

Linhas de Experimentos: Reutilização e Gerência de Configuração em Workflows científicos

Eduardo Ogasawara
Cláudia Werner

Leonardo Murta
Marta Mattoso



Sumário

- Introdução
- Análise de SGWfC pela perspectiva de Engenharia de Software
- Proposta: Linhas de experimentos
- Desenvolvimento *para e com* reutilização
- Protótipo
- Conclusões

Reutilização em workflows científicos

- reexecutar um workflow previamente montado
- buscar workflows já criados

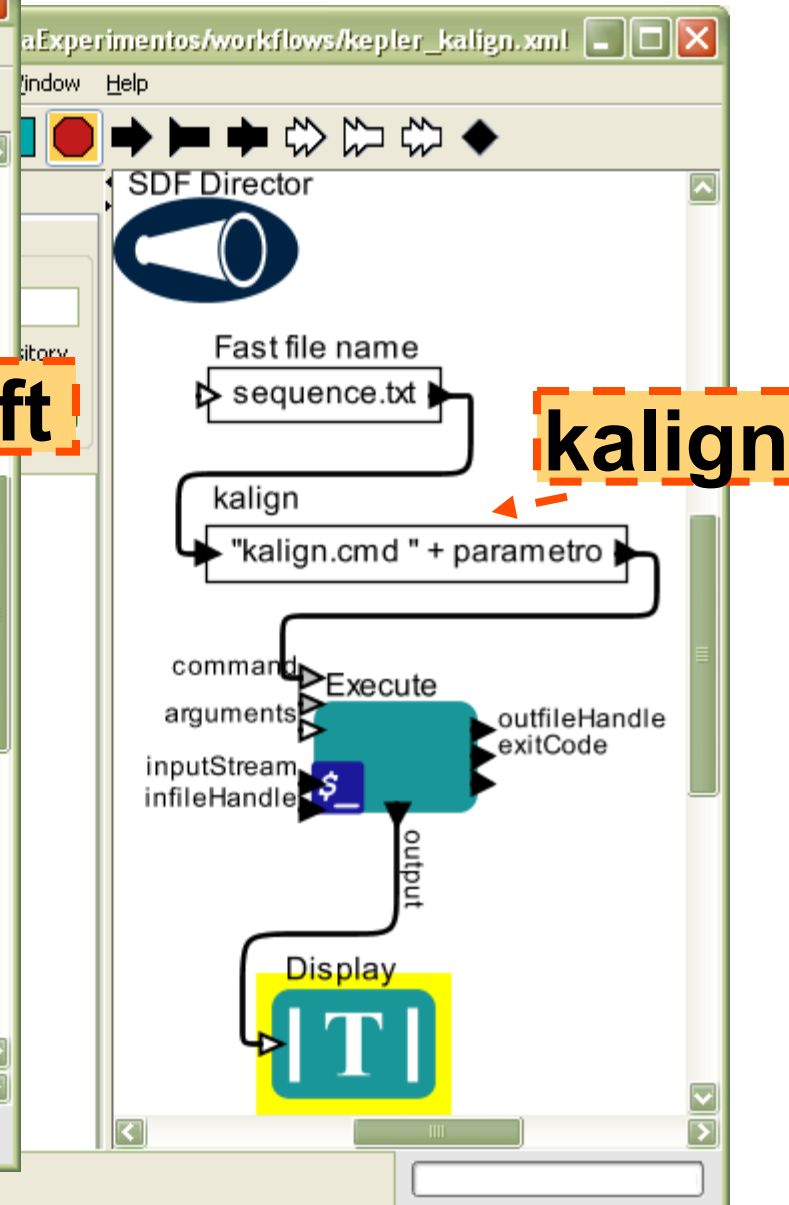
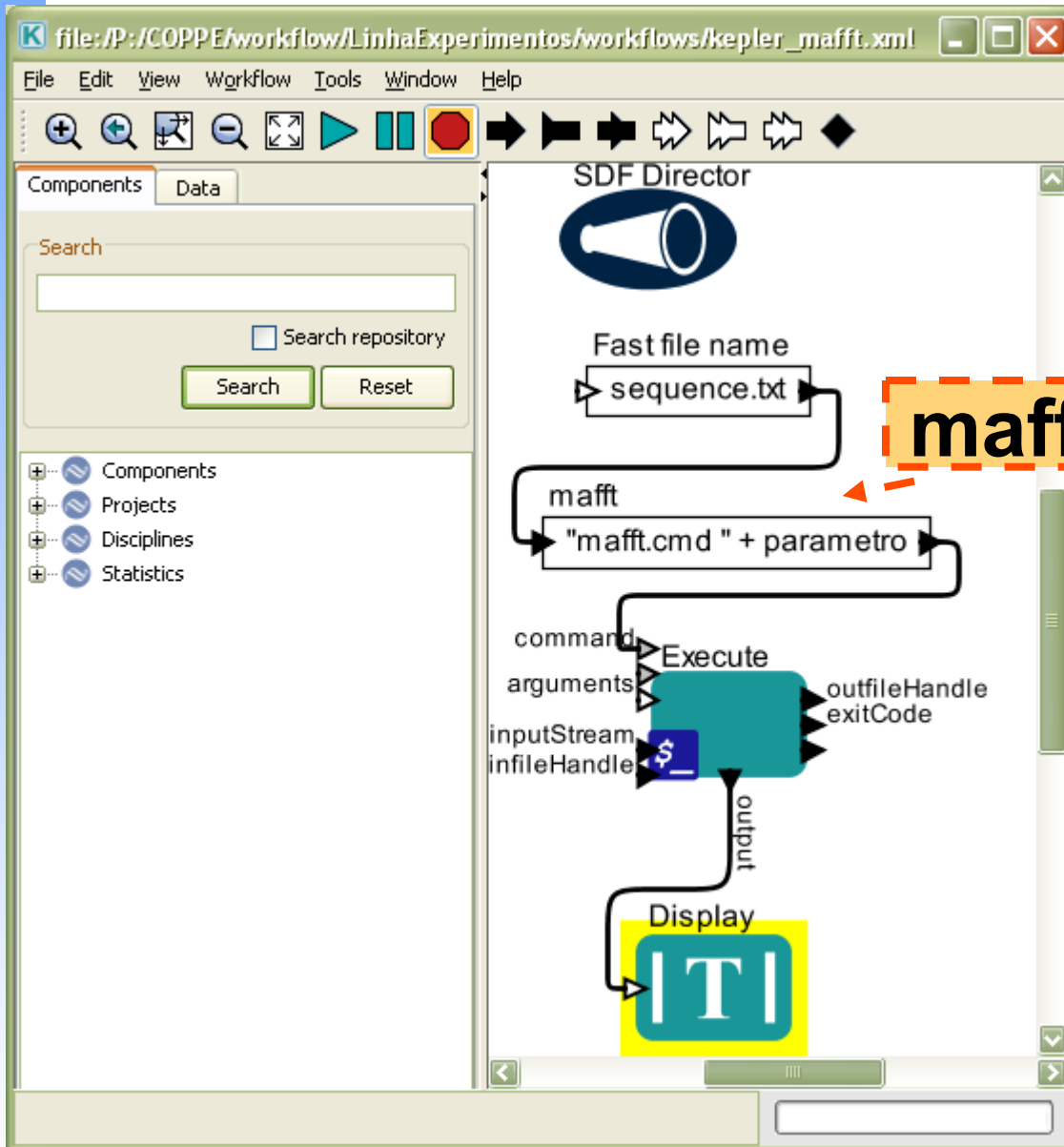
MyExperiment: Um repositório de workflows

