



geosparc
the geospatial architects

Evolution of GIS data

By Joachim Vander Auwera



Abstract

- GIS data is not static. Even without considering the move of the tectonic plates which form the surface of the earth, it is unlikely that data is read-only during its lifespan. The data will change, and still the users expect to have a consistent view on the world. Using transactions in your GIS data store is often no solution as editing and approval may take a long time to finish.
- This talk discusses some options about versioning your data by investigating how this is handled in various systems and looks at some of the implications for the applications and user experience. It looks at the inclusion of versioning of your data, the use and application of business processes and business rules and the effects on scalability.

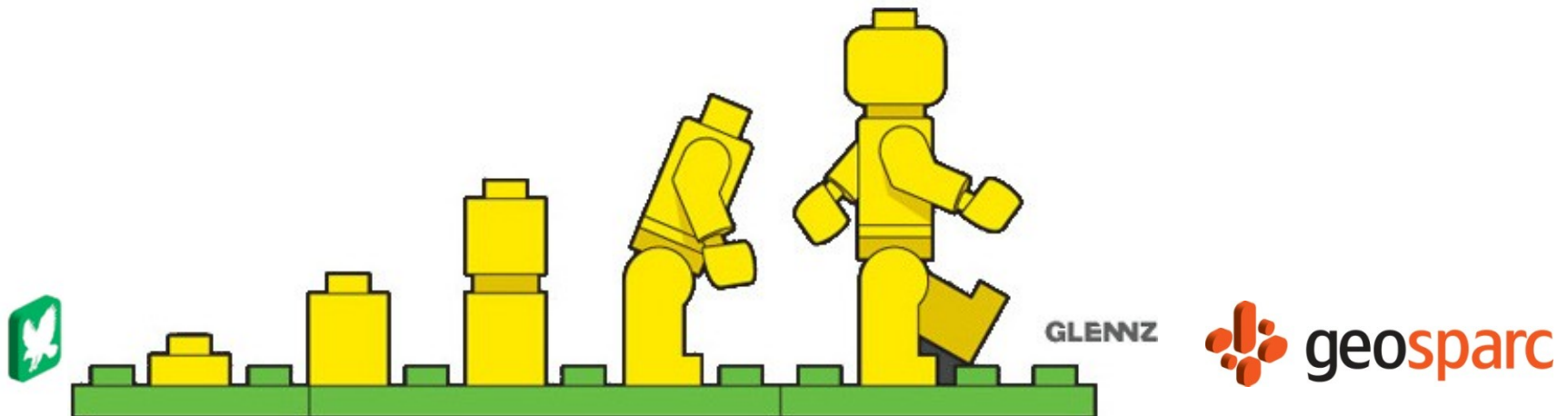
Who am I

- 20+ years commercial programming experience
- 10+ years Java
- Focus :
 - enterprise software
 - code quality
 - maintainability
- Geomajas dev since August 2009



