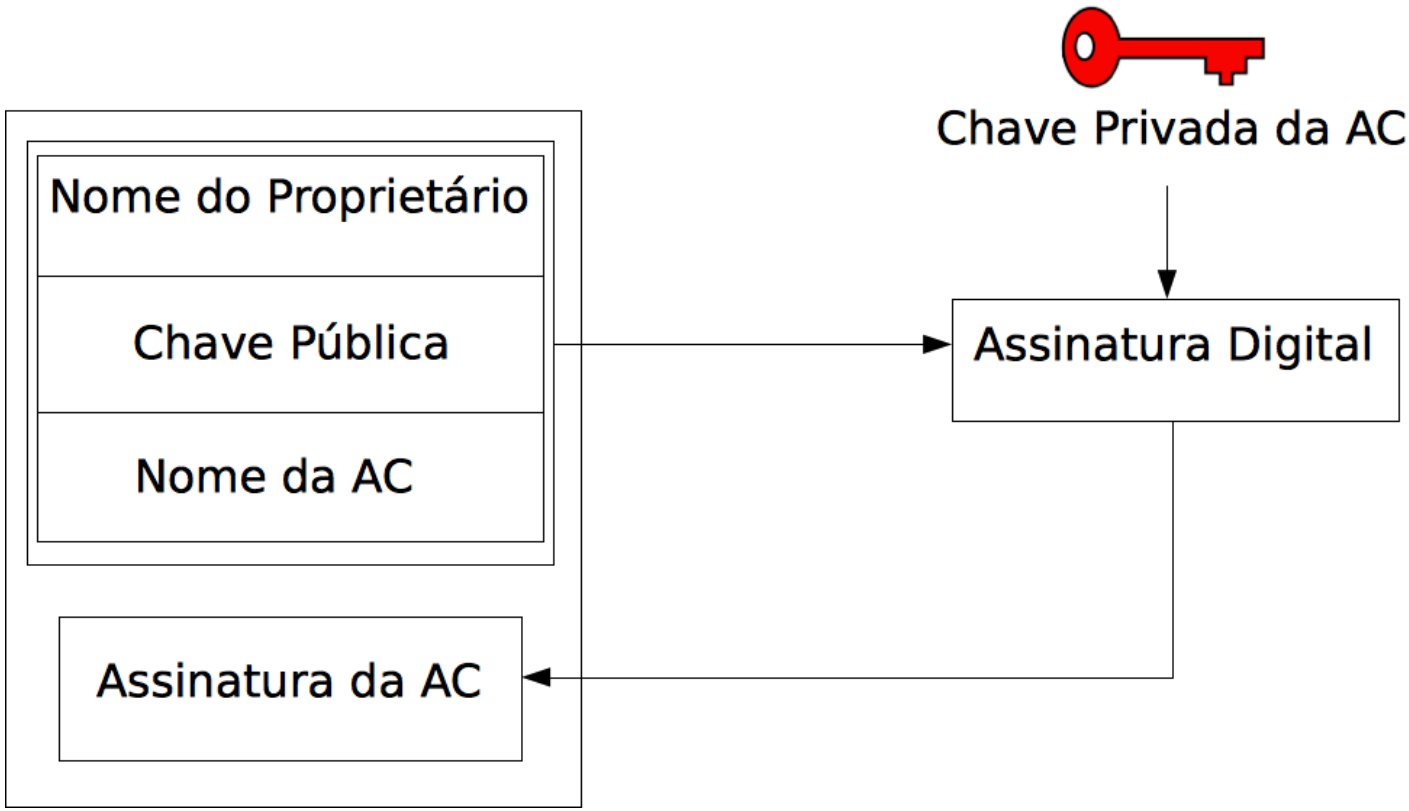
Componentes de uma ICP



- Um sistema de certificação digital deve manter um diretório (em LDAP ou HTTP p.ex.) para publicação de:
  - certificados emitidos;
  - listas de certificados revogados;
  - políticas e práticas de certificação.

- Emissão de um certificado:
  1. A entidade final, E, gera seu par de chaves criptográficas.
  2. E inclui sua chave pública em uma requisição de certificado, e o envia para a AR;
  3. A AR realiza o processo de validação de E, se a validação for positiva ela assina a requisição de certificado e a repassa para a AC;
  4. A AC assina um certificado com as informações contidas na requisição.