The screenshot shows the myExperiment website in a Windows Internet Explorer browser window. The browser's address bar displays the URL <http://www.myexperiment.org/>. The website features a search bar at the top with a dropdown menu set to 'All' and a 'Search' button. Below the search bar, there are three main sections:

- Use myExperiment to...**: A list of actions including 'Find Workflows', 'Find Files', 'Share Your Workflows and Files', 'Create and Find Packs of Items', 'Create and Join Groups', 'Find People and Make Friends', 'Send Messages', 'Get Feedback', 'Tag and Rate things', 'Write Reviews and Comments', and 'Build your Profile and Reputation'.
- Explore**: A section with a 'Find Workflows' button and a workflow diagram. The diagram shows 'Workflow Inputs' (query, program, database) leading to 'species_filter', 'chromosome_filter', and 'blast_arch', which then lead to 'blast_compare'. 'blast_compare' leads to 'Workflow Outputs' (component_output, blast_output).
- Register or Login:**: A registration and login form with fields for 'Username or Email', 'Password', and 'Remember me: '. It also includes an 'Or use OpenID:' section with a field and a 'Login' button. A 'Forgot Password?' link is at the bottom.

At the bottom of the page, there are 'Project Links' including 'myExperiment Wiki', 'Mailing List', 'Give us Feedback', 'myGrid', 'Taverna Workflow Workbench', and 'myExperiment Publications'. The browser's status bar at the bottom shows 'Internet' and '100%' zoom.

Workflows similares



Reutilização no nível de atividades

- Entender atividades como componentes
 - [Gary et al., 1999] [Bowers et al., 2006]
- Trazer semântica entre workflows similares para que o pesquisador possa:
 - Saber que existem variações
 - Explorar variações em suas pesquisas
 - Facilitar a manutenção dos workflows
 - Saber medir quais variações são mais usadas

Estado da arte em reutilização de atividades

- Adição de novas atividades (plug-ins)
 - Kepler, Vistrails, Taverna
- Apoio a subworkflows
 - Kepler e Taverna

Linhas de produtos em Engenharia de Software

- Conjunto de sistemas que compartilham estruturas comuns em termos de aplicabilidade.

[Northrop, 2002][Frakes et al., 2005]

- Sucesso requer planejamento e estruturação dos elementos a sofrer técnicas de reutilização.

Características de linhas de produtos

- Conceito de componentes
- Ativos de reutilização
- Pontos de variabilidade
- Pontos de opcionalidade
- Regras de composição
- Estabelecimento de recortes

Proposta:

Linhas de experimentos

- Representação do encadeamento de atividades para experimentação num nível mais abstrato.
- Separação do processo básico da escolha de um determinado algoritmo ou técnica.