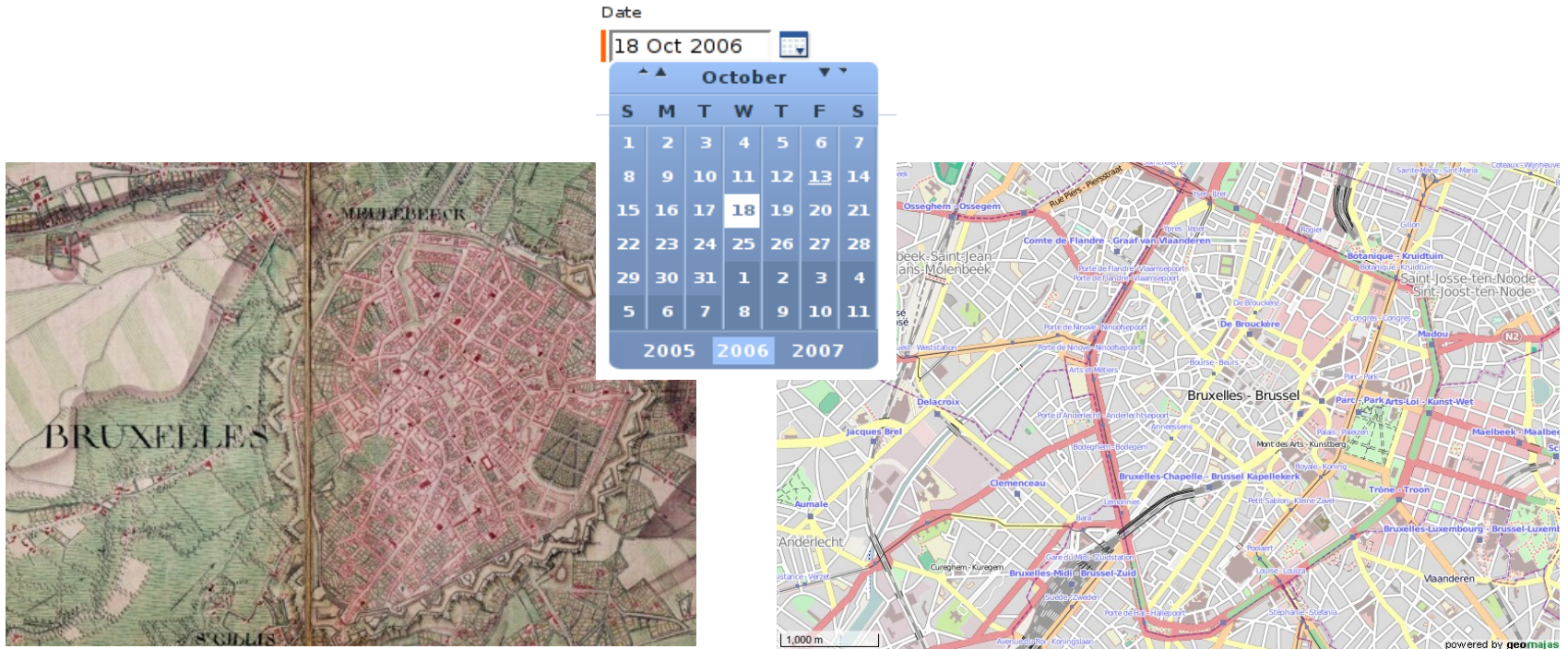
ev·o·lu·tion ?

- **Any process of formation or growth; development:** *the evolution of a language; the evolution of the airplane.*
- A product of such development; something evolved: *The exploration of space is the evolution of decades of research.*
- **A process of gradual, peaceful, progressive change or development,** as in social or economic structure or institutions.
- A motion incomplete in itself, but combining with coordinated motions to produce a single action, as in a machine.
- ...



