INTEGRATING A CUSTOMISED SMT SYSTEM…
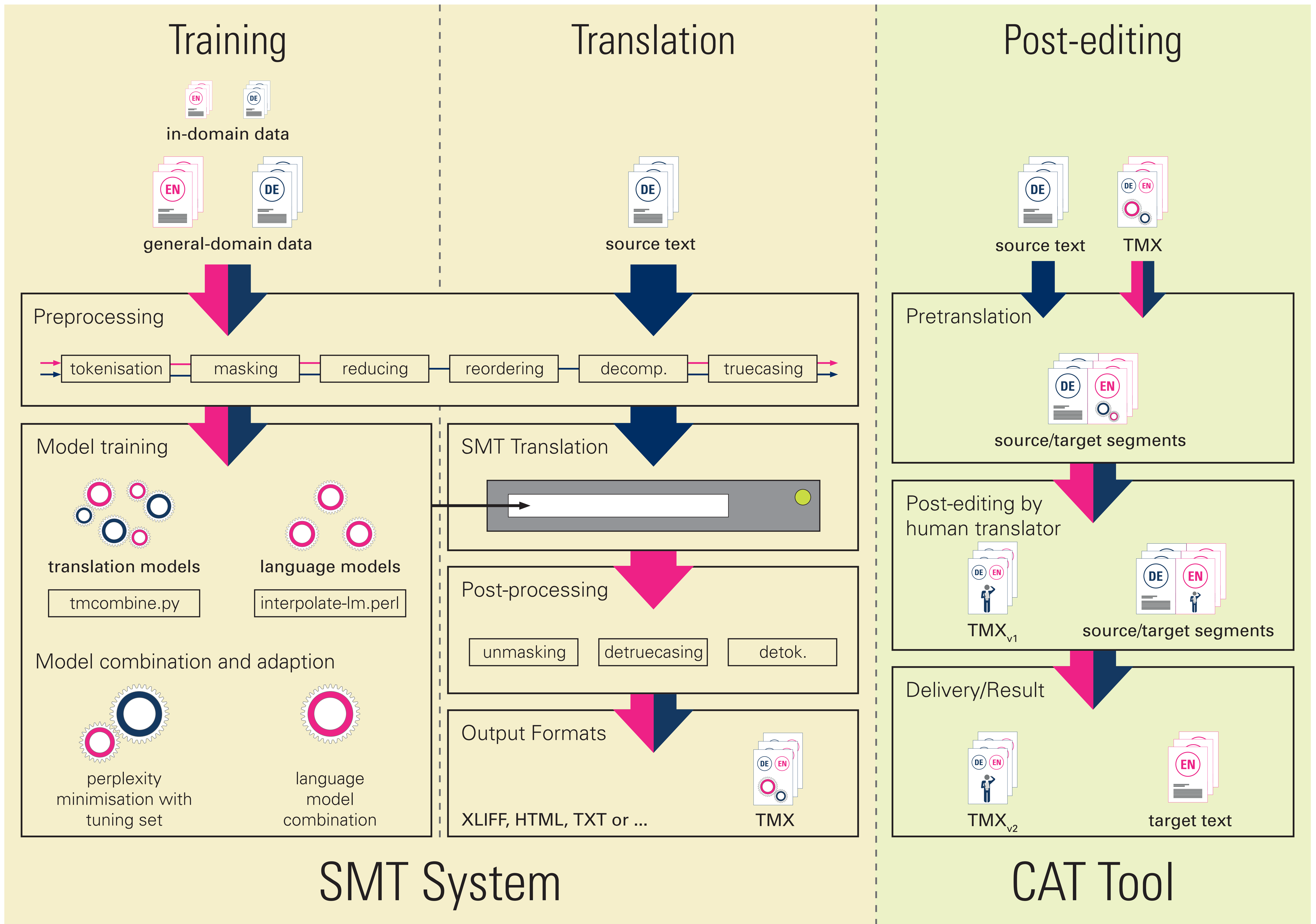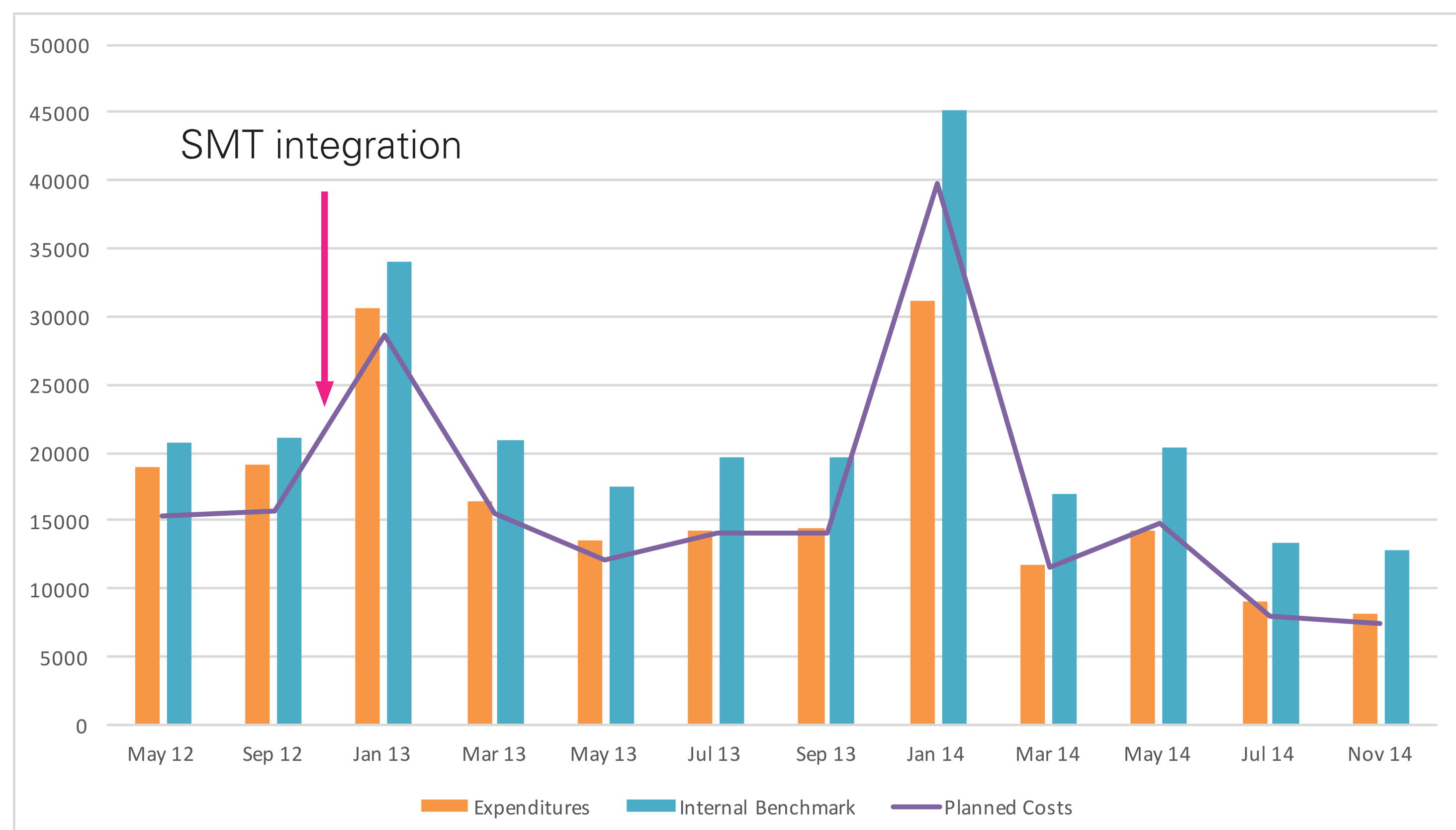
… INTO TRADITIONAL COMPUTER-AIDED TRANSLATION WORKFLOWS