Workflow padrão

Encadeamento de atividades gerais com pontos de variabilidade e opcionalidade.

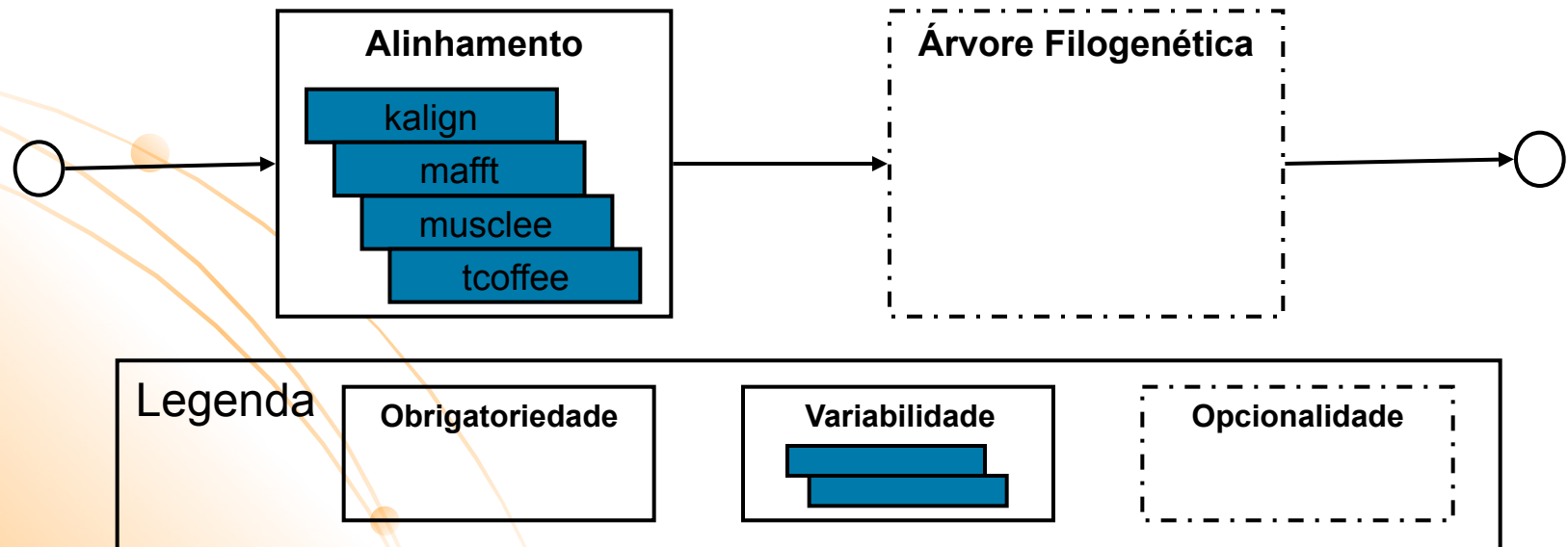


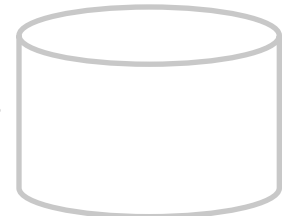
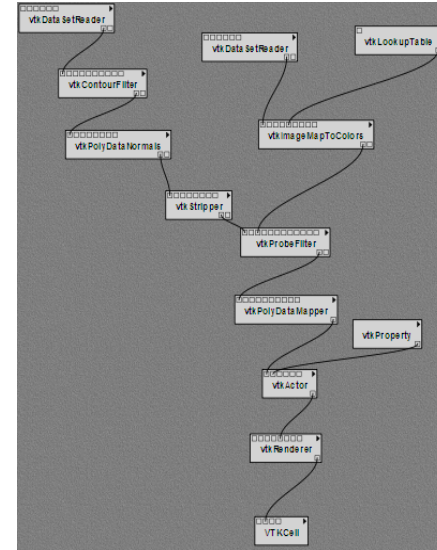
Tabela de equivalência entre linhas de produtos e linhas de experimentos

Linhas de produtos	Linha de produto	Workflow padrão	Linhas de experimentos	
	pontos de variabilidade	Atividades e relacionamentos variantes		
	pontos de opcionalidade	Atividades e relacionamentos opcionais		
	Regras de composição	Regras de encadeamento		
	Recorte	Geração de workflow a partir da linha		
	Features	Recorte guiado		