Use cases

- Allow historic overview – data from yesterday, last week, 5 years ago



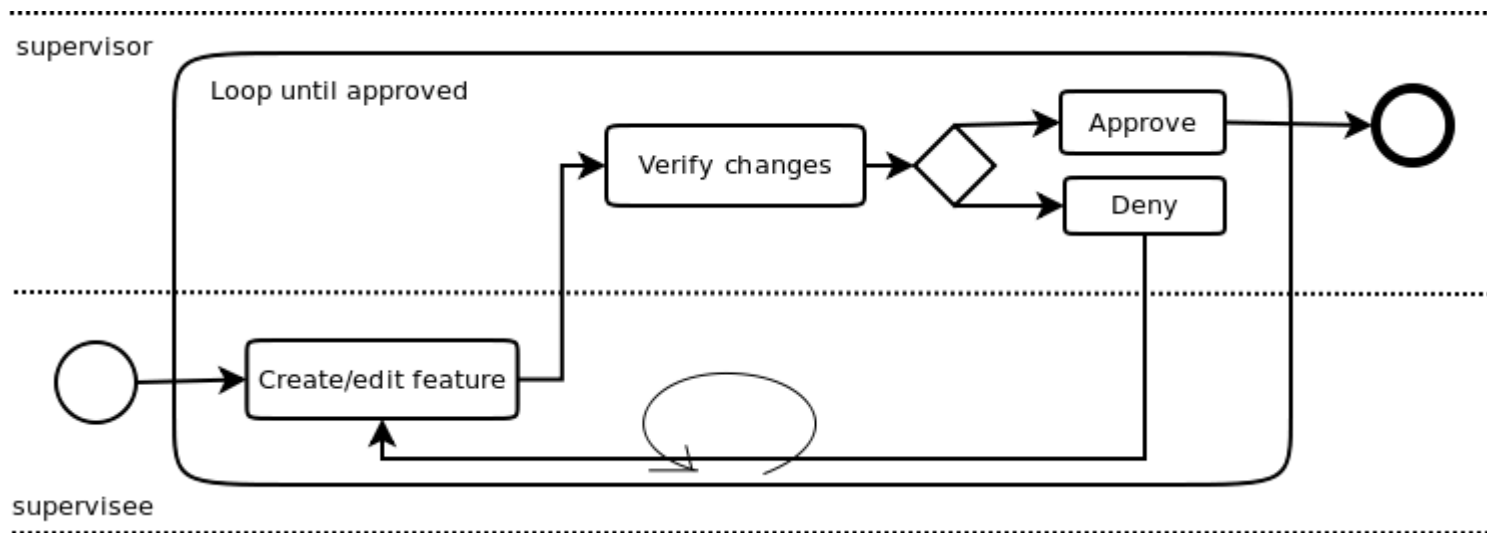
Use cases

- Editing changes can take long



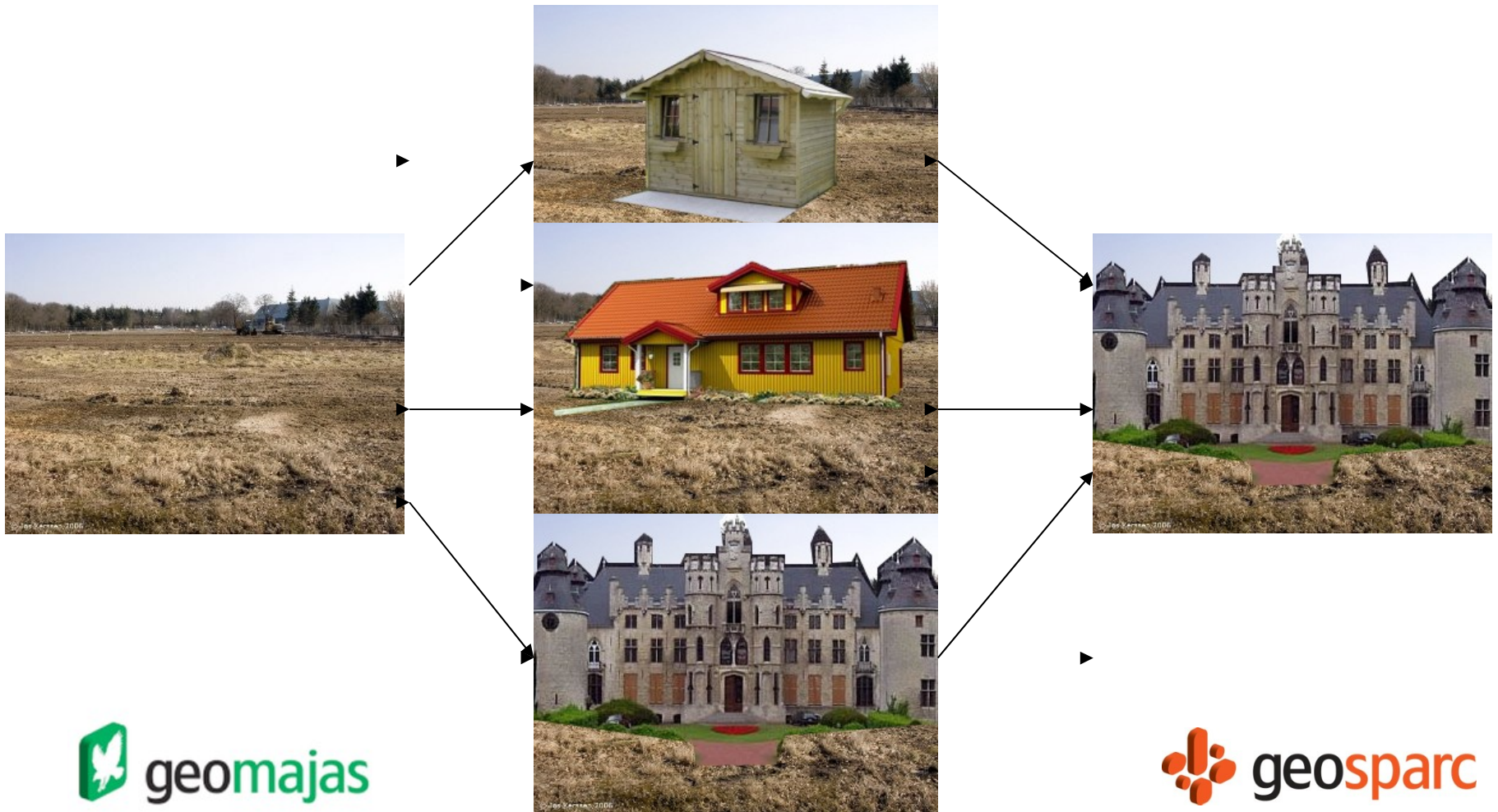
Use cases

- Changes may need approval



Use cases

- Options may be explored in parallel



Use cases, challenges

- Planning
- Auditing, compliance (Sarbanes-Oxley, Basel et al)
- Merging?