Every year, Finnova translates 6,000 customer support messages within software release letters from German to English. The aim of the project was to develop a translation system tailored to translate those texts automatically.

Training

in-domain data

general-domain data

Preprocessing
tokenisation | masking | reducing | reordering | decomp. | truecasing

Model training
translation models

language models

tmcombine.py

interpolate-lm.perl

Model combination and adaption
perplexity minimisation with tuning set

language model combination

Translation

source text

SMT Translation

Post-processing
unmasking | detruecasing | detok.

Output Formats
XLIFF, HTML, TXT or …

TMX

Post-editing

source text | TMX

Pretranslation
source/target segments

Post-editing by human translator
TMX$_{v1}$ | source/target segments

Delivery/Result
TMX$_{v2}$ | target text

SMT System

CAT Tool

Expenditures before and after the SMT system integration

SMT integration

Legend: Expenditures | Internal Benchmark | Planned Costs

X-axis: May 12, Sep 12, Jan 13, Mar 13, May 13, Jul 13, Sep 13, Jan 14, Mar 14, May 14, Jul 14, Nov 14

Highlights

>> Successful cooperation between academics and industry
>> Customisation of an open source SMT system
>> Integration into a CAT-tool workflow
>> Implementation of an algorithm that gives more weight to in-domain data
>> ROI within two years and cost savings of more than 30%

University of Zurich UZH

UZH: Prof. Dr. Martin Volk; Dr. Rico Sennrich; Dr. Mark Fishel; Katrin Affolter; Mathias Müller **Finnova:** Roberto Nespeca; Sandra Roth

finnova