Desenvolvimento para reutilização



Requisitos

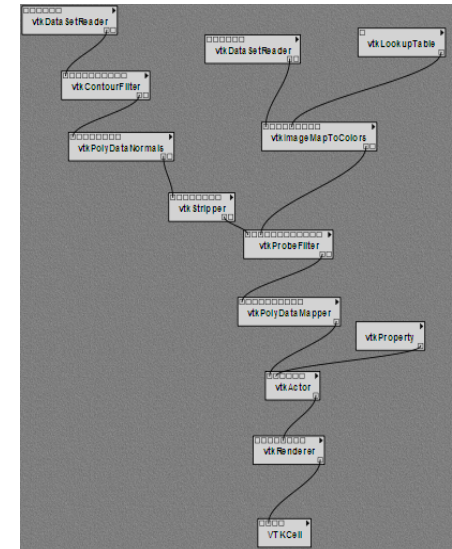
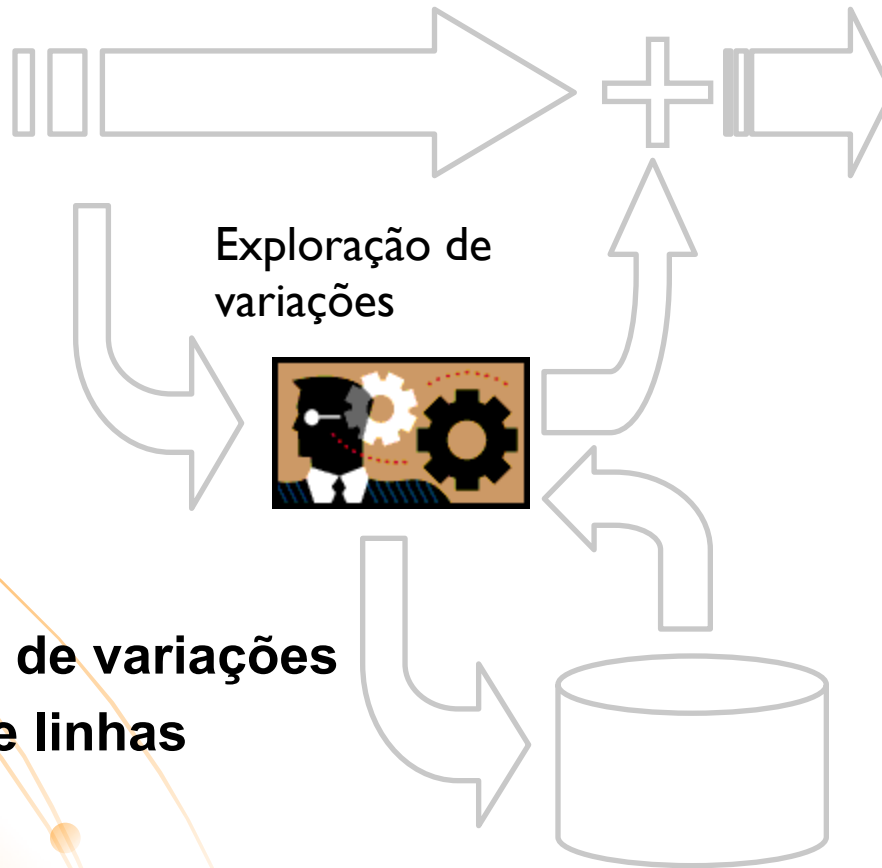


- Há um trabalho inicial maior de planejamento para se compor uma linha de experimentos
- Mudanças na linha podem ser refletidas nos workflows montados