Componentes de uma ICP



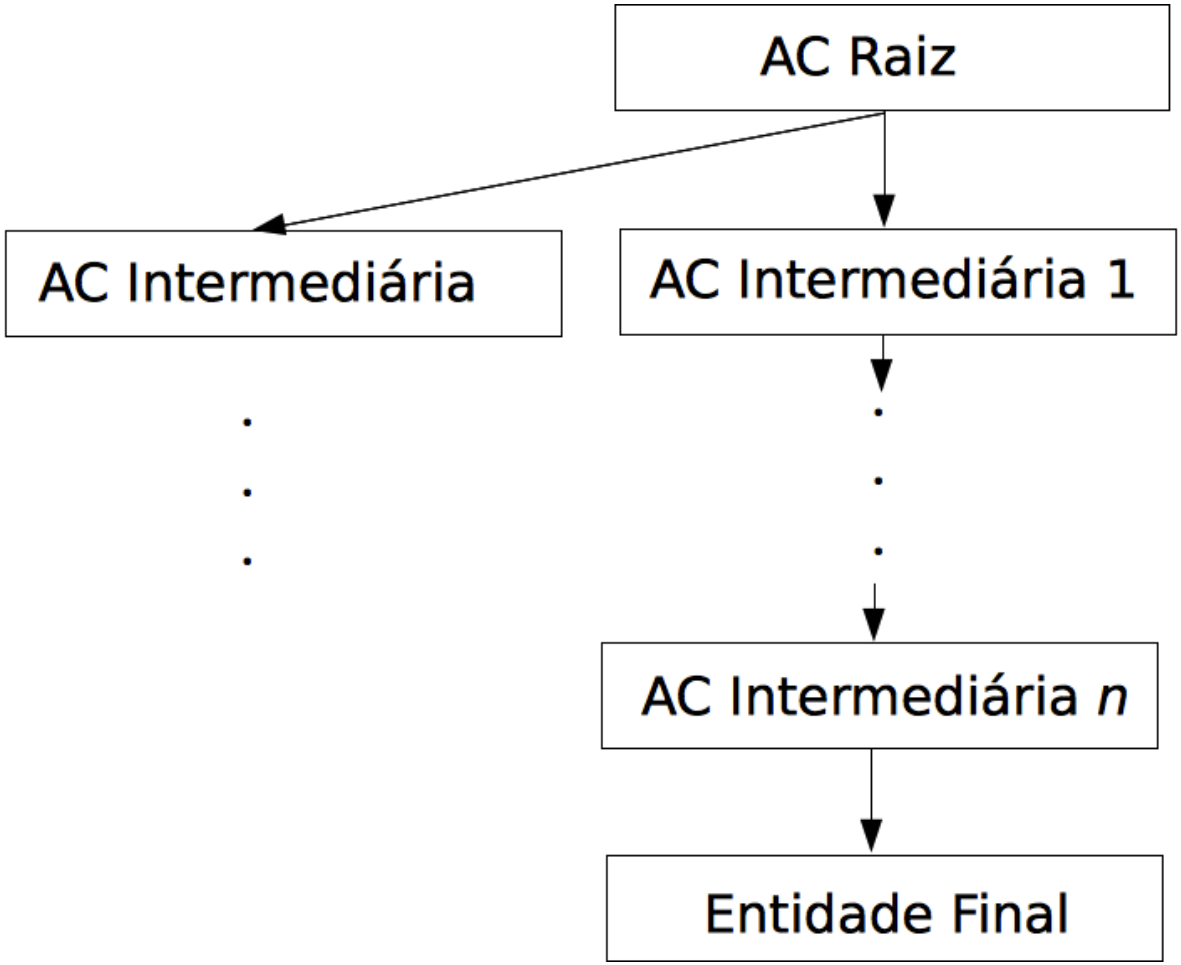
Criptografia e Certificados Digitais

- Eventualmente um certificado digital precisará ser revogado:
  - perda da chave privada;
  - comprometimento da chave privada;
  - mudança nas informações do certificado;
  - desligamento institucional de usuário.