Solutions

Possible solutions

Workflow solution

Long lasting transactions

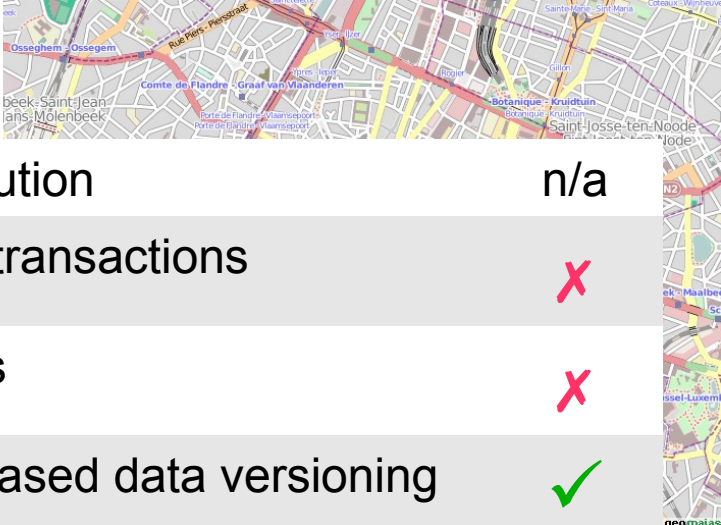
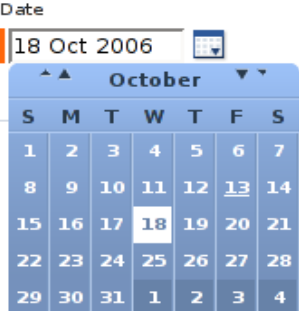

Editing layers

Timestamp based data versioning

Revision based data versioning

Use cases

- Allow historic overview, data from yesterday, last week, 5 years ago

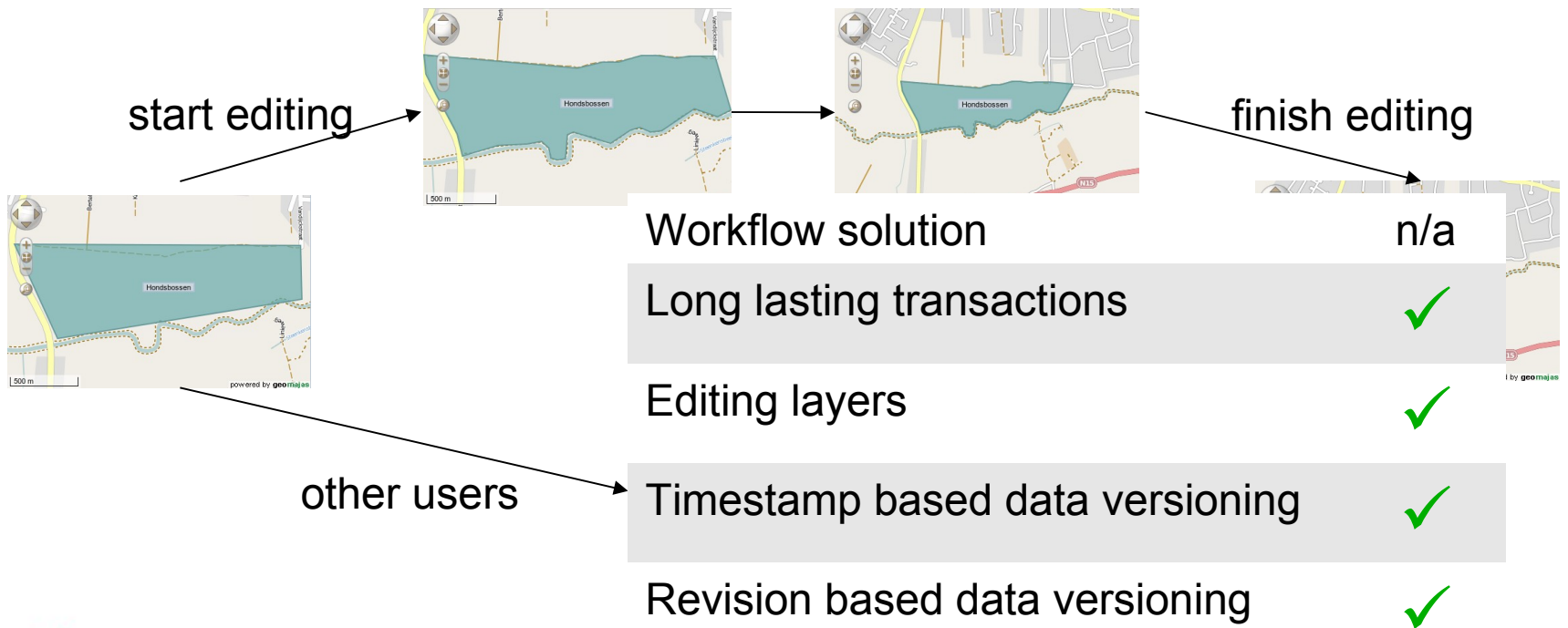


Date						
18 Oct 2006						
October						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Workflow solution	n/a
Long lasting transactions	X
Editing layers	X
Timestamp based data versioning	✓
Revision based data versioning	✓