Desenvolvimento com reutilização

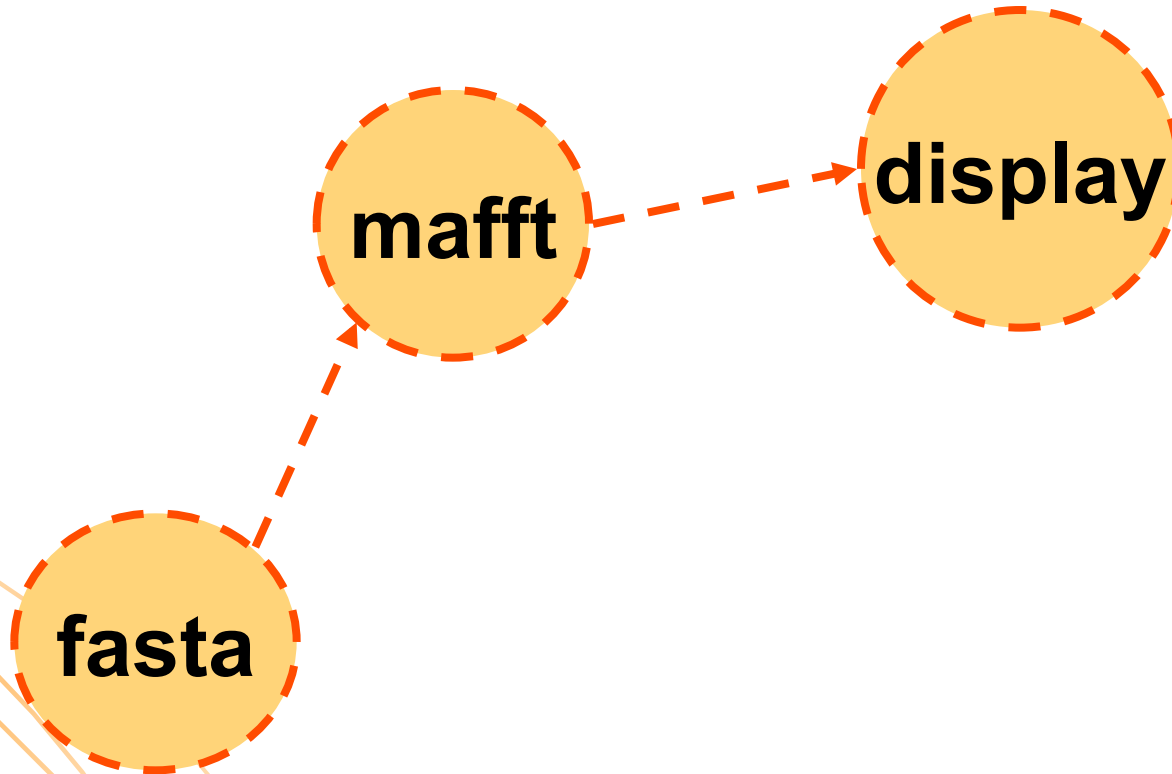


Requisitos

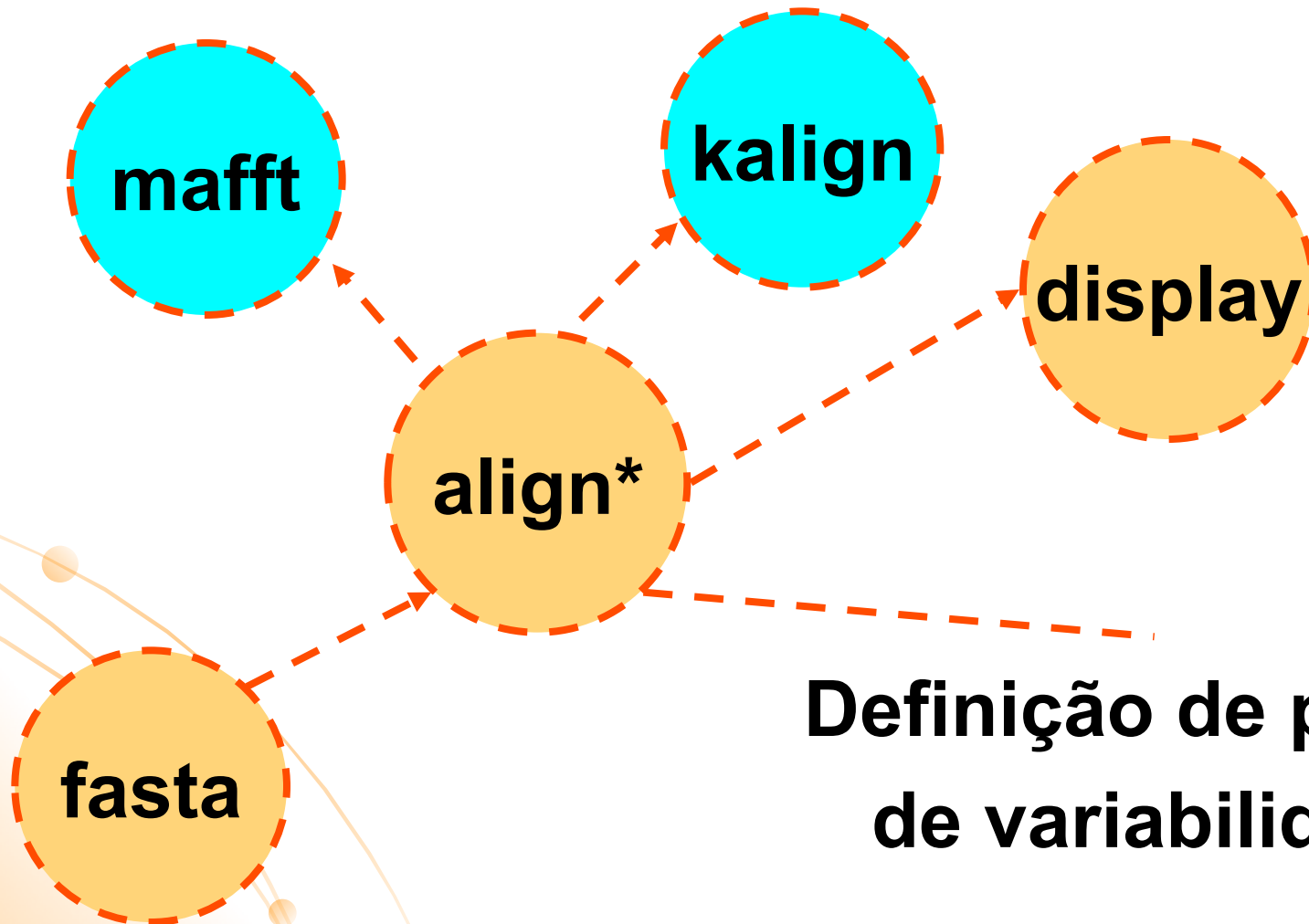


- Exploração de variações
- Recortes de linhas

Construção da linha a partir de um workflow

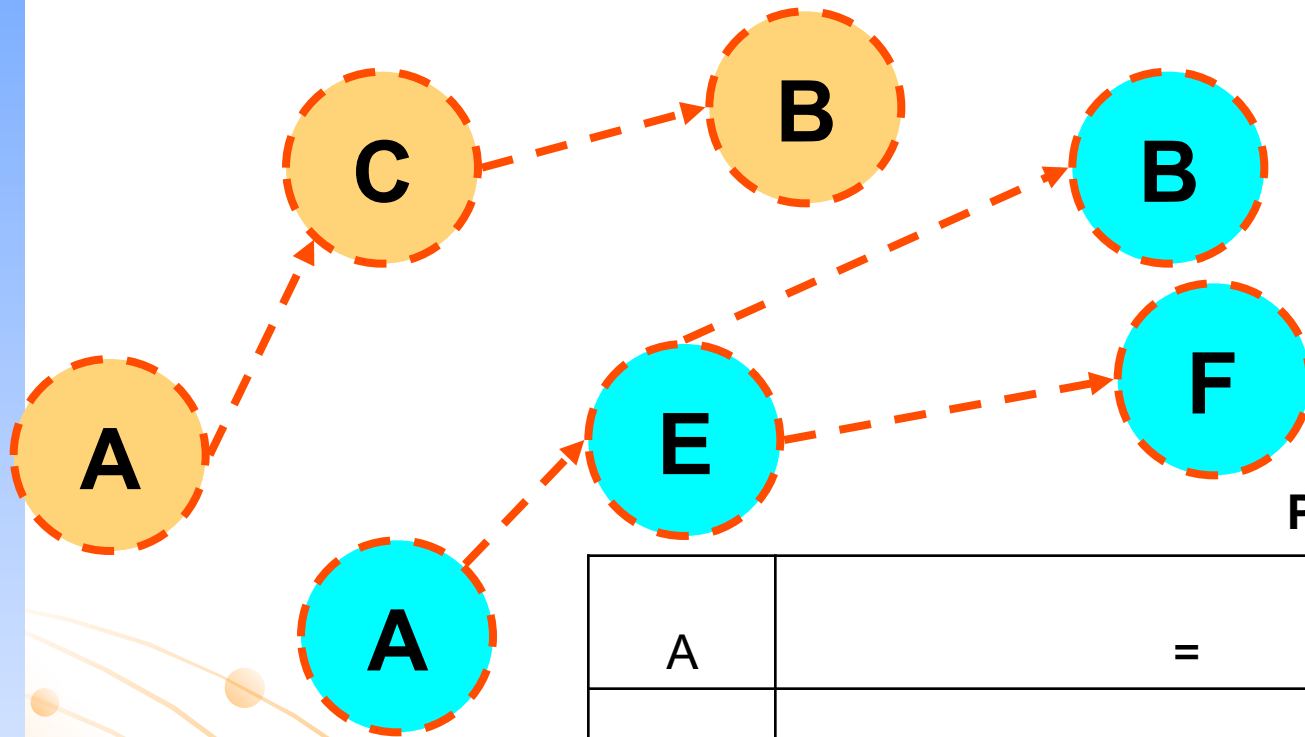