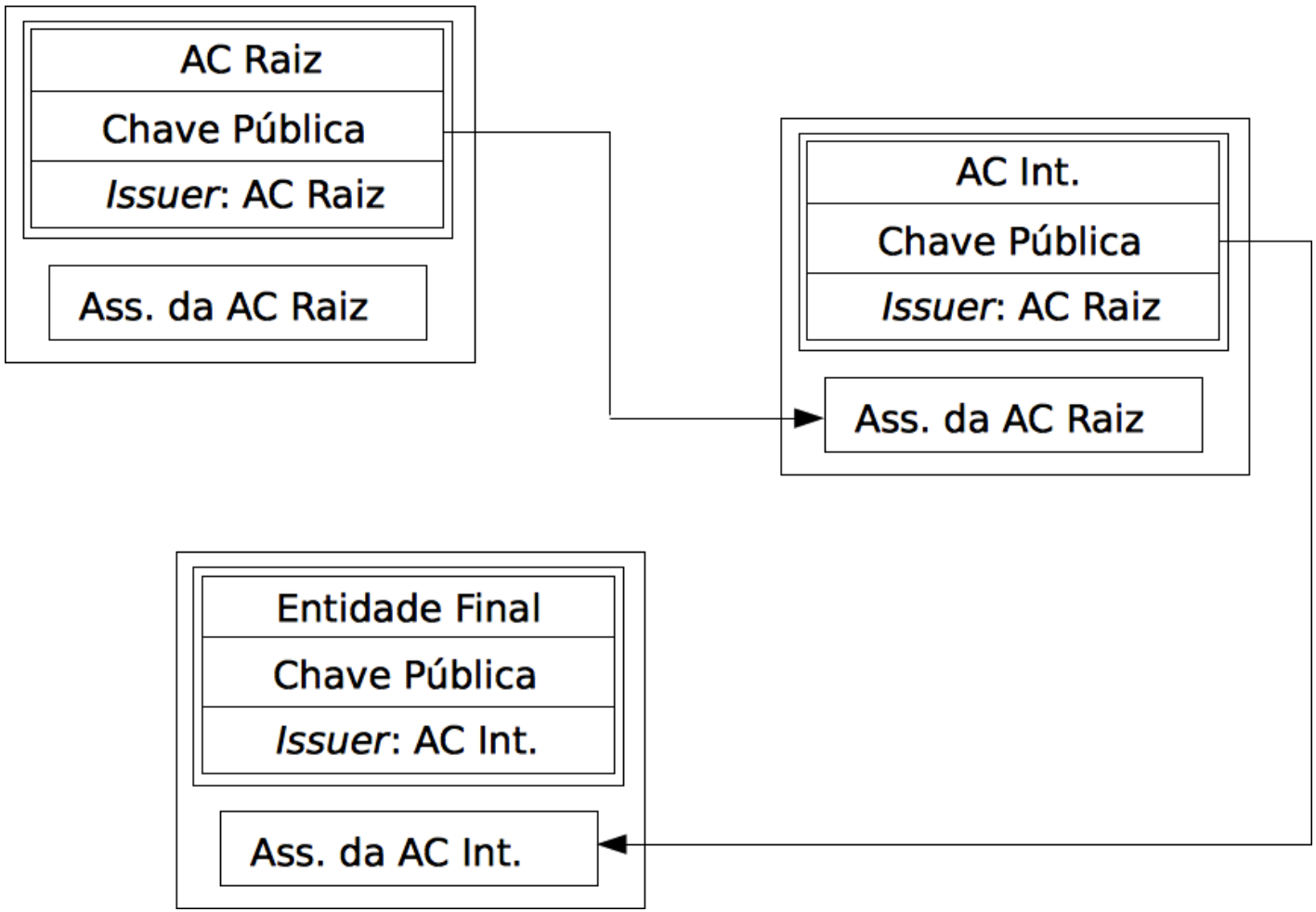
Criptografia e Certificados Digitais

- A revogação pode ser feita diretamente pela AC ou por solicitação da entidade final.
- A AC publica em seu diretório um documento chamado de lista de revogação de certificados (certificate revocation list – CRL).
- O documento contém uma lista de números de série de certificados revogados e é assinado digitalmente pela AC.
- Periodicamente a CRL é atualizada.