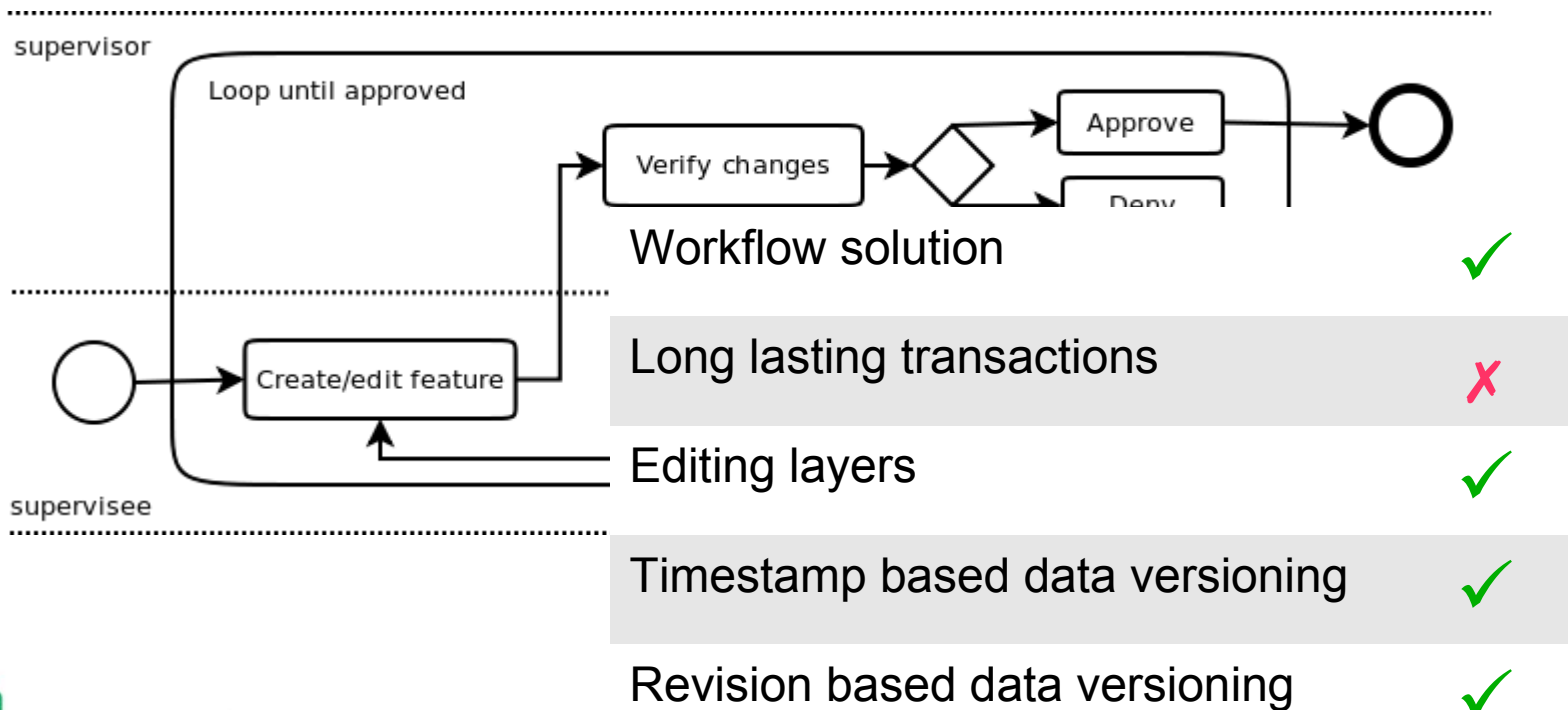
Use cases

- Editing changes can take long



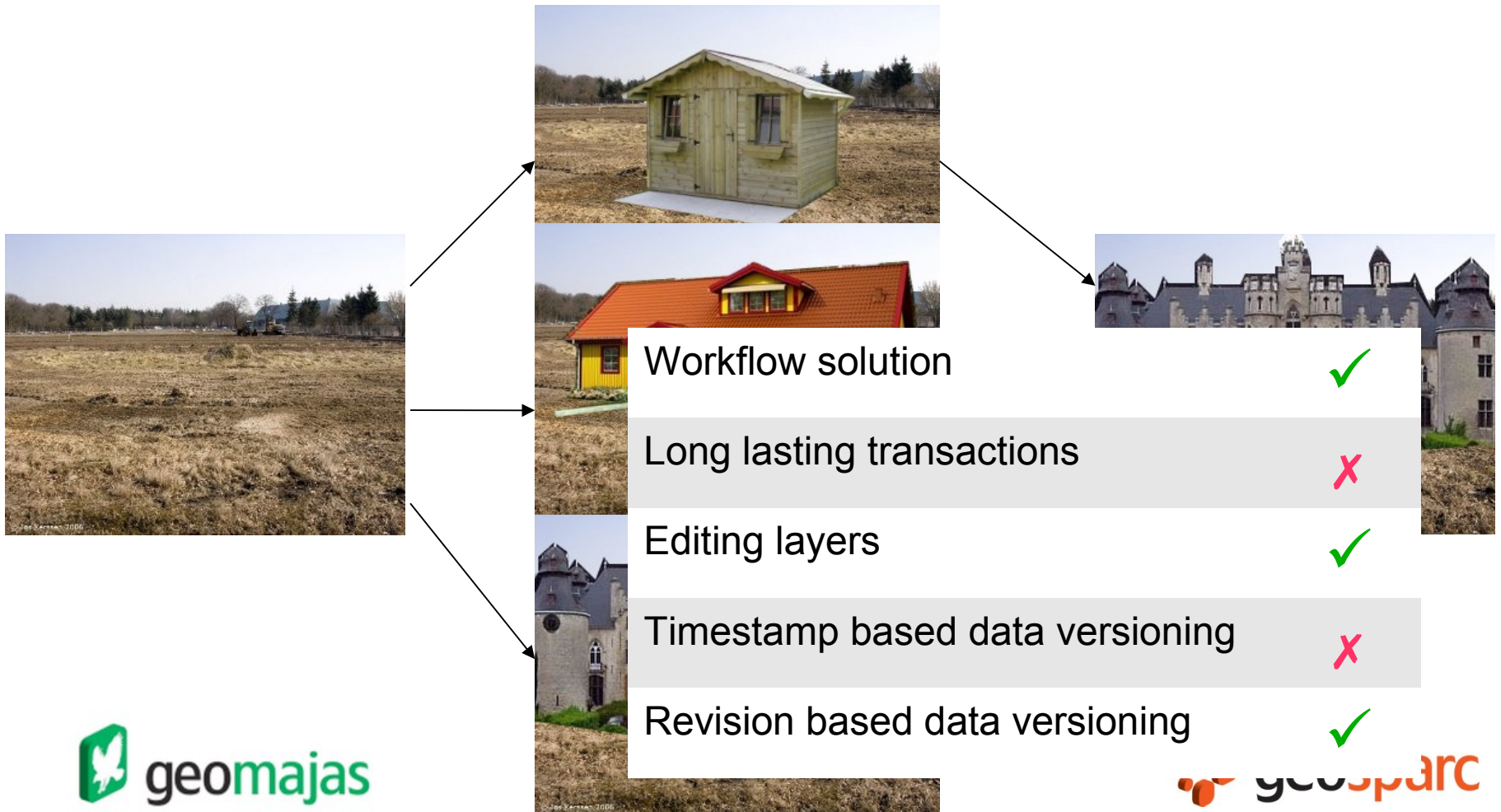
Use cases

- Changes may need approval



Use cases

- Options may be explored in parallel

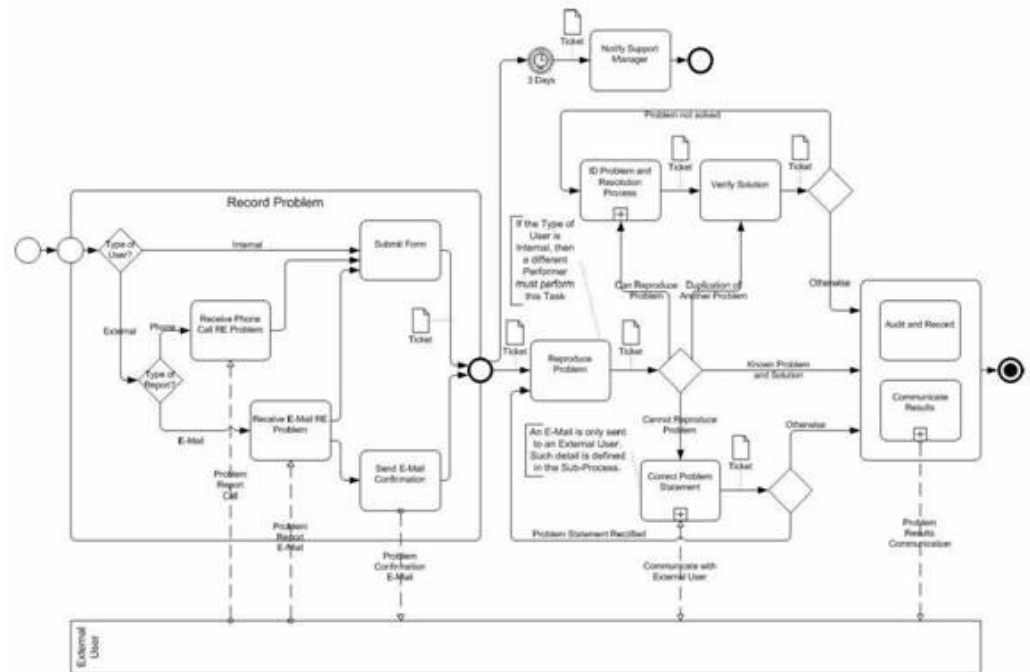


Use cases & solutions, overview

Workflow solution	n/a	n/a	✓	✓
Long lasting transactions	✗	✓	✗	✗
Editing layers	✗	✓	✓	✓
Timestamp based data versioning	✓	✓	✓	✗
Revision based data versioning	✓	✓	✓	✓

Workflow solutions

- Wait states, asynchronous calls