Indicação manual de ponto de variabilidade



**Definição de ponto
de variabilidade**

Dif/Merge de workflows



Busca de subgrafos semelhantes

Tabela de dif/merge
Para construção da linha

A	=	A
B	=	B
C	ponto de variabilidade	E
	incluir atividade opcional	F

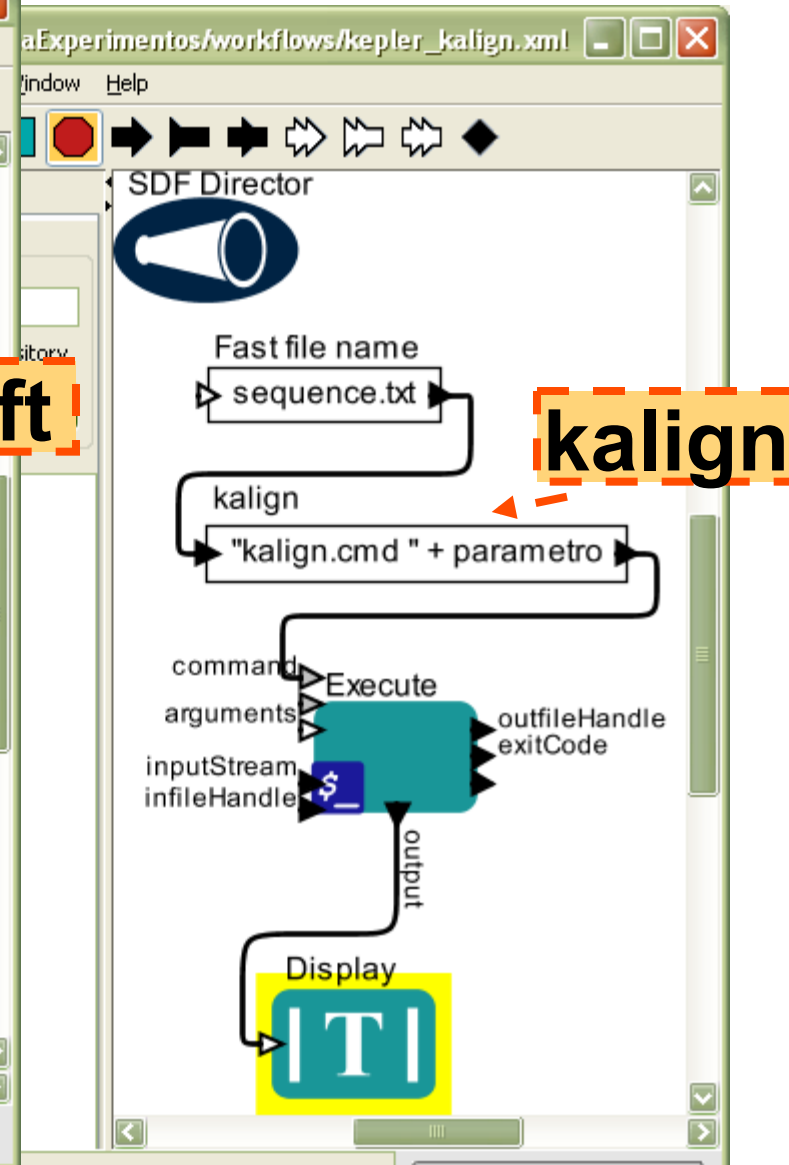
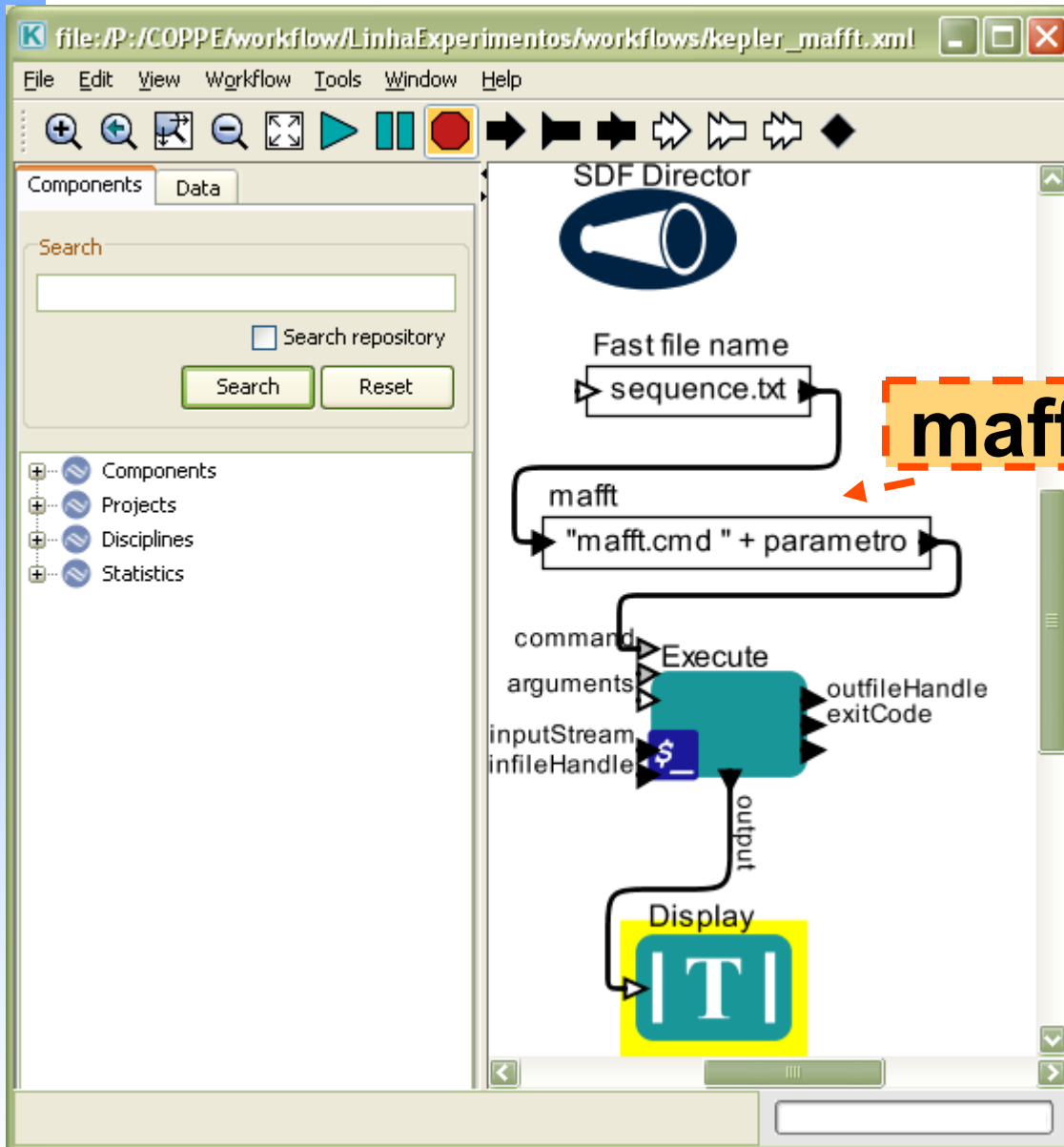
Protótipo

- Objetivo de avaliar os conceitos apresentados.
- Desenvolvido um protótipo em Java
- SGBD Postgres
- XPDL-Specifications do WfMC.