Hierarquias de Certificação



Caminhos de Certificação



Grid Security Infrastructure (GSI)

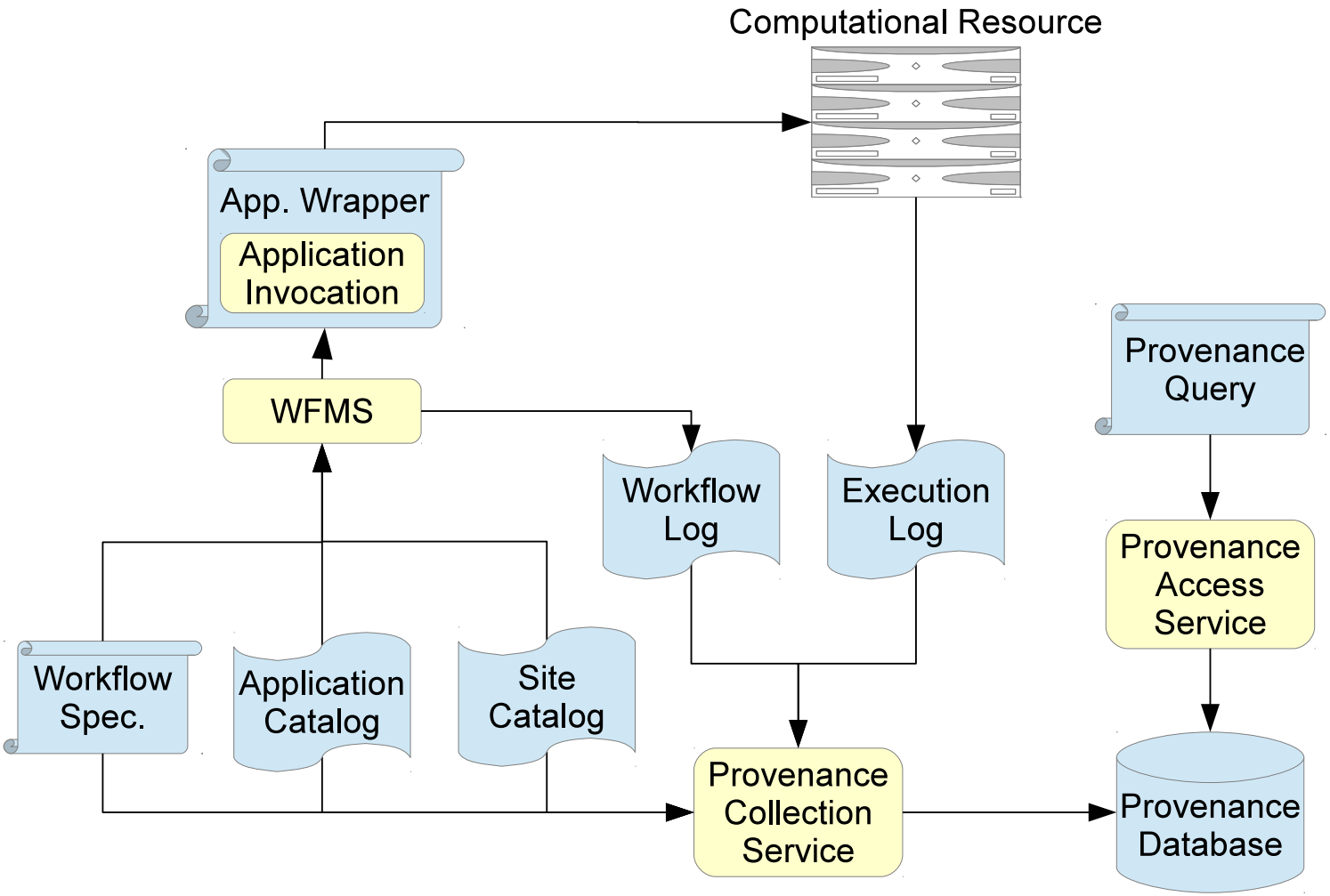
- A GSI utiliza as técnicas apresentadas de criptografia e certificação digital para prover para grades:
  - autenticação;
  - confidencialidade;
  - integridade.
- O DN (*distinguished name*) é único globalmente e é a identidade de um usuário ou host.
- O DN (global) é mapeado para nome de usuários locais em um domínio administrativo.

Kairos: Introduction

- Kairos is given by techniques for applying provenance to protect authorship of computational scientific experiments.
- Main contributions:
  - Evaluation of threats to computational scientific experiments;
  - Extension of Kairos to protect authorship of computational scientific experiments.
  - Prototype implementation and evaluation.