- Standards : BPEL, BPEL4People, BPMN, BPMN2
- Open source solutions
 - Activity
 - JBoss jBPM
 - Intalio
 - Drools flow
 - ProcessMaker
 - Bonita Studio
 - Uengine
 - ...



Versioned data

- Standards : none
- Proprietary database support
- Do it yourself
- Possible solutions
 - timestamps
 - RCS-like versioning with tags, branches and merging

Spatial database solutions

database	native data versioning ?	extensions or plug-ins ?
PostGis	no	* (timetravel, geoserver)
Oracle Spatial	yes	Flashback
DB2	yes (DB2 10)	
MySQL	no	*
MS SQL	yes	Change Data Capture



Do it yourself

- In one table or split table
- One table : fast insert, slow “current” query
- Split table: fast “current” query, slow insert

Person:

Id	Name	Surname
123	John	Smith
857	Brad	Pitt
698	Mary	Doe
...		

Person_AUD:

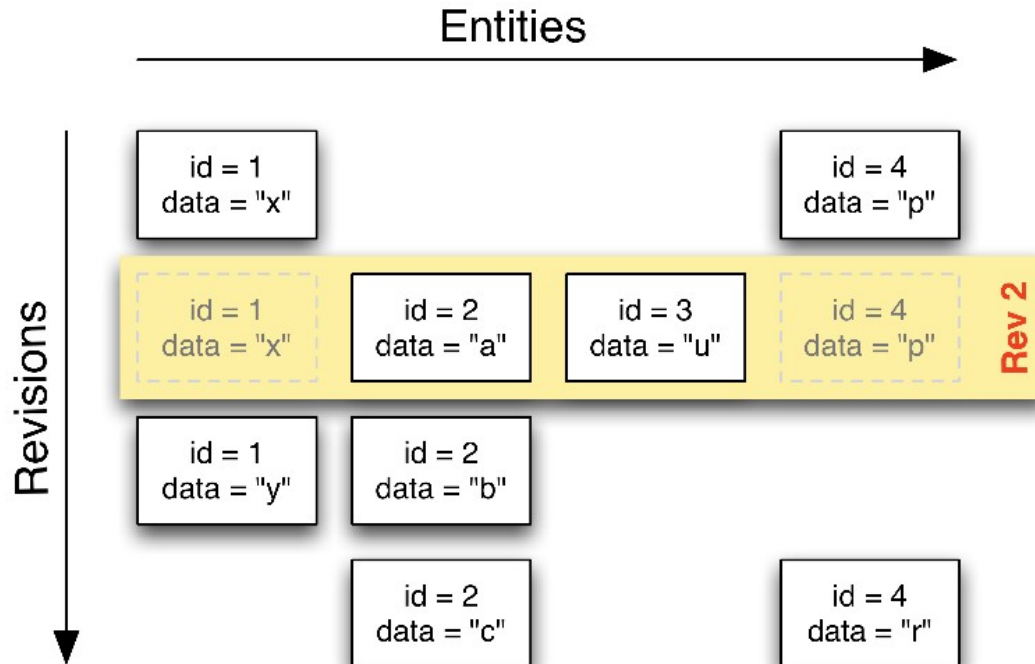
Rev number	Id	Name	Surname	Rev type
64	123	James	Smith	ADD
64	857	Brad	Pitt	ADD
85	123	Peter	Smith	MOD
90	501	Arnold	Schwarzenegger	DEL
90	123	John	Smith	MOD