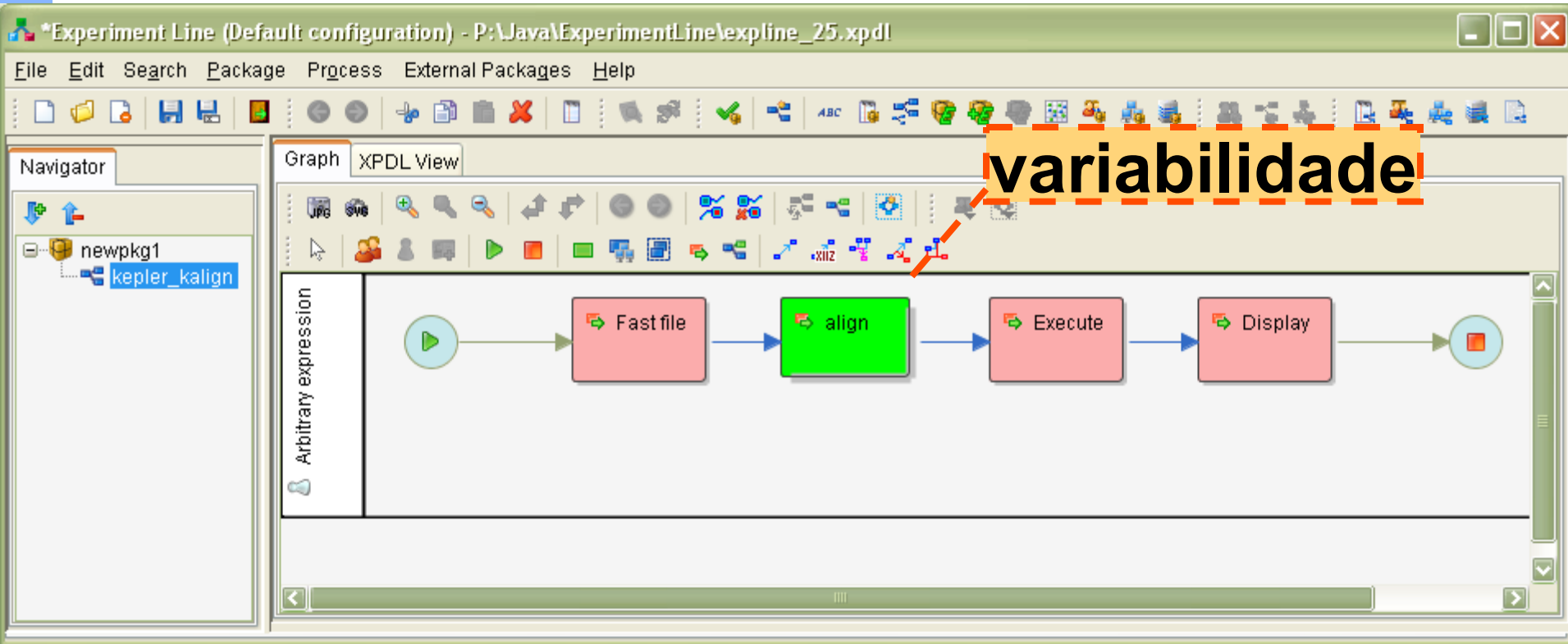
Características do protótipo

- Workflow padrão (nível mais abstrato)
- Criação de linhas de experimentos a partir de um workflow inicial
- Cartuchos para exportação/importação de workflows para Kepler, *Taverna* e *Vistrails*.
- Variabilidade e opcionalidade de atividades e relacionamentos

Retomando ...



Protótipo: gerenciador de linhas



Edição de atividade de alinhamento

Nível de abstração

Activity Property

Name: align

Class: ptolemy.actor.lib.Expression

Description:

Optional activity Active

Ports Variant activities

Convert to normal relation

ID	Name
76	kalign
77	mafft
78	muscle
79	tcoffee

Add Edit Delete Choose

Ok Cancel

Items de variabilidade

Edição de relacionamento

Transition Property

Name:

Class:

Optional relation

Normal relation

From:

To:

Converte em item variável

Conclusões

- Aplicação de conceitos de reutilização para composição de workflows.
- Maior formalização via workflow padrão, que permite:
 - Explorar variações;
 - Facilitar a manutenção;
 - Medir uso a partir dos recortes;

Trabalhos em andamento

- Este trabalho faz parte do projeto de Gerência de Experimentos Científicos em Larga Escala (GEXP)
- Alinhado ao tema de doutorado.
- Aplicação do conceito de linha de experimentos para cadeia de petróleo e bioinformática.
- Exploração mais aprofundada de gerência de configuração.



Linhas de Experimentos:
Reutilização e Gerência de Configuração em
Workflows científicos

Obrigado!

Visite nosso sítio
<http://gexp.nacad.ufrj.br>

Eduardo Ogasawara
eduardo.ogasawara@gmail.com