Gadelha, L., Mattoso, M. (2015). Applying Provenance to Protect Attribution in Distributed Computational Scientific Experiments. Provenance and Annotation of Data and Processes (Lecture Notes in Computer Science, Vol. 8628, pp. 139–151). Springer.

Provenance in e-Science



Gadelha, L. et al. (2012). MTCProv: a practical provenance query framework for many-task scientific computing. Distributed and Parallel Databases, 30(5-6), 351–370.

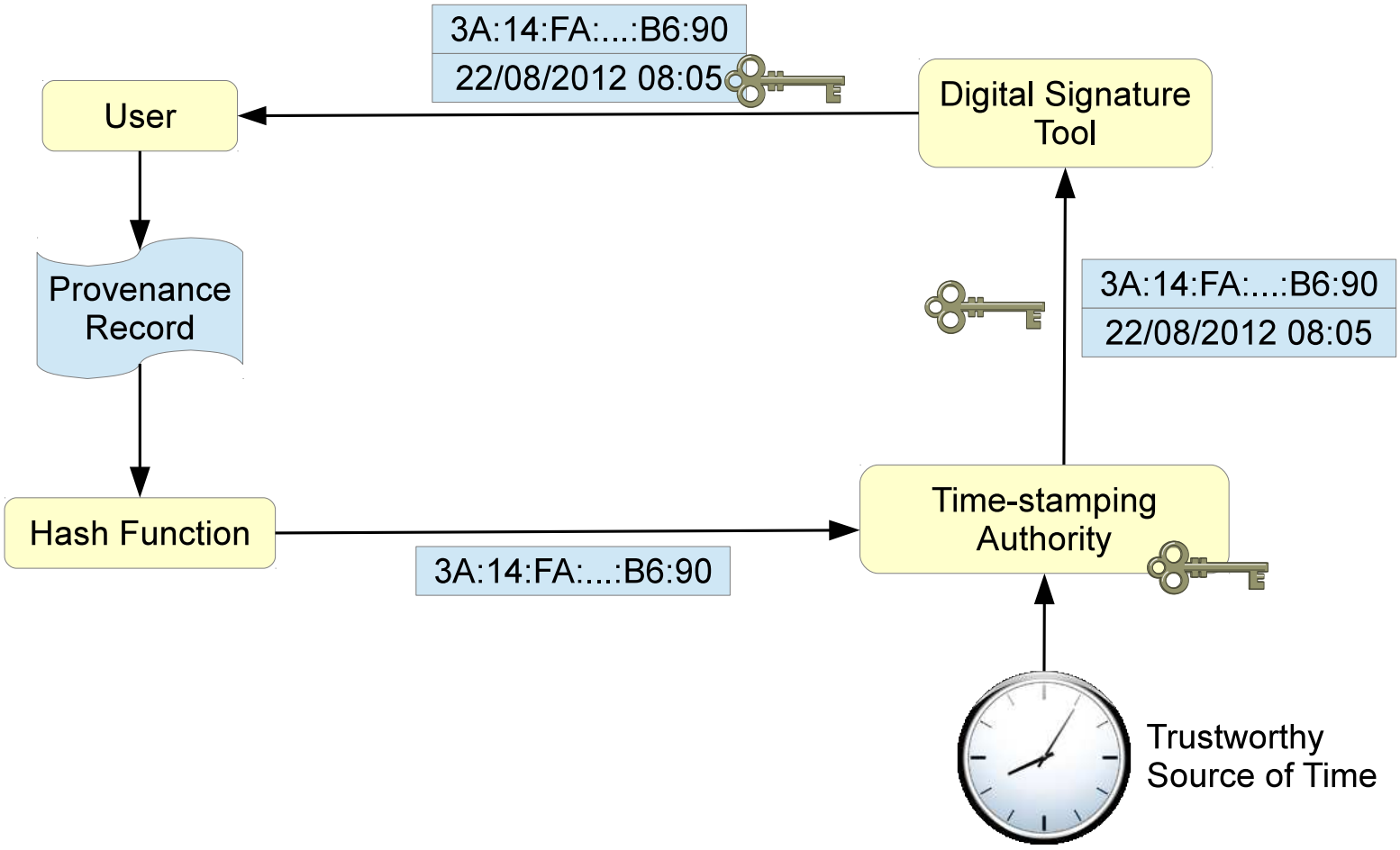
Kairos: Securing Authorship

- Aims the protection of authorship and temporal information of provenance data.
- Use of techniques from public key infrastructures (PKI).
- A trusted entity called Certificate Authority (CA) issues *digital certificates* to other entities.
- Encryption with the private key can be used to perform *digital signatures*.