Simple versioning, timestamp

- Timestamp field, always create new record
- One or two fields:
 - `validFrom`
 - `validFrom` and `validUntil`
- Current query
 - `HAVING validFrom = MAX(validFrom)`
 - `HAVING validUntil IS NULL`
- Timestamp query :
`WHERE validFrom<? AND (validUntil>=? OR validUntil IS NULL)`
- Which timestamp not-approved data, alternatives?
- Visible work in progress?
- How to handle parallel changes?

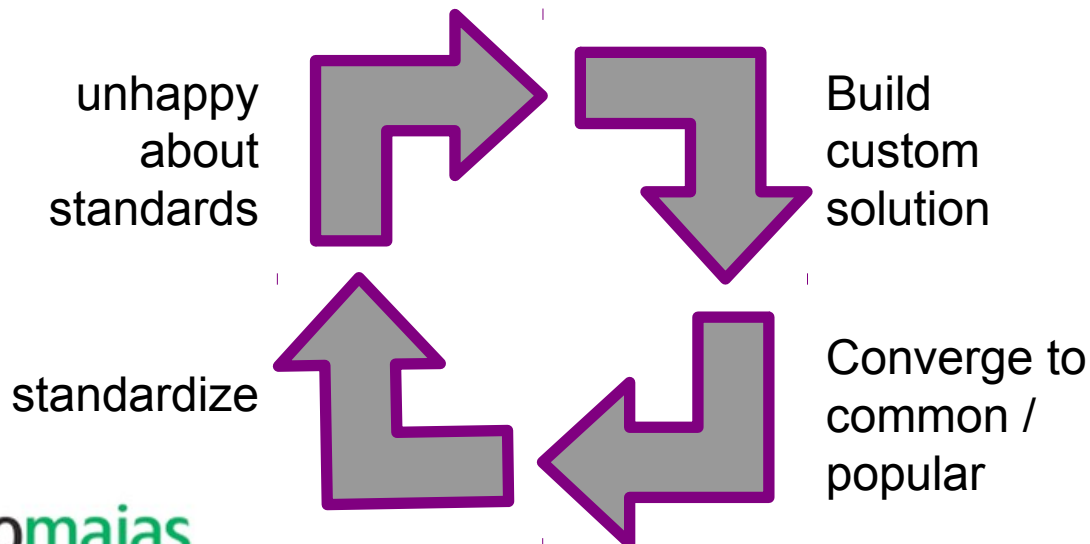
Better versioning, RCS-like

- Subversion or git-like, store revisions with possible branching and tagging
- No native support in any database
- Java : Hibernate + envers



Step back

- Are relational databases and SQL still relevant?
- NoSQL
- Very powerful
- Only partly off-the-shelf



Integration



Roadmap

- Hibernate envers layer
- BPM plug-in

Questions?

Thanks!

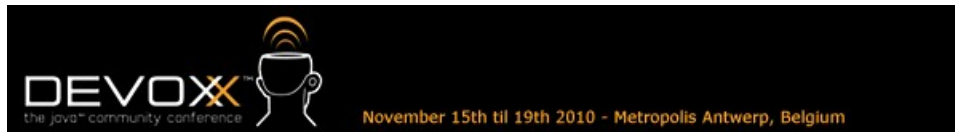
Contact me : joachim@geosparc.com, <http://blog.progs.be/>

Twitter : @joachimvda, @geomajas

www.geomajas.org

www.geosparc.com

Meet us @ Booth # 12



Devoxx, Belgium :

- Geomajas users meeting, November 15, 21h access to all!
- Introduction to the Geomajas web mapping framework, November 16, 13h, room 5



Illustrations

- <http://www.clipartof.com/details/clipart/31485.html>
- <http://store.glennz.com/evolution.html>
- <http://www.phon.ucl.ac.uk/home/wells/blog0802b.htm>
- <http://wieland.wordpress.com/2008/11/28/de-cartes-marchandes-van-de-ferraris-1778-op-europeana/>
- ...