Kairos: Securing Authorship

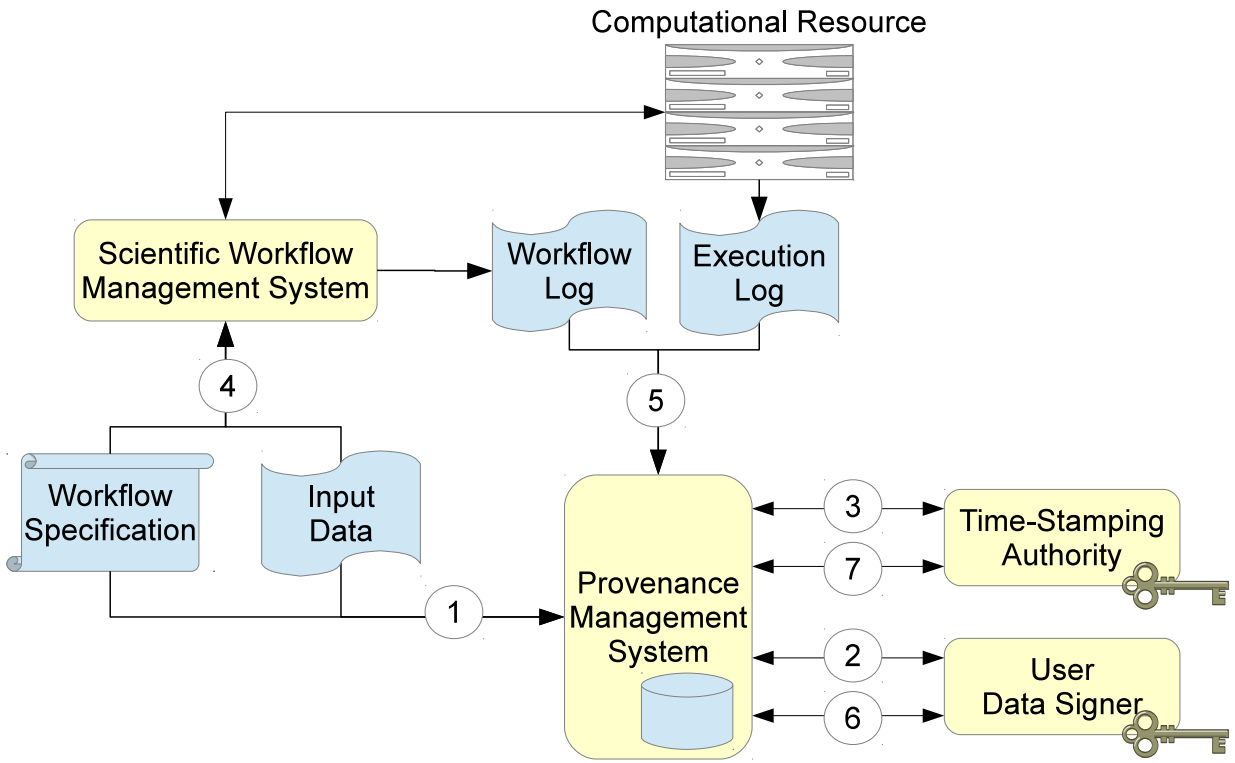
- Time-Stamp Protocol (TSP) can be used to securely determine the date and time in which a data object was generated.
- It requires the availability of a Time-Stamp Authority (TSA).
- The TSA digitally digitally signs tokens containing the hash value of a document and the current date.
- These signed tokens can be used later as evidence that the document existed in the date contained in the token.

Time-stamping protocol



Haber, S. and Stornetta, W. S. How to time-stamp a digital document. Journal of Cryptology 3, 99–111 (1991).

Kairos (extended to prospective provenance)

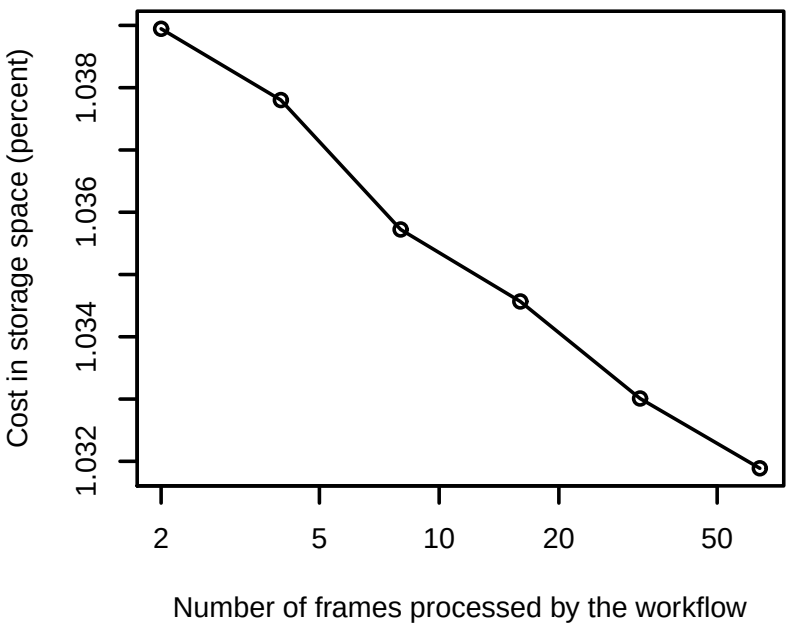


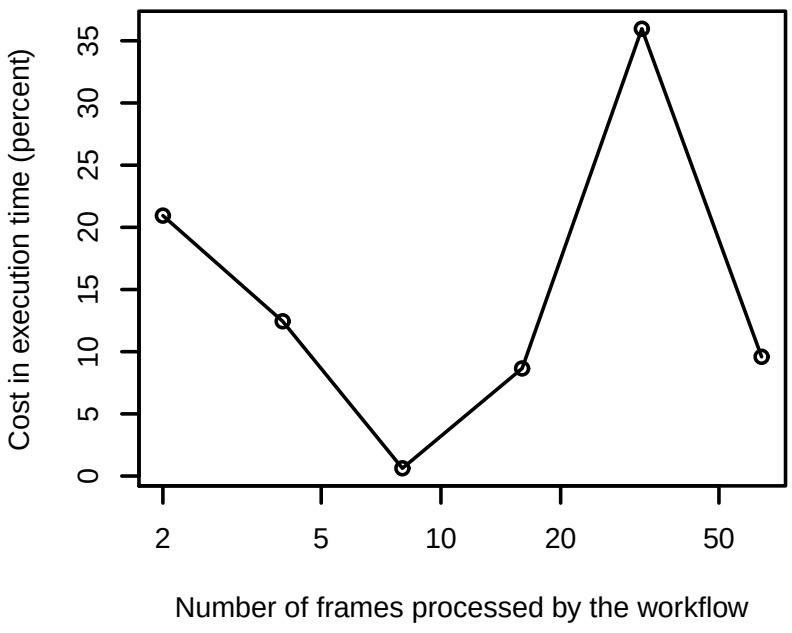
Gadelha, L., Mattoso, M. (2015). Applying Provenance to Protect Attribution in Distributed Computational Scientific Experiments. Provenance and Annotation of Data and Processes (IPAW 2014). Lecture Notes in Computer Science, vol. 8628, pp. 139–151. Springer.

Evaluation

- Kairos implemented in Python as a wrapper to:
  - Swift parallel scripting
  - OpenSSL cryptographic library functions:
    - `smime` for digital signatures.
    - `ts` for the time-stamping protocol.
  - Integrated to MTCProv, Swift's provenance database:
    - Cryptographic data added as annotation to the workflow run.
    - Enables recursive queries for attribution across workflow runs.
- Evaluated with a ray-tracing/video rendering workflow.
- 24-core AMD Opteron server and an 8-core submission host.

Kairos impact





Implementation Issues

- Applicable signature standards:

Data Type	**Signature**	**Time-Stamp**
Plain Text	CMS	TSP
XML document	$\geq$ XAdES-T	

Concluding Remarks

- We surveyed and analyzed security requirements for provenance management systems.
- The security controls implemented are essential to any claim of intellectual property.
- The extension of Kairos allows for better protection of correct authorship attribution since it applies the proposed security controls also at the design phase of the experiment.
- We evaluated of the impact of the proposed techniques in terms of storage space required and execution time, concluding that it is relatively small.